ABSTRACT

Title of Document: A RELIABLE TRAVEL TIME PREDICTION SYSTEM WITH SPARSELY DISTRIBUTED DETECTORS

Nan Zou

Directed By: Dr. Gang-Len Chang, Professor
Department of Civil and Environmental Engineering

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 Traffic 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.

This study aims to develop a travel time prediction system that needs only a small number of reliable traffic detectors to perform accurate real-time travel time predictions under recurrent traffic conditions. To ensure its effectiveness, the proposed system consists of three principle modules: travel time estimation module, travel time prediction module, and the missing data estimation module.

The travel time estimation module with its specially designed hybrid structure is responsible for estimating travel times for traffic scenarios with or without

sufficient field observations, and for supplying the estimated results to support the prediction module.

The travel time prediction module is developed to take full advantage of various available information, including historical travel times, geometric features, and daily/weekly traffic patterns. It can effectively deal with various traffic patterns with its multiple embedded models, including the primary module of a multi-topology Neural Network model with a rule-based clustering function and the supplemental module of an enhanced *k*-Nearest Neighbor model.

To contend with the missing data issue, which occurs frequently in any real-world system, this study incorporates a missing data estimation module in the travel time prediction system, which is based on the multiple imputation technique to estimate both the short- and long-term missing traffic data so as to avoid interrupting the operations.

The system developed in this study has been implemented with data from 10 roadside detectors on a 25-mile stretch of I-70 eastbound, and its performance has been tested against actual travel time data collected by an independent evaluation team. Results of extensive evaluation have indicated that the developed system is capable of generating reliable prediction of travel times under various types of traffic conditions and outperforms both state-of-the-practice and state-of-the-art models in the literature. Its embedded missing data estimation models also top existing methods and are able to maintain the prediction system under a reliable state when one of its detectors at a key location experience the data missing rate from 20% to 100% during uncongested, congested and transition periods.

A RELIABLE TRAVEL TIME PREDICTION SYSTEM WITH SPARSELY DISTRIBUTED DETECTORS

By

Nan Zou

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Advisory Committee:
Professor Gang-Len Chang, Chair
Professor Ali Haghani
Professor Paul Schonfeld
Professor Martin Dresner
Assistant Professor Cinzia Cirillo

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 Traffic 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 this issue, transportation professionals have proposed and implemented a variety of systems for providing travel times in the past two decades. However, most real world systems have provided only travel times of completed trips, or based only on the current traffic conditions, not the en-route trips or for pre-trip planning.

Traditionally, travel time prediction models are based on the 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 and evaluated 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 an en-route trip, 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 estimate the future time-varying traffic demand, however, is also quite a complex task; and it demands the proposed model not only to best use 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 as 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.
- It 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 various recurrent congestion conditions and effectively dealing with related issues during real-time operations.

Intending to embody all 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 incur in a real-world system.

1.3 Organization of the Dissertation

Based on the proposed research objectives, the final dissertation will be organized into 8 chapters. The interrelations among those tasks are illustrated in Figure 1.1. A brief introduction of each 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 architecture 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 three principal modules: travel time estimation module, travel time prediction module, and missing data estimation 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 the traffic conditions into pre-specified groups, and then applies the liner 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 the clusters without sufficient samples. The developed hybrid travel time estimation model has been calibrated and validated with actual detector data obtained from 10 detectors on a 25-mile stretch of I-70 eastbound. The model evaluations include results for individual links and sub-segments that consist of multiple links.

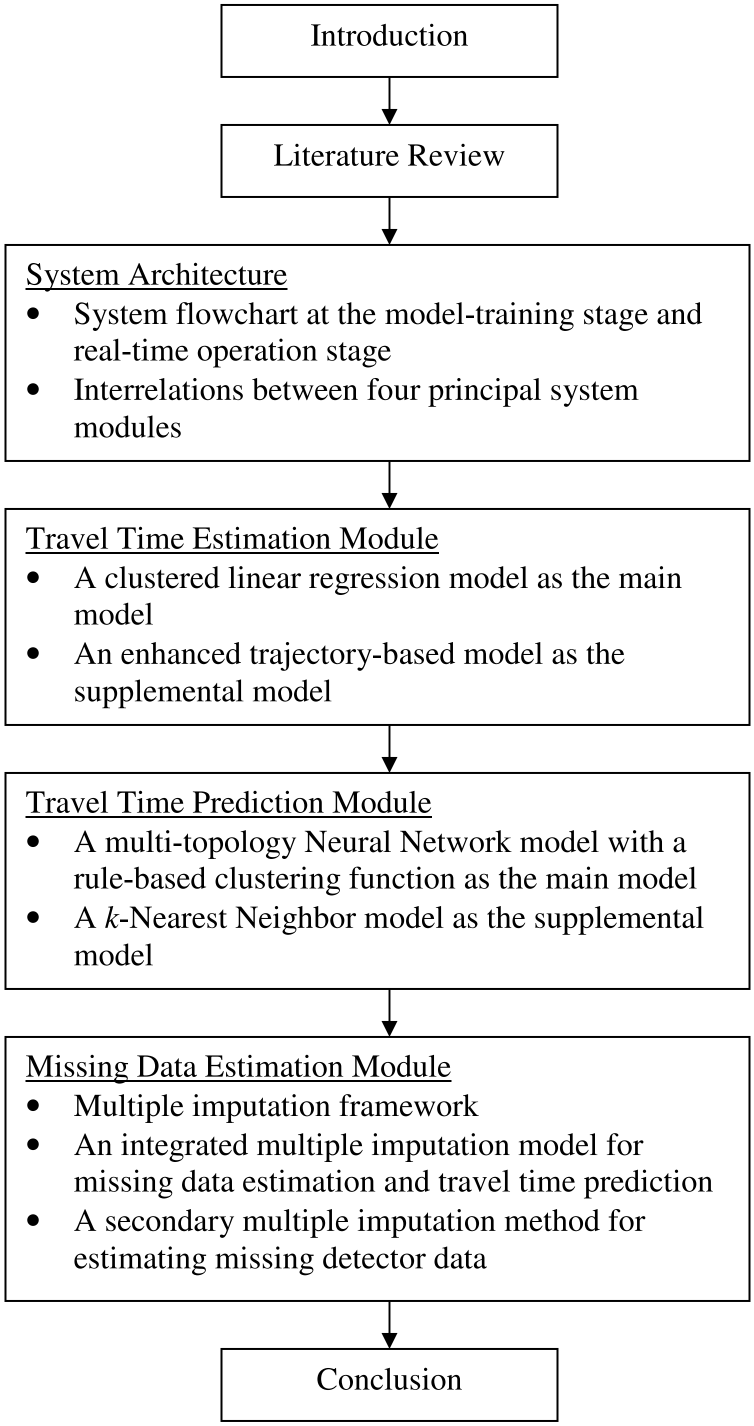


Figure 1-1 Interrelations between primary research tasks

Chapter 5 proposes a hybrid travel time prediction model for freeway segments with a large detector spacing. The hybrid travel time prediction model takes full advantage of a multi-topology Neural Network model with a rule-based clustering function and a *k*-Nearest Neighbor model to provide reliable travel time predictions under recurrent congestion patterns. The multi-topology Neural Network model, which is the main model of the hybrid model, categorize the traffic conditions into pre-defined groups with its embedded clustering rules, and then apply different Multi-layer Perceptron (MLP) or Time Delayed Neural Network (TDNN) model to fit the properties of the recurrent congestion patterns in each cluster. The *k*-Nearest Neighbor model serves as the supplemental component to take advantage of the rich historical travel time information, when available, for a reliable travel time prediction. It has been modified and modeled to take into account traffic characteristics and both daily as well as weekly traffic patterns. The numerical example is based on historical travel times estimated from the same dataset used in Chapter 4. The predicted travel times have been compared with both estimated travel times and actual collected travel times.

Chapter 6 develops a multiple imputation framework for travel time predictions under the impact of missing data, which includes one traditionally modeled multiple imputation method that imputes the missing detector data, and one integrated multiple imputation model that imputes the missing data and predicts the travel time at the same time. Both models overcome the issue of commonly-seen high variations of the detector data in a short period and offer an estimate on the reliability of the imputed or predicted data, which can serve as an important indicator for the developed travel time prediction system to temporarily suspend the outputs for the impacted segments to avoid potential large errors. The missing data patterns from 10 roadside detectors on a 25-mile stretch of freeway segment are summarized in this chapter, along with numerical examples with both missing data imputation models developed in this study and those commonly-used in the literature and real-world systems.

Chapter 7 summarizes the research findings and potential applications for this research to be implemented in ATIS system or traffic control systems. The future research direction is also included in this chapter, which includes integrating a module for detecting non-recurrent congestions and how to model the response from drivers to the prediction results, for example, a variable-rate toll system.

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 in which some input variables are dependent on 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 condition.

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, properly calibrated detectors still cannot be guaranteed to operate at a desirable level of high detection accuracy. Unpredictable traffic count measurement errors 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:

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

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

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

d is the detector ID (numbered from upstream to downstream);

$v_d(t)$ 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 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 area 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). Researchers 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 account of 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 as being used in practice to predict travel times.

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 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 the day of 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.

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

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

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.

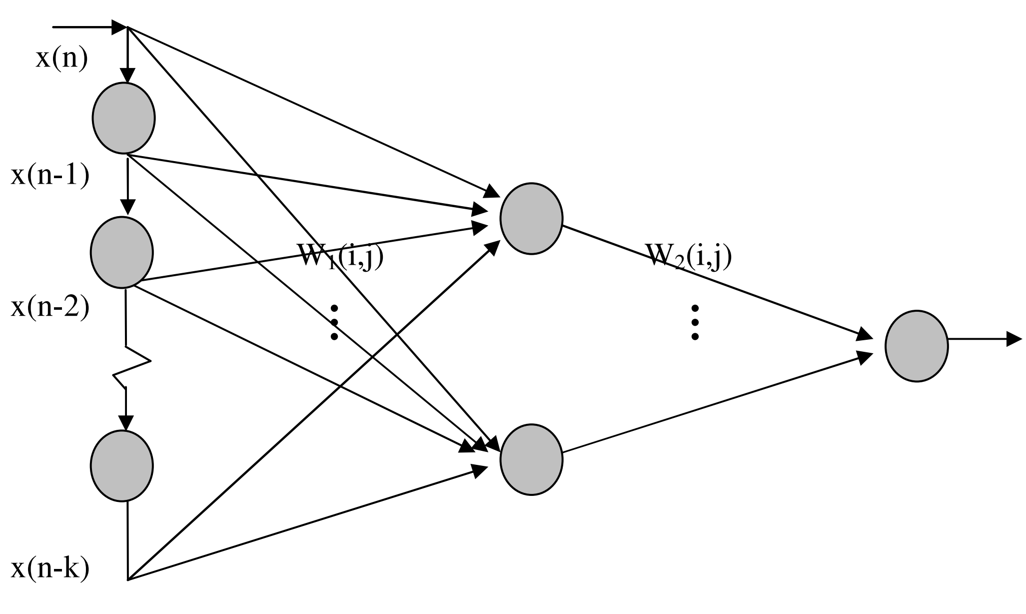


Figure 2-1 Example topology of a Time-Delay Neural Network

Except for the time-delay recurrent Neural Network models, Neural Network 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 and important to its performance; it may be the most significant factor that results in the poor performance of Neural Network models reported in some literature (Smith and Demetsky, 1996; Kirby et al., 1997).

In the literature, 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 reported 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 regular 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 in concept 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 not enough 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 using multiple models. Similar to the nonparametric approaches, hybrid methods 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 procedures for the calibration of such models to various locations.

2.4 Simulated and Real-World Application Systems

Many experimental systems have been implemented worldwide to provide 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 prediction headways 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.

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 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 Conclusion

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 dense detector distribution. 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 reliably cover the targeted freeway segment for travel time prediction, the more likely the responsible agency will be able to afford 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; and (3) estimate missing or delayed data to extend the system's reliability.

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 three principle components: a travel time estimation module, a travel time prediction module, and a missing data estimation 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 system reliability.

Step 2: Long-term data 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 in the real-time operation stage, the travel time prediction module considers weekly traffic patterns. Therefore, there needs to be a fairly long data collection period to make sure 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.

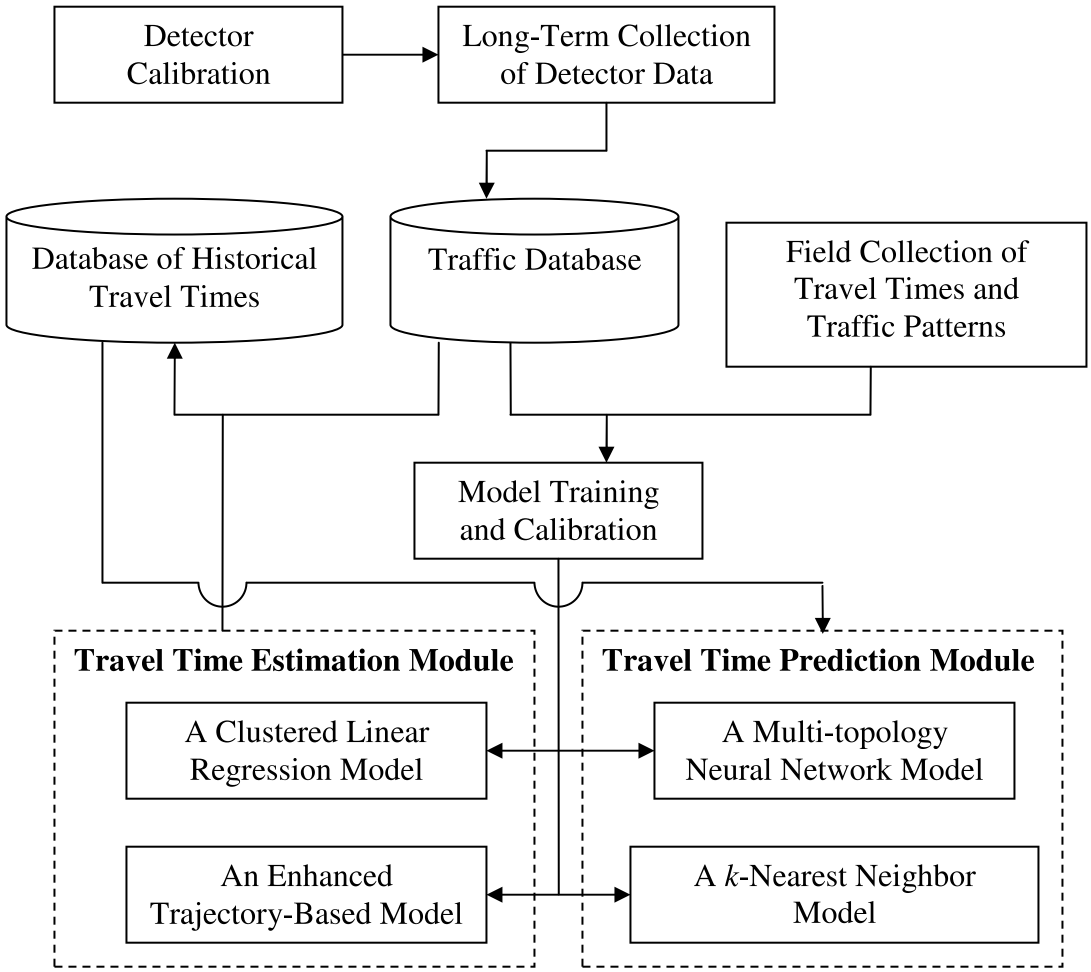


Figure 3-1 System flowchart for the model-training stage

Step 3: Collection of traffic patterns and actual travel times

The proposed system needs information about recurrent traffic patterns to determine critical lanes before its model components are calibrated. The travel time estimation module also requires actual travel time information to calibrate 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 multi-topology Neural Network model requires the analysis of the historical daily traffic patterns in critical lanes to determine its parameters. The *k*-Nearest Neighbor model needs further analysis on weekly traffic patterns.

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

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.

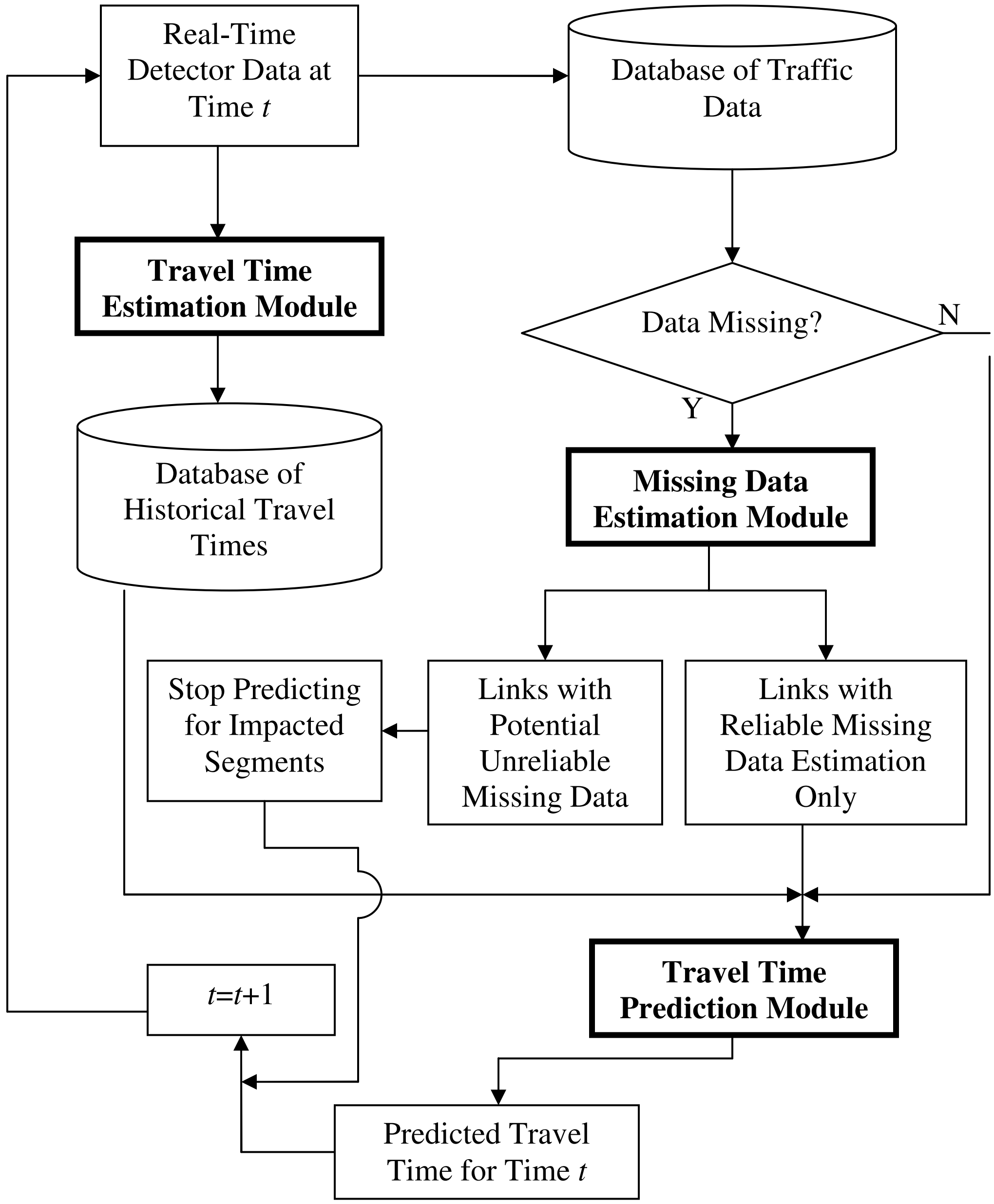


Figure 3-2 System operational flowchart for the real-time operation stage

Step 2: Missing data estimation

The missing data estimation module will perform a test on those identified critical links 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 3: 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 unreliable missing data from traffic detectors.

Step 4: 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 three principal modules: a travel time estimation module, a travel time prediction module, and a missing data estimation 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 combines one multi-topology Neural Network model with a rule-based clustering function and a *k*-Nearest Neighbor model to improve prediction accuracy.

With the clustering rules determined based on the analysis of historical daily traffic patterns, the multi-topology Neural Network is able to group traffic scenarios with similar characteristics and apply a customized Neural Network model to one scenario. When the historical travel time database is rich enough to find *k* historical traffic scenarios similar to the current traffic condition, the developed travel time prediction system can take full advantage of historical travel times and on-line detected traffic conditions with the supplemental *k* Nearest Neighbor model. 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 traffic patterns and geometric features of the target segment.

3.3.3 Missing Data Estimation Module

Missing data is a critical issue that often plagues any on-line system. Most models for on-line systems 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 of the travel time prediction module, the proposed travel time prediction system will stop the prediction on the affected segments until reliable data becomes available.

3.4 Conclusion

This chapter presented operational flowcharts of the travel time prediction system with sparsely distributed detectors in both the model-training stage and the real-time operation stage. To contend with the many technical and compatibility issues, the proposed system consists of three main modules: a travel time estimation module, a travel time prediction module, and a missing data estimation 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, to avoid system interruption.

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 a 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 example traffic conditions commonly seen near a lane drop point. During congested periods, traffic conditions in four subsegments, 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.

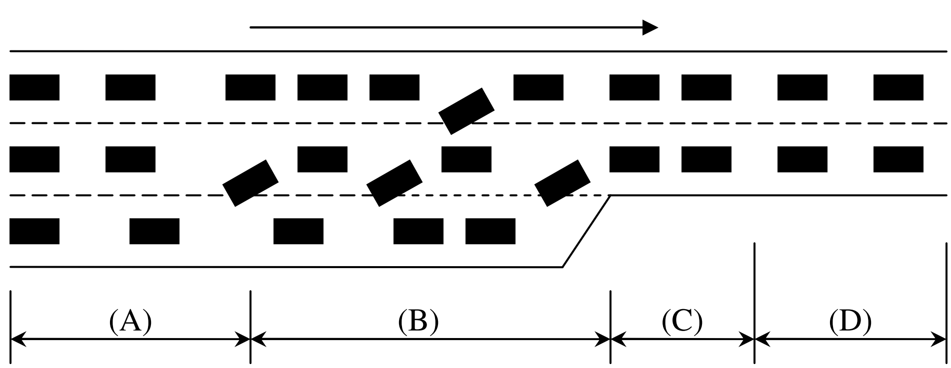


Figure 4-1 Congestion pattern near a lane drop point

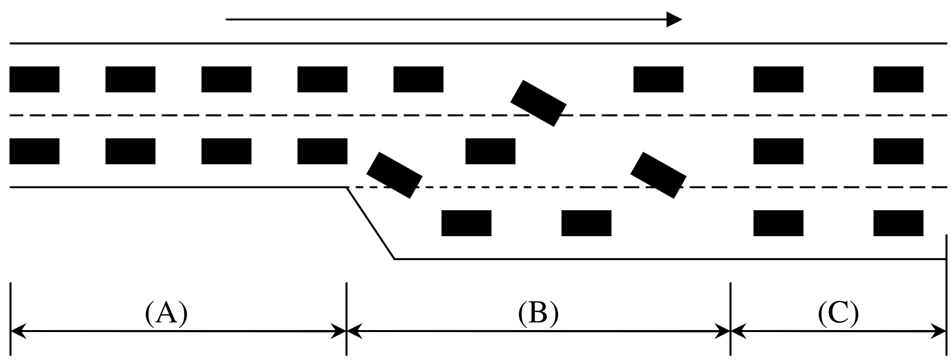
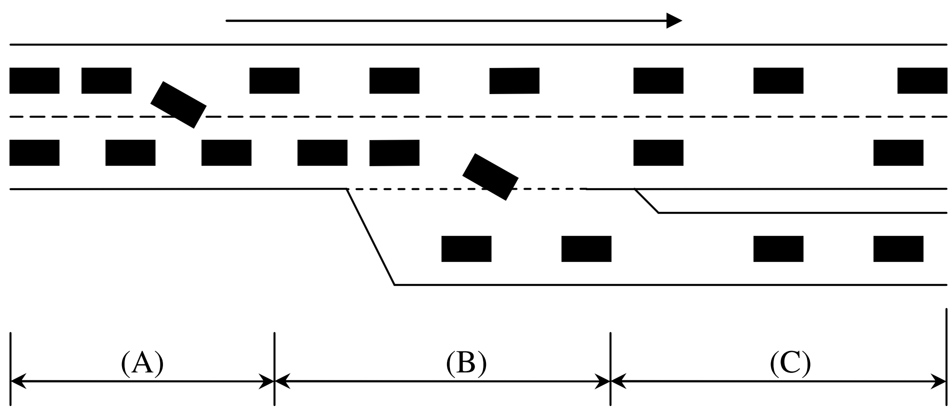
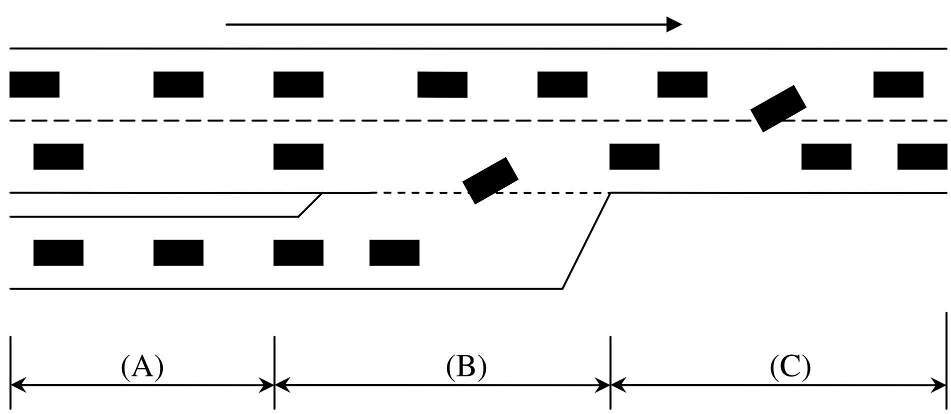


Figure 4-2 Congestion pattern near a lane addition point



(a)



(b)

Figure 4-3 (a) Congestion pattern near an off-ramp;
(b) Congestion pattern near an on-ramp

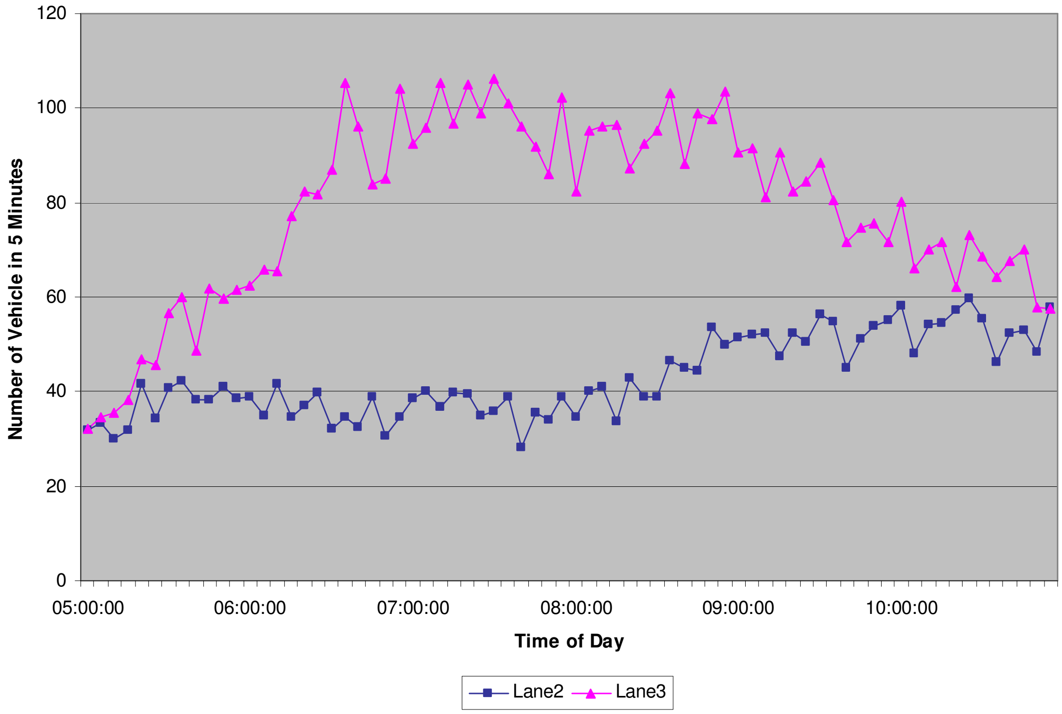


Figure 4-4 Average vehicle counts in 5-minute intervals on four Thursdays in July, 2006 at Exit 87A on I-70

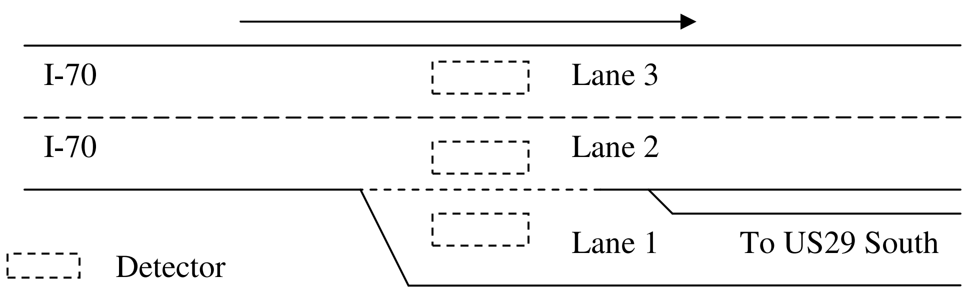


Figure 4-5 Geometry of I-70 at Exit 87A

- 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 <u>A Hybrid Travel Time Estimation Model</u>

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 a 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.

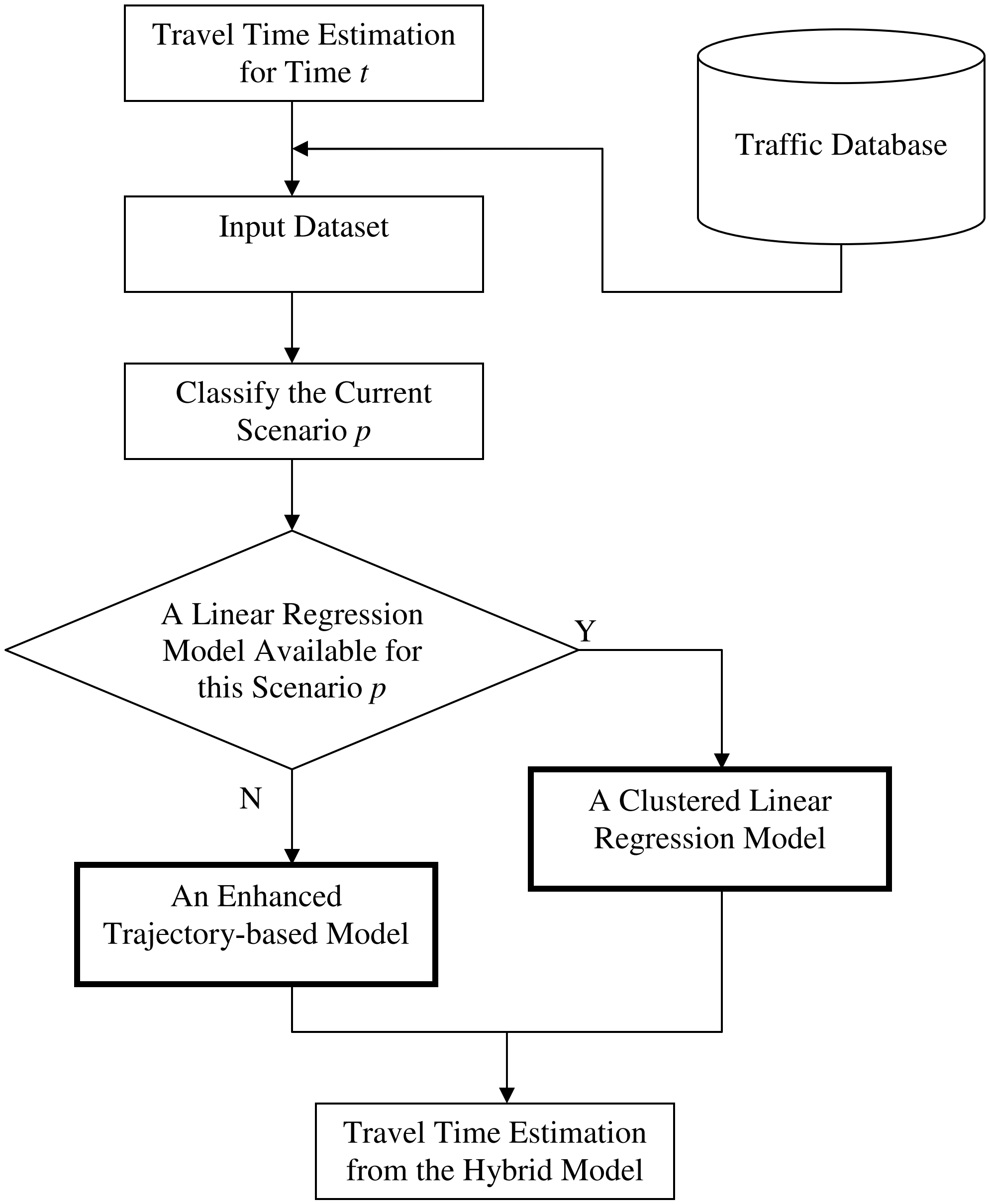


Figure 4-6 Flowchart of the hybrid travel time estimation model

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 d is the detector ID (numbered from upstream to downstream); and

$\tau_d^j(t)$ is the travel time for the vehicle to traverse the j^{th} half of the link (d, d+1) with departure time t (j=1 or 2).

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

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

where L_d is the length of Link (d, d+1).

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; and

$\hat{u}_d^{Thru}(t, \Delta t)$ is the average speed at Detect *d* during time (*t*, *t*+Δt).

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 seems 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 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):

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

where $\hat{u}_d^{Thru}(t,\Delta t)$ is the average speed of through traffic at Detector d during time (t, t+Δt).

a_{ij}^{k} are coefficients;

l_a is lane ID (numbered from right to left);

$\mathbf{CLT}_{d,d+1}^{d}(p)$ is the set of 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;

$\mathbf{CLR}_{d,d+1}^{d}(p)$ is the set of 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.}; and

$u_{d,l_a}(t,\Delta t)$ is the average speed in Lane l_a at Detector d during time (t, t+Δt).

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.

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

where g is the average vehicle length;

$v_{d,l_a}(t,\Delta t)$ is the average flow rate in Lane l_a at Detector *d* during time (t, $t+\Delta t$); and

$o_{d,l_a}(t,\Delta t)$ is the average occupancy in Lane l_a at Detector *d* during time (t, $t+\Delta t$).

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:

$$\begin{aligned}
\tau_d(t) = & \sum_{la\in \mathbf{CLT}^{\mathbf{d}}_{d,d+1}(p)} b^{T,p}_{d,la}\frac{o_{d,la}(t,\gamma^d_p\tau^E_d(p))}{v_{d,la}(t,\gamma^d_p\tau^E_d(p))} + \sum_{la\in \mathbf{CLR}^{\mathbf{d}}_{d,d+1}(p)} b^{R,p}_{d,la}\frac{o_{d,la}(t,\gamma^d_p\tau^E_d(p))}{v_{d,la}(t,\gamma^d_p\tau^E_d(p))} \\
& + \sum_{la\in \mathbf{CLT}^{\mathbf{d+1}}_{d,d+1}(p)} b^{T,p}_{d+1,la}\frac{o_{d,la}(t+\gamma^d_p\tau^E_d(p),(1-\gamma^d_p)\tau^E_d(p))}{v_{d,la}(t+\gamma^d_p\tau^E_d(p),(1-\gamma^d_p)\tau^E_d(p))} \\
& + \sum_{la\in \mathbf{CLR}^{\mathbf{d+1}}_{d,d+1}(p)} b^{R,p}_{d+1,la}\frac{o_{d,la}(t+\gamma^d_p\tau^E_d(p),(1-\gamma^d_p)\tau^E_d(p))}{v_{d,la}(t+\gamma^d_p\tau^E_d(p),(1-\gamma^d_p)\tau^E_d(p))} + b^{0,p}_d
\end{aligned} \tag{4.8}$$

where $b^{T,p}_{d,la}$ is the coefficient of the la^{th} lane in $\mathbf{LT}^*_{d,d+1}(d,p)$ at Detector *d* under the p^{th} traffic scenario for Link (d, $d+1$);

$b_{d+1,la}^{T,p}$ is the coefficient of the la^{th} lane in $\mathbf{LT}_{d,d+1}^{*}(d+1,p)$ at Detector d+1 under the p^{th} traffic scenario for Link (d, d+1);

$b_{d,la}^{R,p}$ is the coefficient of the la^{th} lane in $\mathbf{LR}_{d,d+1}^{*}(d,p)$ at detector d under the p^{th} traffic scenario for Link (d, d+1);

$b_{d+1,la}^{R,p}$ is the coefficient of the la^{th} lane in $\mathbf{LR}_{d,d+1}^{*}(d+1,p)$ at detector d+1 under the p^{th} traffic scenario for Link (d, d+1);

$b_{d}^{0,p}$ is the intercept for the p^{th} scenario for Link (d, d+1);

$\mathbf{LT}_{d,d+1}^{*}(d,p)$ is the set of all critical through lanes at Detector d, which significantly contribute to computing the average through traffic condition in Link (d, d+1) under Scenario p; and

$\mathbf{LR}_{d,d+1}^{*}(d,p)$ is the set of all critical ramp lanes at Detector d, which significantly contribute to computing the average through traffic condition in Link (d, d+1) under Scenario p.

In order to estimate travel times with Eq. 4.8, one needs to estimate γ_{p}^{d}, 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:

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

(4.9)

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. 4.9 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. 4.10 and will then apply Eq. 4.9 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.

$$O(x,t)=\begin{cases} o_d(t+\frac{x-x_d}{u_c^{\max}},t+\frac{x-x_d}{u_c^{\min}}) & ,\text{if } x-x_d<\hat{x} \\ o_{d+1}(t-\frac{x_{d+1}-x}{u_c^{\min}},t-\frac{x_{d+1}-x}{u_c^{\max}}) & ,\text{if } x_{d+1}-x<\hat{x} \\ o_d(t+\frac{\hat{x}-x_d}{u_c^{\max}},t+\frac{\hat{x}-x_d}{u_c^{\min}}) & \\ +\frac{(x-x_d-\hat{x})}{\hat{x}} & \\ \times(o_{d+1}(t-\frac{x-(x_{d+1}-\hat{x})}{u_c^{\min}},t-\frac{x-(x_{d+1}-\hat{x})}{u_c^{\max}}) & ,\text{otherwise} \\ -o_d(t+\frac{\hat{x}}{u_c^{\max}},t+\frac{\hat{x}}{u_c^{\min}})) & \end{cases} \tag{4.10}$$

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

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

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

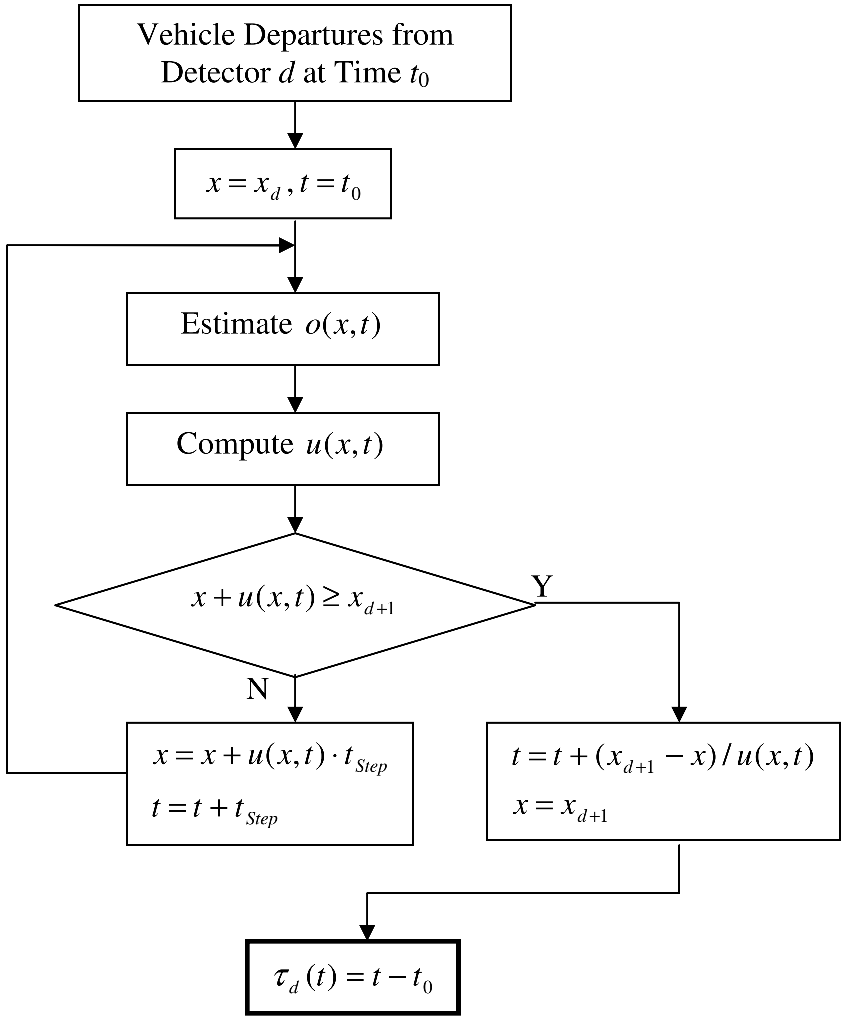


Figure 4-7 Flowchart of the enhanced trajectory-based travel time estimation model

4.6 Numerical Examples

In order to evaluate the effectiveness of the proposed hybrid travel time estimation model on long segments with various geometry features, this study includes a detailed performance analysis based on traffic datasets obtained from real-world detectors. This section will first introduce geometry features, detector locations, traffic patterns and collected travel times of the test site, followed by both link-based and segment-based performance comparison with other models mentioned in the literature to demonstrate the advantage of the proposed model on segments with long detector spacing.

4.6.1 Introduction of the Dataset

The dataset for calibrating and evaluating the proposed hybrid travel time estimation system was acquired from ten roadside traffic detectors and some field surveys on a 25-mile stretch of I-70 eastbound between MD27 and I-695 between January 19th, 2006, and August 2nd, 2006. The locations of the ten detectors were determined based on the geometric features and general traffic patterns obtained from historical volume archives and preliminary site surveys. Figure 4.8 shows the locations of ten detectors along the target freeway segment, numbered from upstream to downstream, and Table 4.2 shows their geographic coordination.

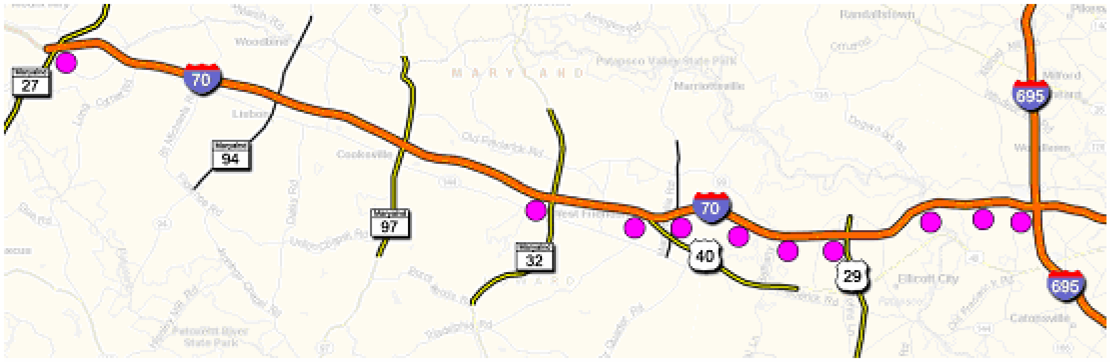


Figure 4-8 Locations of 10 detectors on I-70 eastbound

Table 4-2 Description and geographic locations of ten detectors

Detector ID	Location	Longitude	Latitude
1	About 1000 feet past MD27	-77.163174	39.359605
2	About 500 feet past 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 mileage markers 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

Each traffic detector collects data of traffic count, occupancy and average speed in each lane (except the far left lane at Detector 10) at its location at 30-second intervals. However, this study did not include speed information in either the modeling or validation process because the reliability issues reported by Zou and Wang (2006). Figure 4.9 shows the exact location of each traffic detector.

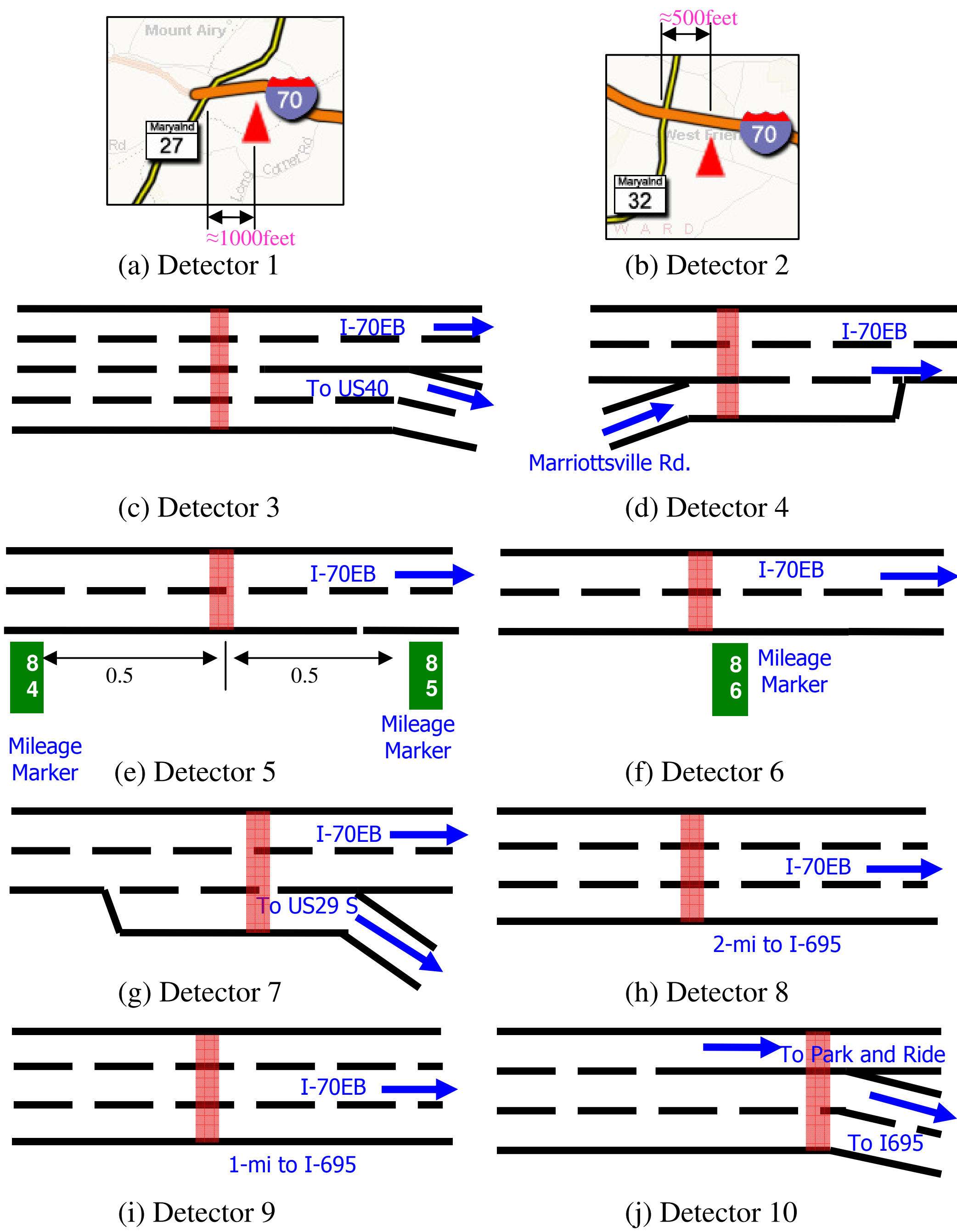


Figure 4-9 (a) to (j) Exact location of each detector

In order to calibrate and validate the developed models, travel time surveys were conducted in the segment either by matching vehicles in videos taken at both upstream and downstream detector locations or by using the GPS devices in the probe vehicles. Table 4.3 shows the schedule of all surveys that have been taken for individual links between two neighboring detectors. Table 4.4 lists all surveys on some subsegments, which consists of more than one link.

Table 4-3 Schedule of all field surveys for individual links

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								
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
4/6/2006	AM			Y						
4/20/2006	AM			Y						
6/13/2006	AM	Y		Y			Y	Y	Y	Y
6/15/2006	PM	Y		Y			Y	Y	Y	Y

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

Table 4-4 Schedule of all surveys for subsegments

Date and Time	**Covered Subsegments**
4/6/2006 AM	(3, 7) and (3, 10)
4/20/2006 AM	(3, 7) and (3, 10)
6/13/2006 AM	Any subsegment between Detector 3 and Detector 10
6/15/2006 PM	Any subsegment between Detector 3 and Detector 10

Note that (d_1, d_2) refers to the roadway segment between Detectors d_1 and d_2, where $d_1 < d_2$.

As shown in Table 4.3, the survey plan was based on the observed daily traffic patterns in the target freeway segment. For example, Links 3-4, 4-5, 5-6, and 6-7 are often very congested in the morning, but are usually not congested in the evening; therefore, no evening surveys were conducted for these segments. In contrast, severe congestion is frequently observed in Links 7-8, 8-9 and 9-10 during both morning and evening peak hours. Therefore, data collection focused on both AM and PM periods for those segments.

Please note that multiple surveys were conducted for certain links to compensate for encountering nonrecurrent congestion patterns, such as accidents. Hence, this study will generally first filter out the data points impacted by incidents/accidents and then calibrate the travel time estimation module using samples in each link that exhibited different recurrent congestion patterns.

4.6.2 Preliminary Analysis of the Dataset

Data availability

Due to communication failures, power outages and/or detector malfunctions, some data was lost during the data collection period. Table 4.5 shows the daily data availability between 2/21/2006 and 3/23/2006 and highlights daily availability of less than 90%. The daily availability is computed as the ratio of number of available data records in a day over the expected number of data points, based on the duration of the data interval (for example, 2,880 data points are expected daily when the duration of the data interval is 30 seconds). Because the travel time estimation module is for off-line use to construct the historical travel time database as mentioned in Chapter 3, this study did not make an additional effort to estimate the missing data for the travel time estimation. The data will be removed from the dataset if a detector experiences a missing rate of more than 5%.

Data reliability

The manufacturer of the traffic detector claims accuracy of more than 95% under light traffic and moderate congestion (occupancy less than 30%). The flow data from the detectors have been validated with the volumes counted from field surveys and found to have less than 5% counting errors during most time intervals. However, the counting errors above 5% but less than 10% were found during some congested periods. Detector calibrations were completed by both the contractor and the manufacturer of the detector.

Table 4-5 Daily availability of detector data between 2/21/2006 and 3/23/2006

%	Detector									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3/23/2006	100	100	100	100	100	100	**89**	99.9	100	100
3/22/2006	100	100	100	100	100	**85.3**	**79.9**	100	100	**43.2**
3/21/2006	100	100	100	100	100	**83.8**	90.1	100	100	**0**
3/20/2006	100	100	100	100	100	**87.9**	91.1	100	100	**0**
3/19/2006	100	100	100	100	100	**89**	94	100	100	**0**
3/18/2006	100	100	100	100	100	**87.6**	97.5	100	100	**0**
3/17/2006	99.7	99.7	99.7	99.8	99.6	93.1	99.5	99.8	99.7	**0**
3/16/2006	100	100	100	100	100	**75.8**	99.8	100	100	**59.8**
3/15/2006	100	100	100	100	100	92.3	99.6	100	100	94.3
3/14/2006	100	100	100	100	100	91.8	99.9	100	100	98
3/13/2006	100	100	100	100	100	93.8	100	100	100	99.9
3/12/2006	100	100	100	100	100	92.8	100	100	100	**85.9**
3/11/2006	100	100	100	100	100	95.3	100	100	100	99.7
3/10/2006	100	100	100	100	100	96.1	100	99.9	100	99.9
3/9/2006	99.8	100	99.7	99.7	100	94.7	99.8	99.9	100	99.4
3/8/2006	100	100	100	100	100	95.5	100	100	100	98.9
3/7/2006	100	99.8	100	100	100	96.2	100	100	100	99.6
3/6/2006	100	100	100	100	100	97.3	100	100	100	99.3
3/5/2006	96.5	95.4	96.7	96.3	95.9	91.4	94.9	96	96	99.1
3/4/2006	100	99.9	100	100	100	96.3	100	100	100	99.1
3/3/2006	100	100	100	100	100	98.3	100	99.9	100	99.5
3/2/2006	100	100	99.9	100	99.9	97.6	100	94.3	99.9	99.9
3/1/2006	100	99.8	99.9	100	100	99.2	100	100	100	100
2/28/2006	100	100	100	100	100	99.5	100	100	100	99.9
2/27/2006	100	99.9	100	100	100	99.4	100	100	97.8	99.9
2/26/2006	100	100	100	100	100	100	100	100	99.9	100
2/25/2006	100	100	100	100	100	100	100	100	100	100
2/24/2006	100	100	100	99.9	**33.7**	99.8	100	100	100	100
2/23/2006	100	99.9	100	100	**13.1**	100	100	99.9	100	100
2/22/2006	99.8	100	99.8	99.8	**71.2**	99.9	99.9	99.8	99.9	99.8
2/21/2006	99.9	100	99.9	99.9	99.9	100	99.9	100	**7.2**	99.9

Volume Drifting

As mentioned in Section 2.2.1, volume drifting is an important issue that prevents the flow-based travel time estimation models from being implemented in a real-world system. Daily volumes at detector pairs (4, 6) and (8, 9) from 6/27/2006 to 7/2/2006 have been summarized in Table 4.6. Neither detector pair has a ramp in between (Figure 4.8); therefore they should report the same daily volumes if every vehicle that has passed the detector stations has been detected correctly. However, the daily volume differences and percentages reported by detectors showed nonsystematic patterns.

Table 4-6 Comparisons of daily volume counts between two detector pairs

	2006-06-27	2006-06-28	2006-06-29	2006-06-30	2006-07-01	2006-07-02
Daily Volume at Detector 4	37040	39121	41595	42707	35190	29891
Daily Volume at Detector 6	37903	39695	42373	43410	35117	29741
Difference	863	574	778	703	-73	-150
Relative Difference	2.33%	1.47%	1.87%	1.65%	-0.21%	-0.50%
Daily Volume at Detector 8	45332	49022	50160	50670	39469	34806
Daily Volume at Detector 9	44979	48945	49796	50449	39314	34784
Difference	-353	-77	-364	-221	-155	-22
Relative Difference	-0.78%	-0.16%	-0.73%	-0.44%	-0.39%	-0.06%

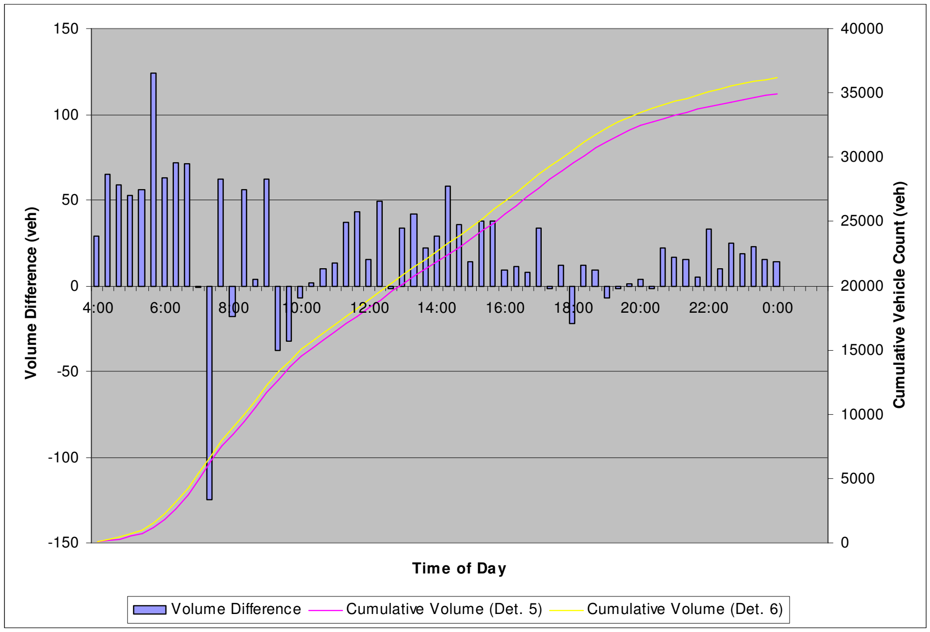
4.6.3 Model Evaluation for Individual Links

Due to the complex geometry features in the study area, a 25-mile freeway segment of I-70 between MD27 and I-695 (Figure 4.8 and Figure 4.9), none of the existing travel time estimation models that take data from traffic detectors can be applied to estimate travel times for all links. However, some links have simpler geometry features and therefore can be used to compare the performance of the proposed hybrid travel time estimation model to the existing approaches.

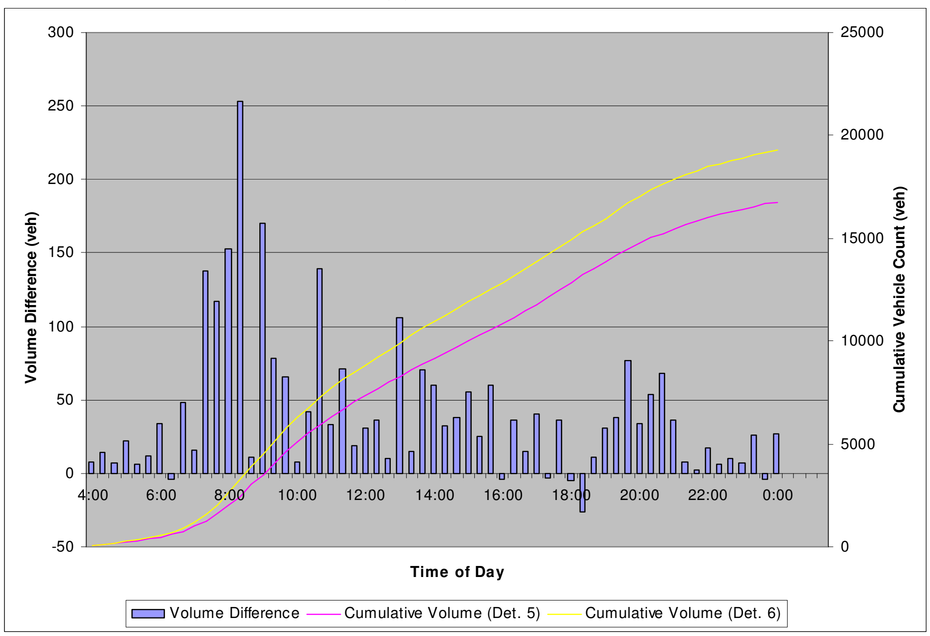
This study selects two links — between Detector 5 (Figure 4.9(e)) and Detector 6 (Figure 4.9(f)) and between Detector 6 and Detector 7 (Figure 4.9(g)) — to compare the performance of travel time estimation models on these links.

<u>Performance Comparison on Link (5, 6)</u>

No ramp or other geometry change (e.g., lane addition or lane drop) exits between Detectors 5 and 6. One on-ramp lane merges into the I-70 mainline segment of two lanes at Detector 4 (Figure 4.9[d]), about 1.12 miles upstream of Detector 5. The nearest geometry change downstream of Detector 6 is about 0.96 miles at Detector 7. Because neither detector in Link (5, 6) is very close to the location of the geometry change, the traffic can be treated as evenly distributed across two through lanes at both detector locations. A flow-based method (Nam and Drew, 1996) has been implemented on this segment and included in the comparison. As proposed by Nam and Drew (1996), an hourly volume-adjusting factor will be introduced to reduce the impact caused by the volume-drifting issues. The performance comparison also includes the original linear piecewise trajectory-based travel time estimation model (van Lint and van der Zijpp, 2003), which requires speed information as an input variable. In the comparison, the speed information will be computed with Eq. 4.10, which estimates speed for the supplemental enhanced trajectory-based model developed in this study.



(a) 1/19/2006



(b) 2/28/2006

Figure 4-10 Distribution of the difference in the detected volume data between Detector 5 and Detector 6 aggregated over each 20-minute interval, and cumulative vehicle counts from 4:00AM to the end of day on (a) 1/19/2006 and (b) 2/28/2006

Figure 4.10 illustrates the distribution of the difference in the volume data between Detector 5 and Detector 6 over each interval of 20 minutes from 4:00AM to the end of day on 1/19/2006 and 2/28/2006, together with the cumulative vehicle counts at both detectors. It shows that the differences vary significantly, especially during peak hours.

The attempt to implement a flow-based model for the Link (5, 6) failed. As shown in Figure 4-11, the model even estimated travel times between 7:00AM and 9:00AM as being less than zero. The estimations it provided for other time periods showed very high fluctuation. This is probably due to the unsystematic distribution of detection errors of the traffic flow.

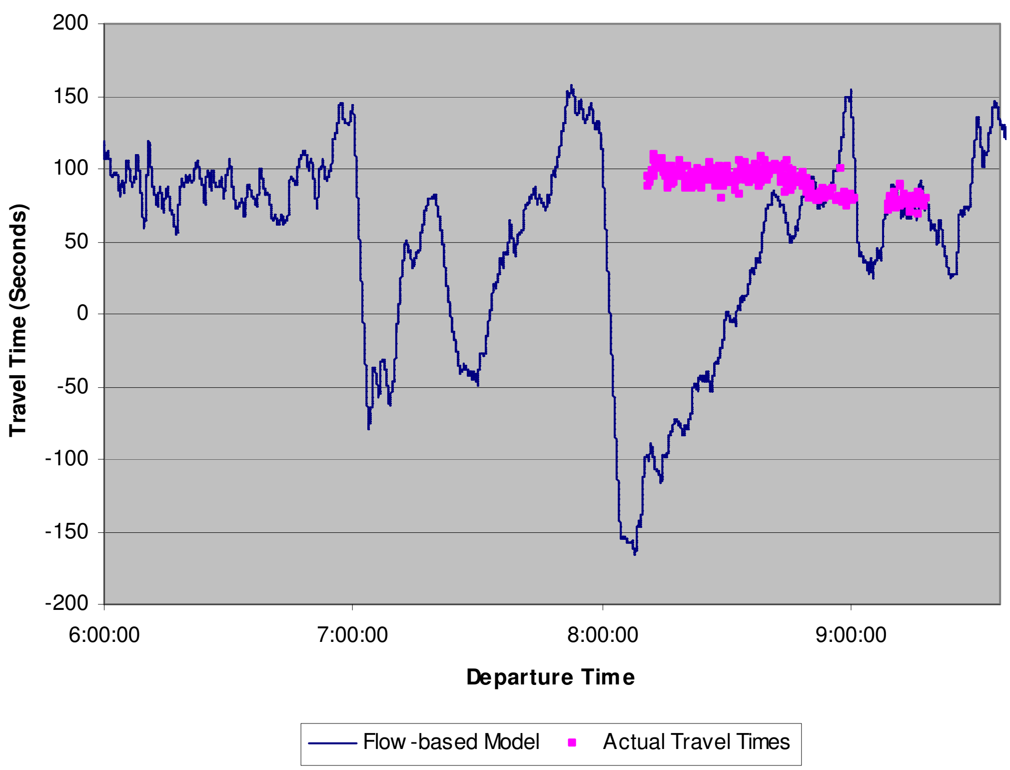


Figure 4-11 Estimated travel times vs. actual travel times between Detectors 5 and 6 on January 19th, 2006

The clustered linear regression (CLR) model and the enhanced trajectory-based (ETB) model, as well as the original piecewise linear speed-based (PLSB) model have been implemented successfully on the Link (5, 6).

As shown in Table 4-7, the developed CLR model categorizes traffic conditions on Link (5, 6) into four scenarios. Scenario 2 congestion is usually caused by high merging volume at Detector 4. An uncongested condition at Detector 7 helps the traffic disperse when vehicles traverse Link (5, 6). In Scenario 3, traffic congestions occurred at both the upstream and downstream of Link (5, 6). Note that, due to traffic merging into the right mainline lane at Detector 4, traffic at Detector 5 is actually not uniformly distributed. Therefore, the criteria are different for lane 1 and lane 2 at Detector 5. All other traffic scenarios that do not have enough observations are categorized into Scenario 4.

Table 4-7 Traffic Scenarios for Link (5, 6)

ID	Description of the Scenario	Detector 5		Detector 6	
		Occ. in Ln. 1	Occ. in Ln. 2	Occ. in Ln. 1	Occ. in Ln. 2
1	No congestion on the link	≤12	≤10	≤10	≤10
2	Congestion at Detector 5; no congestion at Detector 6	>12	>10	≤10	≤10
3	Congestion at both Detectors 5 and 6	>12	>10	>10	>10
4	Other	Other combinations			

Note that the occupancy refers to the average occupancy of a 4-minute period starting from 2 minutes before the current time. The unit of occupancy is %.

Table 4.8 shows all the parameters for Eq. 4.10 for Link (5, 6), which is required by both ETB and PLSB models. The parameters were determined through field observations, following Zou and Wang's (2006) approach.

This study first explores the performances of CLR, ETB and PLSB models in Traffic Scenarios 2 and 3 individually. Then an overall comparison between the proposed hybrid model (HM) and PLSB will be presented.

Table 4-8 Parameters for Equation 4.10 for Link (5, 6)

Parameter	Value
u_{free}	67.23 mph
u_{cong}	10 mph
u_{min}	1 mph
o_{free}	10
o_{cong}	30
o_{max}	45
M	2.3
N	1.5

Traffic Scenario 2

There were a total of 446 samples of actual travel times under the Traffic Scenario 2 on Link (5, 6) collected on January 19th, 2006 (Thursday), and February 28th, 2006 (Tuesday). About 92% of the samples (411 observations) were randomly selected to construct the dataset for calibrating the CLR model. The rest of the samples (35 observations) were used to evaluate the model's performance. The model identifies the only critical lane: Lane 2 at Detector 5 in Scenario 2 on Link (5, 6). Eq. 4.11 shows the CLR model for Scenario 2 on Link (5, 6).

$$\tau_5(t) = 68.171 + 4.561\frac{o_{5,2}(t,270)}{v_{5,2}(t,270)} \qquad (4.11)$$

when $o_{5,1}(t-120,240) > 10$, $o_{5,2}(t-120,240) > 12$, $o_{6,1}(t-120,240) \le 10$

and $o_{6,2}(t-120,240) \le 10$.

In Scenario 2, the observed travel times were distributed between 78 seconds and 124 seconds. Table 4-9 shows comparisons of all observations, samples with shorter travel times (≤ 95 seconds, 20 observations) and those with longer travel times (> 95 seconds, 15 observations) for all three models. Eq. 4-12 defines the comparison indicators, the average absolute error (AAE) and the average absolute relative error (AARE). Overall, both CLR and ETB provided better performance than PLSB. Among these, ETB had a lower AAE, and CLR model performed better with the increase of travel time.

$$\text{AAE} = \frac{1}{N}\sum_{n=1}^{N} | \tau_n - \hat{\tau}_n |$$

$$\text{AARE} = \frac{1}{N}\sum_{n=1}^{N} \frac{| \tau_n - \hat{\tau}_n |}{\tau_n} \tag{4.12}$$

where N is the number of data samples available for comparison,

n is the index of the data sample,

τ_n is the n^{th} observed travel time, and

$\hat{\tau}_n$ is the travel time from the model.

Table 4-9 Overall performance comparison for Scenario 2 on Link (5, 6) with field data

	All Samples (35 Observations)		Travel Times ≤ 95 sec. (20 Observations)		Travel Times > 95 sec. (15 Observations)	
	AAE (Sec.)	AARE (%)	AAE (Sec.)	AARE (%)	AAE (Sec.)	AARE (%)
CLR	5.63	6.57	6.16	8.31	4.46	4.26
ETB	5.14	5.43	3.71	4.19	7.2	7.08
PLSB	6.17	6.49	5.47	5.86	7.67	7.33

AAE: Average absolute error.
AARE: Average absolute relative error.

Traffic Scenario 3

The Traffic Scenario 3 covers situations in which congestion exists at both Detector 5 and Detector 6. A total of 340 observations fell into Scenario 3, 307 of which were randomly selected to calibrate CLR; the other 33 observations were used to evaluate all three models. In this Scenario, CLR identifies its two dependent variables as the average traffic conditions at Detector 5 and Detector 6. Eq. 4.14 shows the CLR model for Scenario 3 on Link (5, 6).

$$\tau_5(t) = 35.626 + 6.952\frac{o_5(t,390)}{v_5(t,390)} + 2.301\frac{o_6(t,90)}{v_6(t,90)} \tag{4.13}$$

when $o_{5,1}(t-120,240) > 10$, $o_{5,2}(t-120,240) > 12$, $o_{6,1}(t-120,240) > 10$

and $o_{6,2}(t-120,240) > 10$.

Table 4-10 summarizes the comparison results, which show that CLR provided significant improvement over ETB and PLSB in this scenario, with an AAE of 6.60 seconds, which is less than 25% of that from the PLSB model, and an AARE of 4.79%. ETB was able to provide an AAE of 19.48 seconds, which was about 7 seconds less than that of PLSB model.

Table 4-10 Overall performance comparison for Scenario 3 on Link (5, 6)

	All Samples (33 Observation)	
	AAE (Sec.)	AARE (%)
CLR	6.60	4.79
ETB	19.48	13.35
PLSB	26.33	17.65

AAE: Average absolute error.
AARE: Absolute relative error.

Traffic Scenarios 1 and 4

60 observations of travel times were between 74 seconds and 91 seconds in Scenario 1. The 151 observations of travel times in Scenario 4 are from 75 seconds to 169 seconds. As shown in Table 4-11, ETB outperformed PLSB in both scenarios.

Table 4-11 Overall performance comparison for Scenarios 1 and 4 on Link (5, 6)

	Scenario 1 (60 Observations)		Scenario 4 (151 Observations)	
	AAE (Sec.)	AARE (%)	AAE (Sec.)	ARE<10% (%)
ETB	2.67	3.33	7.22	6.54
PLSB	2.92	3.67	8.69	7.66

AAE: Average absolute error.
AARE: Absolute relative error.

Overall

Overall, the developed hybrid model, which uses a CLR model as the main model and an ETB model as the supplemental model, reliably estimated travel times for Link (5, 6). The hybrid model had an overall average absolute error of 6.02 seconds for all traffic scenarios and PLSB model had an average absolute error of 9.22 seconds. The flow-based model failed to provide reliable estimates of travel times in Link (5, 6).

<u>Performance Comparison on Link (6, 7)</u>

There is no ramp or other type of geometry change between Detectors 6 and 7. However, flow-based models cannot be applied to this link due to the violation of the first-depart-first-arrive assumption caused by uneven congestion patterns between the ramp lane and through lanes at Detector 7 (Figure 4-9g) during peak hours. Therefore, only PLSB model was implemented in the comparison.

As shown in Table 4-12, the developed CLR model categorizes traffic conditions on Link (6, 7) into 7 scenarios. Congestion in Scenario 2 is usually caused by the dispersion of heavy upstream congestion with no congestion at the downstream detector of the segment (Detector 7). Scenarios 3 to 8 are six types of traffic patterns that have congestion at Detector 7 as being caused by the through lane only, the ramp lane only, or the combination of both types of lanes. As mentioned in previous sections, only lanes 1 and 3 at Detector 7 are included as critical lanes.

Traffic Scenario 2

On January 20th, 2006 and February 28th, 2006, there were a total of 31 samples were observed in Scenario 2 on Link (6, 7). This study applies an evaluation method to first calibrate model parameters, using 28 randomly selected samples, and then tested the model with the remaining 3 samples. This evaluation process was repeated for 10 times to calculate the final average performance of CLR model. Eq. 4.15 shows the parameters determined from one of the 10 evaluations. The ETB model and PLSB model have been evaluated with the same dataset.

$$\tau_6(t) = 22.241 + 7.688 \frac{o_{6,2}(t,270)}{v_{6,2}(t,270)} \tag{4.14}$$

when $o_{6,1}(t-120,240) > 8$, $o_{6,2}(t-120,240) > 10$, $o_{7,1}(t-120,240) \le 8$ and $o_{7,3}(t-120,240) \le 10$.

Table 4-12 Traffic scenarios for Link (6, 7)

ID	Description of the Scenario	Detector 6		Detector 7	
		Occ. in Ln. 1	Occ. in Ln. 2	Occ. in Ln. 1	Occ. In Ln. 3
1	No congestion on the link	≤8	≤10	≤8	≤10
2	Congestion at Detector 6; no congestion at Detector 7	>8	>10	≤8	≤10
3	No congestion at Detectors 6 and congestion at Detector 7 caused by the off-ramp (Lane 1)	≤8	≤10	>8	≤10
4	No congestion at Detectors 6 and congestion at Detector 7 caused by the through lane (Lane 3)	≤8	≤10	≤8	>10
5	No congestion at Detectors 6 and congestion at Detector 7 caused by both the through lane and the off-ramp	≤8	≤10	>8	>10
6	Congestion at Detectors 6 and congestion at Detector 7 caused by the off-ramp (Lane 1)	>8	>10	>8	≤10
7	Congestion at Detector 6 and congestion at Detector 7 caused by the through lane (Lane 3)	≤8	>10	≤8	>10
8	No congestion at Detectors 6 and congestion at Detector 7 caused by both the through lane and the off-ramp	≤8	≤10	>8	>10
9	Other	Other combinations			

Note that the occupancy refers to the average occupancy of a 4-minute period starting from 2 minutes before the current time.

Table 4-13 Overall performance comparison of accuracy for Traffic Scenario 2 on Link (6, 7)

	AAE (Sec.)	AARE (%)
CLR	3.36	2.84
ETB	11.67	9.67
PLSB	11.00	8.33

AAE: Average absolute error.
AARE: Absolute relative error.

Table 4-13 summarizes the performance comparisons of all three models in Scenario 2 on Link (6, 7), in which the observed travel times were between 63 seconds and 94 seconds. The CLR model had the least AAE of 3.36 seconds, which is less than 1/3 of those from the ETB and PLSB models, which showed similar performance.

Traffic Scenario 6

On January 20th, 2006, and February 28th, 2006, there were a total of 22 samples observed in Scenario 6 on Link (6, 7). Similar to Scenario 2, this study calibrated the linear regression model in this scenario with 19 randomly selected samples and then evaluated the model with the remaining 3 samples. Due to the limited samples for model validation, this study repeated the random selection process 10 times for accuracy validation. The CLR model has the least AAE of 6.00 seconds. The ETB model and the PLSB model have AAEs of 14.50 and 16.00 seconds, respectively.

Other Traffic Scenarios

All traffic scenarios other than Scenarios 2 and 6 had too few observed actual travel times to calibrate the CLR model. Therefore, only the ETB and PLSB models were implemented for comparison. As shown in Table 4-14, the ETB model developed in this study was able to provide an AAE of 5.36 seconds compared to 10.22 seconds provided by the PLSB model. The ETB model provided the AARE of 7.94%, which outperformed that of 15.03% from PLSB model.

Table 4-14 Overall performance comparison for the traffic scenarios other than 2 and 6 on Link (6, 7)

	AAE (Sec.)	AARE (%)
ETB	5.36	7.94
PLSB	10.22	15.03

AAE: Average absolute error.
AARE: Average absolute relative error.

4.6.4 Model Evaluation on Multiple Links

The model evaluation on multiple links covered the subsegment between Detector 3 (at the split of I-70 and US40) and Detector 10 (at the start of the ramp to I-695) on the I-70 freeway segment. This subsegment often experiences heavy congestion in the morning peak hours on Tuesdays and Thursdays. Therefore, this study conducted two travel time surveys in the morning peak hours on April 6th, 2006 (Thursday) and April 20th, 2006 (Thursday), for the subsegment. The true travel times were obtained by matching vehicles from two videos taken at the beginning and end of the subsegment. 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 4-12 shows the distribution of collected data samples in the subsegment from Detectors 3 to Detector 10 during the survey periods.

This 10-mile long subsegment consists of four interchanges and seven ramps (Figures 4-8 and 4-9). Complex geometric features and high variation in traffic volumes have made this subsegment difficult to develop a travel time estimation model. This research categorized the congestion patterns into different levels based on the range of travel times, so as to have a detailed evaluation of the performance of the developed hybrid travel time estimation model under various traffic conditions. As shown in Figure 4-12 and Table 4-15, congestion was much heavier on April 6th, 2006, which had a maximum travel time of 1290 seconds (21.5 minutes) on the subsegment, which has a free-flow travel time of 520 seconds (8.7 minutes). Data collected on April 20th, 2006, showed that most travel times were between 800 and

1,000 seconds, which exhibited quite a fluctuating pattern between 7:15AM and 8:00AM.

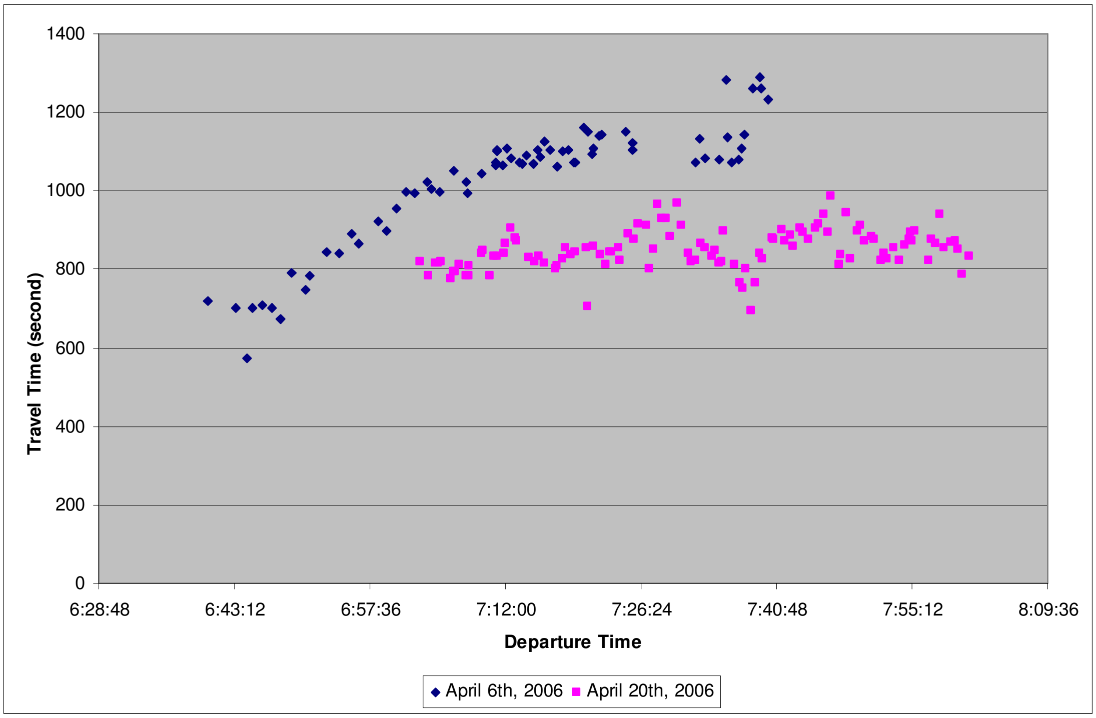


Figure 4-12 The distribution of collected travel times on April 6th, 2006 and April 20th, 2006

Tables 4-15(a) and (b) summarize the performances of the developed estimation model and PLSB model on the subsegment from Detector 3 to Detector 10 against the field data collected on two different days. Figures 4-13(a) and (b) show the distribution of estimated and actual travel times vs. departure time for two days, where the estimated travel times from the developed hybrid model showed a similar trend to the actual travel times and the PLSB model failed to do so. The results from the developed hybrid travel time estimation model showed satisfactory performance over 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 hybrid model was still able to estimate travel times with an error of less than 70 seconds. In heavily congested cases, in which most travel times are greater than twice the free-flow travel time (520 seconds), the developed hybrid model still provided estimates with an AAE of less than 90 seconds. In contrast, the PLSB model had AAEs of more than 2 minutes in all categories. The PLSB model produced a very high AARE of 44.3% under heavy congestion.

Table 4-13(c) shows the overall evaluation results for the transition periods (travel times between 520 and 800 seconds), moderate congestion (travel times between 800 and 1000 seconds) and heavy congestion (travel times greater than 1,000 seconds). For all 185 collected travel times, the hybrid travel time estimation model successfully yielded the estimated travel times with acceptable accuracy, i.e., within 90 seconds.

Table 4-15 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
Hybrid Model	AAE (sec)	51.9	60.3	83.6
	AARE (%)	7.3%	6.6%	7.5%
PLSB	AAE (sec)	122.0	329.6	493.8
	AARE (%)	16.7%	35.5%	44.3%

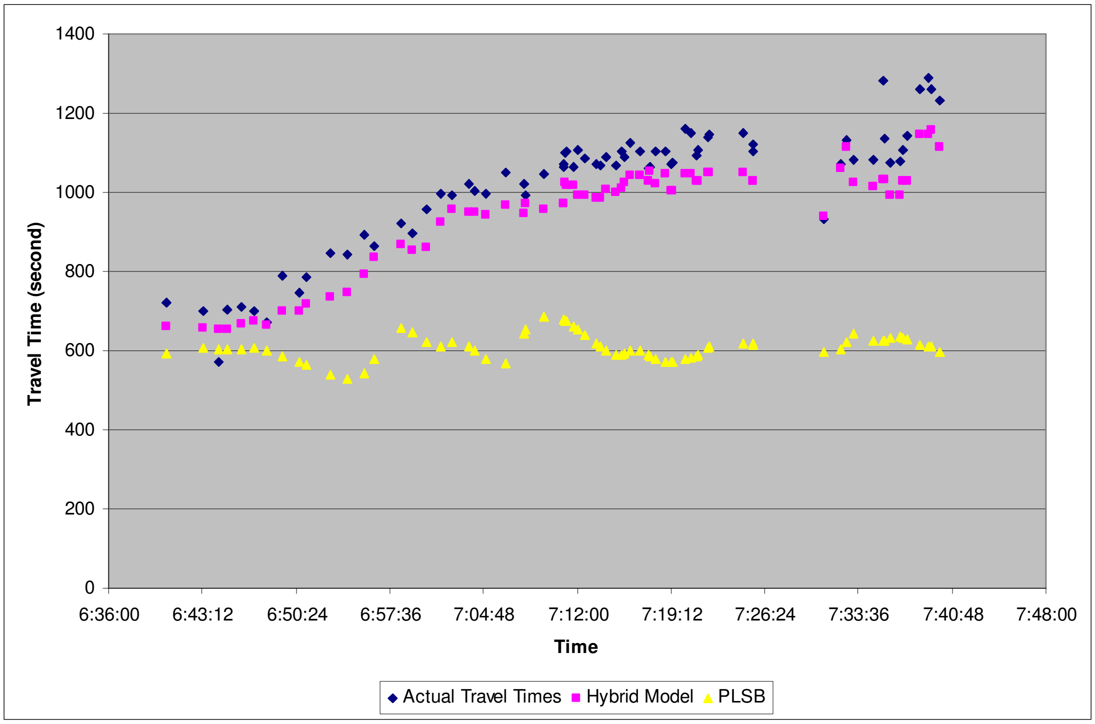
(b) Performance evaluation of travel time estimation module on the subsegment from Detector 3 to Detector 10 on April 20th, 2006

		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
Hybrid Model	AAE (sec)	65.2	49.4	73.0
	AARE (%)	8.7%	5.8%	7.8%
PLSB	AAE (sec)	153.5	243.7	335.2
	AARE (%)	19.7%	28.7%	36.0%

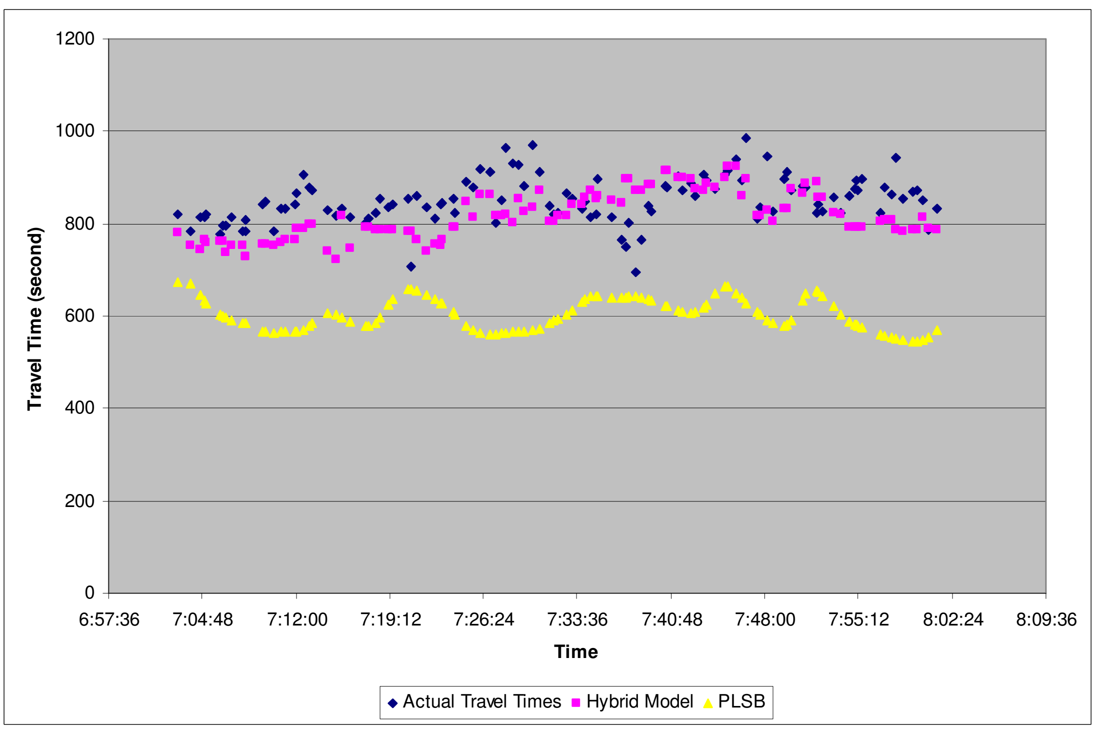
(c) Overall performance evaluation of travel time estimation module on the subsegment from Detector 3 to Detector 10 on April 6th and April 20th, 2006

		Travel Time Range (sec)		
		520 to 800	800 to 1000	> 1000
Sample Size		23	113	49
Maximum Travel Time (sec)		796	998	1290
Average Travel Time (sec)		742.3	847.2	1109.1
Hybrid Model	AAE (sec)	59.4	54.1	83.6
	AARE (%)	8.1%	6.2%	7.5%
PLSB	AAE (sec)	139.8	266.6	493.8
	AARE (%)	18.4%	30.5%	44.3%

AAE: Average absolute error.
AARE: Average absolute relative error.



(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 4-13 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 Conclusion

This chapter presented 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 CLR 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 ETB 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 PLSB models to provide reliable travel time estimations on long links.

An extensive comparison between the collected and estimated travel times clearly indicated that the developed hybrid model was 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 hybrid model may have yielded a relatively large error, but it remained within the range of one minute. The traditional PLSB model implemented for comparison failed to provide reliable estimations. Overall, the developed model 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 Long Detector Spacing

5.1 Introduction

Due to the 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 Traffic 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, reliability of available traffic detectors etc.

As discussed in Section 2.3, many studies have developed travel time prediction models for highway segments with simple geometric features and densely distributed traffic detectors (i.e., every half-mile). The large number of detectors required for those models have limited their potential applications, in view of diminishing resources for infrastructure development. This study intends to develop a travel time prediction model that can provide reliable travel time predictions under 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 congestion, this study proposes a hybrid model structure that employs a multi-topology Neural Network model with a rule-based clustering function for the situation when the historical traffic and travel time databases are not rich enough, and a *k*-Nearest Neighbor model for traffic scenarios that have a sufficient number of similar historical travel times. Both models have been developed to take full advantage of geometric features and historical traffic patterns in order to provide reliable travel time predictions using widely spaced detectors. In the multi-topology Neural Network model, the rule-based clustering mechanism categorizes traffic conditions into three scenarios — congestion in morning peak hours, congestion in evening peak hours and congestion-free periods — and then the model will apply a Neural Network model with the topology of either Multilayer Perceptron (MLP) or Time Delayed Neural Network (TDNN) for that specific traffic scenario to predict the travel time. 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 to switch the prediction model. During operation, this system can continuously update the historical travel time database with estimated travel times of most-recently completed trips and some critical parameters in both models. Figure 5.1 shows the flowchart of the proposed model for travel time prediction.

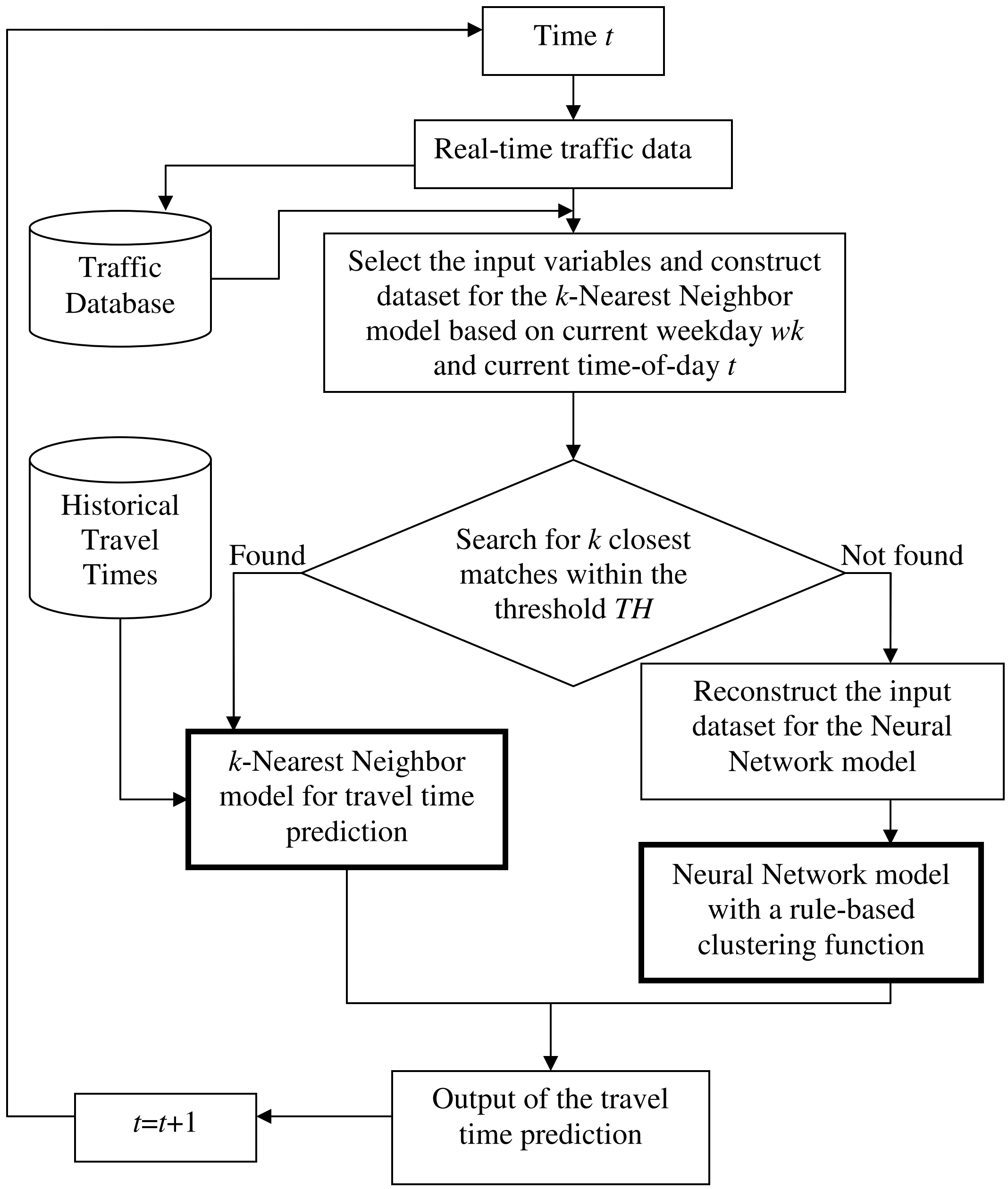


Figure 5-1 Flowchart of the hybrid travel time prediction model

The travel time prediction system will first construct the input dataset of the *k*-Nearest Neighbor model from the current real-time traffic data. If at least *k* historical cases exist within the similarity threshold, *TH*, from the current condition, then the hybrid model's output will be the prediction result of the *k*-Nearest Neighbor model, which is the average of those *k* best historical matches. Otherwise, the prediction system will reorganize the input data for the multi-topology Neural Network model with a rule-based clustering function and then output its prediction result. 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 Neural Network model with a rule-based clustering function and the *k*-Nearest Neighbor model respectively.

5.3 A Multi-topology Neural Network Model with a Rule-based Clustering Function

As reported in the literature, a single-topology Neural Network model is capable of providing reliable predictions in the transportation field under certain conditions (Dougherty, 1995). However, the prediction accuracy of Neural Network models for predicting traffic or travel times varies dramatically with the existence of nonrecurrent congestions and the congestion severity. Therefore, this study develops a multi-topology Neural Network model structure with a rule-based clustering function that can fully take advantage of the efficiency of different Neural Network topologies under various traffic patterns. The following section will describe the rule-based clustering function, the selection of input variables, and the topology of the Neural Network model used under each type of traffic scenario.

A rule-based clustering function

As is well recognized in the literature and observed in the real world, traffic conditions in a long segment may vary significantly during the morning peak hours, evening peak hours and off-peak hours due to the complex interactions of many factors with time-varying natures, such as demand patterns in each link, origin-destination distribution at ramps, drivers' responses to potential congestions etc. Some efforts reported in the literature have clustered the space of input variables and then trained the Neural Network model to obtain parameters in each cluster. A common approach to obtain the clustering criteria is to analyze the historical traffic patterns and determine the average time-of-day boundaries for morning and/or evening peak hours. However, such clustering functions do not really take into account the fact that congestion, rather than of time-of-day, is the most direct factor affecting the travel times.

This study develops a rule-based clustering function that categorizes traffic conditions on a long freeway segment with three commonly seen scenarios: congestion in morning peak hours, congestion in evening peak hours and congestion-free periods. The developed clustering function has the following features:

- The rule-based clustering function considers both daily traffic patterns and weekly traffic patterns to determine a preliminary boundary of peak hours for each weekday.
- The clustering function determines traffic scenarios based only on traffic conditions in critical lanes, which include both mainline lanes and ramp lanes, to reduce the disturbance from fluctuating traffic conditions in those lanes that do not directly contribute to any impact on travel times.

- In order to reduce the frequent scenario switching that occurs during transition periods, the clustering function requires traffic conditions to maintain stability for a period of time to confirm a scenario change.
- During real-time operation, the model can concurrently update some threshold parameters of the clustering functions — for example, the preliminary weekly boundaries of peak hours.

The logic of the developed rule-based clustering function for determining $p_d(t)$, the current traffic scenario at the current time t on link (d, d+1), is as follows:

IF $t \geq TML_d^{wk}$ and $t \leq TMU_d^{wk}$ THEN

IF $\exists la, o_{d^*,la}(t-j) > OM_{d,la}$ for all j, where $la \in \mathbf{CLM}_{d,d+1}^{d^*}$ and $0 \leq j \leq THNN$, THEN

$p_d(t) = 1$ (morning congestion)

ELSE

IF $o_{d^*,la}(t-j) \leq OM_{d,la}$ for all la and j, where $la \in \mathbf{CLM}_{d,d+1}^{d^*}$ and $0 \leq j \leq THNN$, THEN

$p_d(t) = 0$ (off-peak period)

ELSE

$p_d(t) = p_d(t-1)$

END IF

END IF

ELSE

IF $t \geq TEL_d^{wk}$ and $t \leq TEU_d^{wk}$ THEN

IF $\exists la, o_{d^*,la}(t-j) > OE_{d,la}$ for all *j*, where $la \in \mathbf{CLE}_{d,d+1}^{d^*}$ and $0 \le j \le THNN$, THEN

$p_d(t) = -1$ (evening congestion)

ELSE

IF $o_{d^*,la}(t-j) \le OE_{d,la}$ for all *la* and *j*, where $0 \le j \le THNN$, THEN

$p_d(t) = 0$ (off-peak period)

ELSE

$p_d(t) = p_d(t-1)$

END IF

END IF

ELSE

$p_d(t) = 0$ (off-peak period)

END IF

END IF

where, TML_d^{wk} and TMU_d^{wk} are the lower and upper time boundaries for morning peak hours in link (*d*, *d*+1) on weekday *wk* in the historical traffic patterns;

TEL_d^{wk} and TEU_d^{wk} are the lower and upper time boundaries for evening peak hours in link (*d*, *d*+1) on weekday *wk* in the historical traffic patterns;

$0:00 \le TML_d^{wk} < TMU_d^{wk} \le TEL_d^{wk} < TEU_d^{wk} < 24:00$;

$d^* = d$ or $d+1$;

$OM_{d,la}$ is the occupancy threshold at lane *la* at detector *d* in the morning;

$OE_{d,la}$ is the occupancy threshold at lane *la* at detector *d* in the evening;

$\mathbf{CLM}_{d,d+1}^{d^*}$ and $\mathbf{CLE}_{d,d+1}^{d^*}$ are sets of critical lanes at detector d^* in link (*d*, *d*+1) in the morning and in the evening respectively; and

THNN is the required duration for the traffic condition to maintain congested or uncongested stably;

In real-time operations, one can apply the above rules to each link in the target freeway segment to determine the associated traffic scenarios, $p_d(t)$ (*d*=1 to *D*-1), for the current time *t* and then determine which Neural Network model to use with $P(t)$ defined in Eq. 5.1.

$$P(t) = \sum_{d=1}^{D-1} p_d(t) \tag{5.1}$$

The model for congestion in morning peak hours will be used when $P(t) > 0$ and $p_d(t) \geq 0$ (*d*=1, 2, …, *D*-1); the model for congestion in evening peak hours will be used when $P(t) < 0$ and $p_d(t) \leq 0$ (*d*=1, 2, …, *D*-1); otherwise, the model for congestion-free periods will be used.

Note that the model assumes common traffic patterns with the existence of morning peak hours and evening peak hours in this study. However, it is possible for a segment to have more complicated congestion patterns that cannot be categorized as morning congestion or evening congestion. In such a case, one may define one congestion scenario for all congestion patterns and one congestion-free scenario, or further cluster the traffic conditions with site-specific congestion characteristics.

The selection of input variables

Neural network models have been widely used in transportation studies because they are proficient at recognizing patterns while being easy to apply without modeling the physical relations between input and output variables. It is reported in the literature that topology modification according to real-world relations between input variables may improve the performance of a Neural Network model (van Lint et al., 2002). Previous efforts at travel-time related studies with Neural Network models are mostly based on densely distributed detectors (i.e., less than 0.5-mile apart). In order to provide satisfactory reliability with long detector spacing, this study includes careful analysis of all available information and its relations.

In most intelligent transportation systems (ITS) for travel time prediction with traffic detectors, such as the framework presented in Chapter 3, the following information is commonly available:

- Geometric features of the entire segment;
- Current departure time;
- Traffic information detected at the current time, including volume count, occupancy and speed;
- Traffic information detected before the current time, including both short-term historic data (i.e., less than one hour before the current time) and long-term historic data (i.e., data from previous day); and

- Historic travel times, which include all trips that have completed before the current time.

Of the above information, geometric features are fixed, and therefore cannot be included as input variables to the Neural Network models directly. However, such information, along with historic traffic patterns, will be the basis for determining critical lanes, as discussed in Chapter 4. As mentioned previously, speed data does not serve as an input variable in any form due to the unreliability of speed measurements from most commonly-used roadside traffic detectors. This study develops the following general guidelines for determining the input variables to the Neural Network model in each cluster.

1. Include the current time of day as an input variable for congested clusters to capture the recurrent congestion patterns which have a time-varying nature.
2. In each cluster, identify all critical lanes in which traffic conditions are directly related to the congestion patterns.
3. Combine the same type of data in all through lanes into one variable at a detector location, if traffic conditions are uniformly distributed across all through lanes in one traffic scenario.
4. Include a number of consecutive intervals of traffic data in a period of time, *TP*, for each critical lane from the current time to the past as a set of time-series input variables that represent the "current conditions."
5. Contain both the historical average of travel times with departure times within a period of time, *TP*, before the same time of day and the historical

average of travel times that departed within a period of time, *TP*, after the current time of day as input variables.

To improve the model's ability to predict during transition periods, *TP*, an important parameter, must be no less than the average duration for traffic conditions to switch from a congestion-free to congested scenario or vice versa. The following shows how *TP* is determined in this study.

First, denoting *k* as the index of each occurrence in the historical database of a traffic scenario switch between congested and uncongested conditions, , one can obtain a set of historical time spots, $\mathbf{TK} = \{t_k\}, (k = 1, 2, \ldots K)$, in which each t_k satisfies that,

$$p_{d_k}(t_k) \neq p_{d_k}(t_k - 1)$$
$$p_{d_k}(t_k) \cdot p_{d_k}(t_k - 1) = 0 \tag{5.2}$$

where, d_k is the detector location at which the k^{th} switch of scenario occurred in the historical traffic database.

Then, define $ts_d(t)$ to describe the stability of the traffic.

$$ts_d(t) = \begin{cases} 0, & \begin{array}{l} \text{if } \exists \hat{t}_1, \hat{t}_2, la_1 \text{ and } la_2, \\ (o_{d,la_1}(\hat{t}_1) - OM_{d,la_1}) \times (o_{d,la_2}(\hat{t}_2) - OM_{d,la_2}) < 0, \\ t - THNN \leq \hat{t}_1, \hat{t}_2 \leq t, \\ TML_d^{wk} \leq t \leq TMU_d^{wk}, \\ \text{or} \\ \text{if } \exists \hat{t}_1, \hat{t}_2, la_1 \text{ and } la_2, \\ (o_{d,la_1}(\hat{t}_1) - OE_{d,la_1}) \times (o_{d,la_2}(\hat{t}_2) - OE_{d,la_2}) < 0, \\ t - THNN \leq \hat{t}_1, \hat{t}_2 \leq t, \\ TEL_d^{wk} \leq t \leq TEU_d^{wk}, \end{array} \\ 1, & \text{otherwise} \end{cases} \tag{5.3}$$

For each *k*, denote TP_k as the time for the traffic condition to switch from a stably congested scenario to a stably uncongested scenario or to switch the other way. TP_k satisfies that for any $t' \, (t_k - TP_k < t' < t_k - 1)$,

$$p_{d_k}(t') = p_{d_k}(t_k - 1) = p_{d_k}(t_k - TP_k)$$

$$ts_{d_k}(t_k - TP_k) \times ts_{d_k}(t_k) = 1$$

$$ts_{d_k}(t') = 0 \tag{5.4}$$

Then, *TP* can be determined as follows.

$$TP = AVG\{TP_k\} \tag{5.5}$$

<u>Topology of the Neural Networks used in this study</u>

The Multilayer Perceptron (MLP) has demonstrated that it can perform well at predicting travel times in the transportation field with densely placed detectors. Its flexibility about required input variables and its easy training procedures made MLP one of the most popular Neural Network topologies in the transportation-related literature. However, there are very limited efforts in the literature to optimize MLPs to provide acceptable levels of accuracy for freeway segments with widely spaced detectors. With a rule-based clustering function based on the careful analysis of historical weekly traffic patterns, this study develops one MLP for each clustered traffic scenario with a different set of input variables from critical lanes only to best utilize the good prediction ability of MLP for a freeway segment with sparsely distributed detectors.

Although a careful analysis can improve a MLP's performance by keeping only the information that directly interacts with the travel times, the model's inability to model time-series data limits its potential for better accuracy. As mentioned in 2.3.2, many researchers have applied a Time Delayed Neural Network (TDNN) for transportation prediction. The TDNN, which models time-series data with a tap-delay line, has a short-term memory unit attached to the input node to take advantage of the temporal relation in the input data stream. In the literature, most TDNN models for transportation forecast are for one stream of time-series data only — for example, volume in one lane at a detector location. In addition to MLP, this study develops an alternative Neural Network topology which accommodate both time-series and non-time-series data. First, the volume and/or occupancy data in each critical lane has been modeled as one input node with its own tap-delay line. Then, traditional input nodes are added to the topology for non-time-series data to the topology to form the complete Neural Network model. Figure 5-2 illustrates the topology of the enhanced Neural Network developed in this study. Capable of catching the trend of time-series data with its short-term memory elements, the developed alternative Neural Network model has potential to provide more reliable predictions during transition periods.

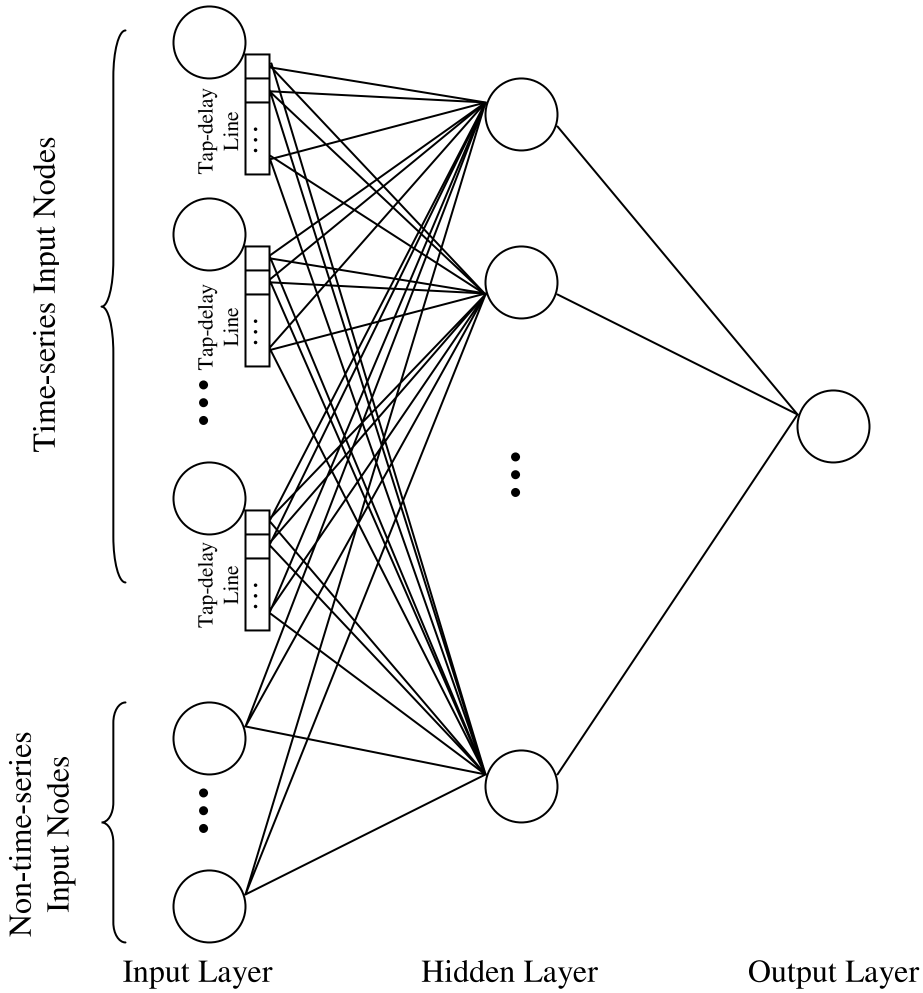


Figure 5-2 Topology of the enhanced Neural Network that combines both time-series and non-time-series inputs

5.4 *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 weighing 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, for travel time prediction, this definition 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 proposes the following sequence to compute the distance between the current and the historical case.

The proposed model first categorizes traffic conditions with detected occupancy information. One can then use the following equation to define three types of traffic conditions, free-flowing, heavily congested, and moderately congested.

$$TC_d^{l_a}(t, t+\Delta t) = \begin{cases} -1 & \text{, when } o_d^{l_a}(t, t+\Delta t) \le OF_d^{l_a} \\ 1 & \text{, when } o_d^{l_a}(t, t+\Delta t) \ge OC_d^{l_a} \\ 0 & \text{, otherwise} \end{cases} \quad (5.6)$$

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, respectively, for lane l_a at detector d.

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

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (p_i^* - q_i^*)^2} \quad (5.7)$$

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

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

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

<u>Selection of the input variables</u>

Most existing applications of *k*-Nearest Neighbor Models 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 to be disturbed by on-ramp or off-ramp flows. For example, a through lane next to an off-ramp lane may be avoided by most through traffic 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, such 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.

<u>Searching window and time range</u>

Both the 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 performing a reliable prediction with a longer horizon usually requires more historical and/or on-line data. To predict the travel time on one segment with multiple links, one needs to predict traffic conditions at a detector location which is closer to the departure point and with a shorter prediction horizon than those detectors that are farther away from the departure point. Therefore, the searching window of traffic information at each detector may increase with the increase in distance from the origin point. To ensure the computing efficiency, one must 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 resulting in different travel times. For example, it is possible for two cases with similar detected traffic conditions to have different travel times. Very often, a morning case and an evening case may have similar detected traffic flows, but they go to different destinations. Therefore, for better prediction accuracy, the searching procedure should only look for historical cases within a reasonable range from the current time of day. Hence, one can add this constraint to Eq. 5.7 to obtain the following equation.

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (\hat{p}_i - q_i^*)^2} \tag{5.8}$$

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 time-of-day traffic patterns at detector d. For example, $T_{th}(d,t)$ may be different in morning peak hours, evening peak hours and off-peak hours.

Besides the use of time-of-day information, this study further modifies Eq. 5.8 to improve the models' reliability by searching for cases that are in a weekday which usually has similar traffic patterns. Weekdays with similar traffic patterns are first grouped together into S sets. One can then modify Eq. 5.8 to obtain Eq. 5.9:

$$mdis = \sqrt{\sum_{i=1}^{k} w_i (\hat{p}_i - q_i^*)^2} \tag{5.9}$$

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

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

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

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

Weighting factors are used in the model to reflect the contribution of traffic conditions in each critical lane 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 into 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 of each variable and the time-varying searching range for each weekly traffic scenario by analyzing the historical traffic patterns.

5.5 Numerical Examples

To demonstrate the potential of the developed hybrid travel time prediction model in a real-world application with large detector spacing, this study includes numerical examples using the same dataset from 10 roadside detectors on a 25-mile stretch of I-70 introduced in Chapter 4 (See Figure 4-8). The target 10.06-mile subsegment selected for the numerical examples is located between an origin about 1.04 miles west of Detector 3 and the destination, I-695 (Figure 4-9j). A consulting company collected travel time data with probe vehicle method in both morning peak and evening peak hours for the 4 days from May 16th to May 19th, 2006. The headway of the data collection was between 4 to 6 minutes each day. These surveys recorded about 7 to 14 travel time samples each day and yielded a total of 98 actual travel times. Several accidents were observed by the data collector or recorded by the accident response team during these days, including during the evening peak hours on May 18th and May 19th. There were also several time periods with missing data during this 4-day period. This study removed the travel times under the accident impacts or the missing data and obtained a final dataset of actual travel times with 70 samples.

This section will first analyze the performance of the developed hybrid model for travel time prediction along with comparisons to some commonly-implemented models based on the estimated travel times in the database, which were used to train all the models. This is followed by an analysis of overall system performance using the collected actual travel times under recurrent traffic conditions.

Models for Comparison

The numerical examples include two commonly implemented models: the simple current-constant-speed-based (CCSB) prediction model found in most real-world systems as mentioned in Section 2.4.2, and the time-varying coefficient model

(TVC) (Zhang and Rice, 2003) implemented in two simulated systems (Chen et al., 2003; Rice and van Zwet, 2004) with Eq. 2.10.

The CCSB model assumes that the vehicle's in-link speed between each adjacent detector pair will be same as the average of currently detected upstream and downstream speeds. The travel time of one vehicle departing the origin at time *t* can be obtained by Eq. 5.10.

$$\tau_{d_1,d_2}(t) = \sum_{d=d_1}^{d_2-1} \frac{2L_d}{u_d(t) + u_{d+1}(t)} \tag{5.10}$$

where d_1 is the origin detector ID;

d_2 is the destination detector ID;

$\tau_{d_1,d_2}(t)$ is the travel time from Detector d_1 to Detector d_2 with the departure time *t*; and

L_d is the link distance between the detector pair (*d*, *d*+1), and

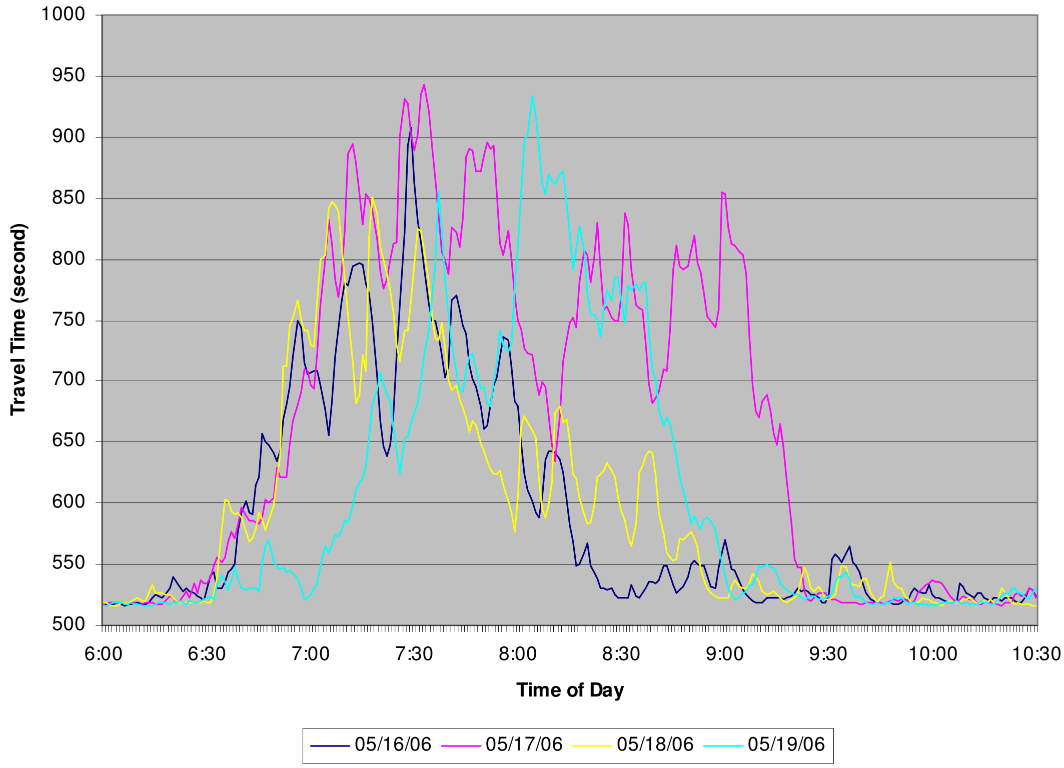
$u_d(t)$ is the speed detector at Detector *d* at time *t*

In addition to the developed hybrid travel time prediction model, its main prediction model, which is a multi-topology Neural Network model with a clustering function, and the supplemental model, an enhanced *k*-Nearest Neighbor model, have been implemented separately to explore their individual performances with various sizes of the historical database.

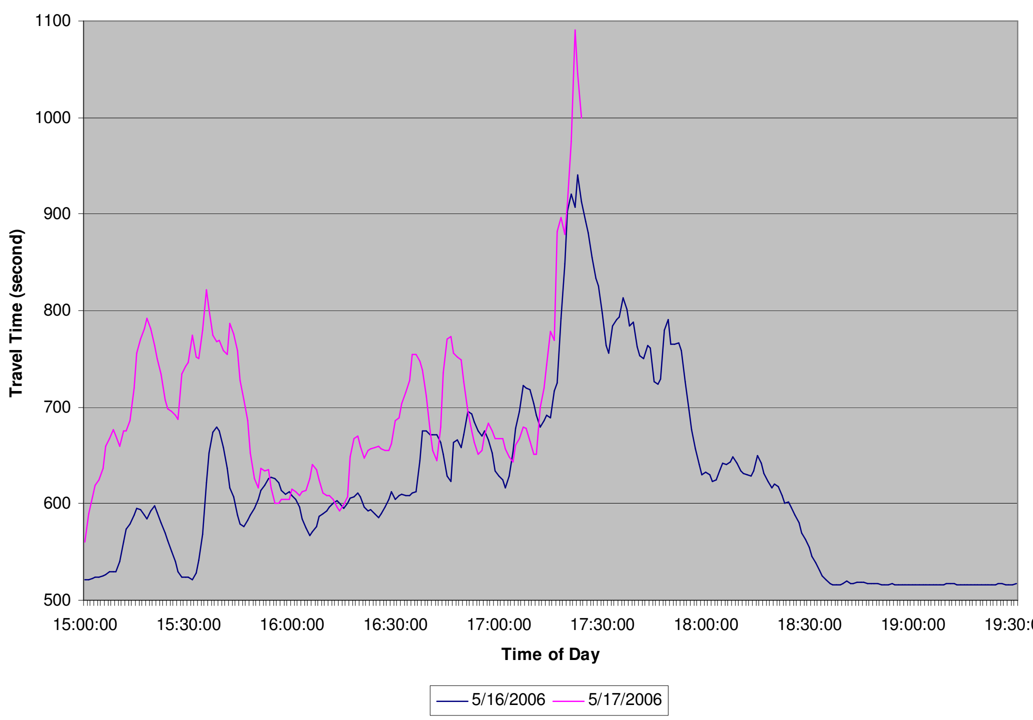
Performance for the Entire Week

Figures 5-3a and 5-3b show the distributions of recurrent travel times in morning peak hours and evening peak hours on the 4 different weekdays from May 16th to May 19th, 2006, excluding evening peak hours on May 18th and May 19th due to accidents and some periods due to missing data. The morning peak patterns have similar shapes on the sample days with, however, quite different starting and ending times. Generally, the congestion level is higher in the evening peak hours, with largest travel times more than double of those in the morning peak hours during the sample days.

The CSSB model is the only one that does not rely on any historical data. Other models all require some data to calibrate their parameters. In order to explore the impact of the size of the historical database, this study first includes complete data for a period of 4 weeks between April 7th and May 14th, 2006 to calibrate time-varying coefficient model, the *k*-Nearest Neighbor model, the Neural Network model and the hybrid model, here called TVC4, kNN4, NN4, HM4, respectively. Another set of models TVC10, kNN10, NN10, and HM10, are calibrated with a dataset of 10 weeks between February 9th and May 14th, 2006 (Table 5.1). Note that missing date periods were filtered out first from both datasets.



(a) Morning peak hours



(b) Evening peak hours

Figure 5-3 Distributions of travel times in time periods with no impact of accidents or missing data

Table 5-1 List of all model IDs

	4 Weeks of Training Data	10 Weeks of Training Data
Hybrid model developed in this study	HM4	HM10
Neural Network model in the developed hybrid model	NN4	NN10
k-Nearest Neighbors model in the developed hybrid model	kNN4	kNN10
Constant current speed-based model	CCSB	
Time-varying coefficient model	TVC4	TVC10

Among all the models calibrated with 4 weeks of data, HM4 and NN4 performed the best over the included periods on all 4 sample days. As shown in Table 5-2, NN4 has an average absolute error, as defined in Eq. 4.12, of 53.88 seconds, which is the best among all single models. HM4 provided better performance with the same 4-week historical database. TVC4 had an average absolute error of almost 3 minutes and an average absolute relative error (Eq. 4.12) of 28.10%.

Table 5-2 Performance of all models in all samples days

Model	Average Absolute Error (second)	Average Absolute Relative Error (%)
CCSB	77.92	10.89
TVC4	173.99	28.10
TVC10	65.64	9.44
kNN4	64.38	9.04
kNN10	60.86	8.56
NN4	53.88	7.81
NN10	48.68	7.07
HM4	48.84	6.92
HM10	45.69	6.53

With a 10-week historical dataset, NN10 was able to provide better accuracy than NN4, trained with a 4-week dataset, and benefited HM10, which was the best model, with an average absolute error of less than 46 seconds. Note that the average travel time for these samples days is 842.79 seconds (14.05 minutes) on this subsegment.

Table 5-3 Performances of all models in each included peak-hour period

Average Absolute Error (seconds)	5/16 AM	5/16 PM	5/17 AM	5/17 PM	5/18 AM	5/19 AM
CCSB	56.37	73.93	106.62	106.84	63.95	71.97
TVC4	186.04	127.27	232.82	128.47	166.45	168.96
TVC10	39.64	105.05	83.84	121.86	41.38	36.99
kNN4	34.42	84.09	81.48	126.45	34.10	58.85
kNN10	31.71	71.08	79.46	127.66	33.68	54.25
NN4	31.81	68.18	64.77	93.47	32.80	53.39
NN10	30.70	65.38	55.64	75.96	36.83	43.92
HM4	29.44	54.00	58.37	87.17	28.95	49.82
HM10	29.09	52.10	53.75	75.96	35.26	36.69

Table 5-3 details the comparison between all sample days in each included peak-hour period. The performances of the hybrid model are consistent across all periods. The TVC10 model was able to provide acceptable performance in one time period, however generated an average absolute error more than 80% larger than the hybrid model in several time periods. Note that the performance of the *k*-Nearest Neighbors model was not very stable due to its heavy reliance on similar historical scenarios. However, with its embedded ability to measure the potential errors, the hybrid model was still able to improve over the Neural Network model.

Table 5-4 shows a detailed comparison during free-flow traffic conditions (the lowest observed free flow travel time was about 520 seconds), with travel times less than or equal to 580 seconds, moderate congestion, with the travel times between 580 and 900 seconds (inclusive), and heavily congested traffic. In light traffic conditions, the *k*-Nearest Neighbor model had the best performance, mainly because of its enhanced searching function that takes special treatment for the light traffic. Under moderate congestion, the Neural Network model contributed the most to the hybrid model structure. This implies that Neural Network model was able to recognize the changes in the congestion pattern, which therefore gave it the best performance among all single models. Under heavy congestion, the NN10, HM10 and TVC10 were the best models, providing similar performances.

Table 5-4 Performances of all model in three congestion scenarios

	Model	Average Absolute Error (seconds)	Average Absolute Relative Error (%)
TT≤580 seconds	CCSB	13.45	2.52
	TVC4	161.56	30.70
	TVC10	12.65	2.37
	kNN4	7.48	1.39
	kNN10	7.40	1.38
	NN4	17.66	3.36
	NN10	16.14	3.06
	HM4	10.66	2.00
	HM10	10.05	1.89
580<TT≤900 seconds	CCSB	111.53	15.92
	TVC4	186.50	26.92
	TVC10	108.42	15.56
	kNN4	98.98	14.29
	kNN10	93.42	13.49
	NN4	75.22	10.79
	NN10	67.31	9.68
	HM4	69.81	10.00
	HM10	63.63	9.18
TT>900 seconds	CCSB	255.48	29.78
	TVC4	190.61	22.66
	TVC10	128.46	15.15
	kNN4	191.04	22.25
	kNN10	178.83	20.81
	NN4	141.44	16.43
	NN10	129.72	15.12
	HM4	138.73	16.10
	HM10	129.51	15.09

Note: TT = Travel Time

Overall, the developed hybrid model was able to provide acceptable performance with average absolute errors of about 10 seconds, 1 minute and 2 minutes under free-flow conditions, moderate congestion, and heavy congestion, respectively. The developed hybrid model had similar performances with both 4-week and 10-week historical travel times, which could help shorten the duration of the system training stage for system implementation.

Performance Comparison with Actual Travel Times

To demonstrate the potential of the developed travel time prediction system, this study compared all 70 collected actual travel times with the outputs from HM4 and HM10. Table 5-5 shows the performance comparison on all 70 samples and in each defined category. The developed hybrid model can provide acceptable accuracy with a 10-week training dataset. With a shorter duration of training data of 4-week, the developed hybrid still provided an average absolute error of less than 2 minutes in all observed traffic scenarios.

Table 5-5 Performance comparison of the travel time prediction system

	Average Travel Time (seconds)	HM4 AAE (seconds)	HM10 AAE (seconds)	Number of Samples
All samples	655.67	56.58	51.69	70
TT≤580	532.58	15.74	15.11	24
580<TT≤900	703.86	80.45	72.02	36
TT>900	949.67	113.43	95.29	10

Note: TT = Travel Time
AAE = Average Absolute Error

5.6 Conclusion

This study developed a hybrid travel time prediction model for reliable real-time travel time prediction. A multi-topology Neural Network model with a rule-based clustering function serves as the main model; this module fully takes into account the target roadway's geometry features and daily congestion patterns. The clustering function categorizes the traffic conditions using information obtained from critical lanes that affect the travel time the most, and then apply an appropriate Neural Network model. Based on the available historical travel times obtained from the travel time estimation module, the prediction module was able to switch to a *k*-Nearest Neighbor model for traffic scenarios that had a sufficient number of similar historical cases. The numerical examples showed that the hybrid model provided acceptable travel time results. The hybrid model can provide reasonably good performance with a 4-week historical dataset. With a 10-week dataset for calibration, the hybrid model managed to have an average absolute error of less than 130 seconds in the high congestion. The comparison against 70 actual travel time samples showed that the developed system can provide reliable travel time predictions on a road segment with complex geometric features and highly fluctuating traffic conditions.

Chapter 6: An Integrated Multiple Imputation Approach for Contending with Missing Data in the Travel Time Prediction

6.1 Introduction

As is well recognized, predicting travel times with sparsely distributed detectors is a very challenging task due to the complex interactions between many factors, such as large fluctuations of traffic conditions at detector locations and the large spacing between adjacent detectors. Chapters 4 and 5 presented several models to contend with these issues under normal operating conditions, where detectors can provide reliable traffic characteristic data. However, quite often in real-world deployment, detectors may not function as reliably as expected and may produce various types of missing data patterns, which may either degrade the accuracy of the predicted travel time or prevent the system from executing its functions due to an unacceptable level of reliability. Hence, it is essential for any real-time operational model to have an effective module for dealing with data missing scenarios.

A reliable module for estimating missing data needs to be able to (1) fully take into account of the site specific geometric features and traffic patterns to maximize its performance; (2) ensure the proper functioning of the travel time prediction model under various missing data scenarios, and (3) provide a reliability indicator so that the primary model can determine if the prediction function should cease under the detected data-missing scenario.

The next section will first introduce some missing data patterns revealed by the field demonstration study and their impacts on predicted travel times under existing models discussed in the literature. This is followed, in Section 6.3, by a comprehensive review of available missing data imputation methods in the literature. Section 6.4 and Section 6.5 will present two multiple imputation approaches developed for the travel time prediction model in Chapter 5 and apply them to a dataset collected from a field demonstration project. Section 6.6 will summarize the chapter.

6.2 Impact of Missing Data on Predicting Travel Times

In a real-time travel time prediction system, missing data can be categorized as short-term or long term; it is usually caused by data delay and/or data loss. A communication error often contributes to short-term missing data, whereas failure of a device, such as the traffic detector or the data storage device, often contributes to a long-term missing data. This section will illustrate some data-missing patterns and their possible impacts on the existing real-time travel time prediction systems

6.2.1 Common Missing Data Patterns

The dataset from 10 detectors illustrated hereafter was taken from the field demonstration project between February 9th and June 4th, 2006, which contain various commonly seen patterns of missing data.

Long-term Missing Data

On average, one detector has experienced a data-missing rate of more than 10% on 6.4 days. In total, for 39 out of 116-day field demonstration period (33.6% of the period of operation), at least one detector suffered a daily missing data rate exceeding 10%. Table 6-1 shows a detailed summary of the average data availability and the total missing data duration for each detector on days having a missing data rate of more than 10%. The low daily data availability implies that the long-term missing data pattern tends to be continuous throughout the entire day. Using a travel time prediction model that strictly requires the input dataset to contain no missing data during its real-time operations will result in a significant amount of time in which the system cannot function, when using most traffic detection systems.

Table 6-1 Average data availability and total data missing duration for each detector during those days when the missing data rate exceeded 10%

Detector	1	2	3	4	5	6	7	8	9	10
Number of Days	3	2	2	3	9	8	14	2	11	10
Average Daily Availability (%)	30.0	25.5	25.5	57.7	47.6	63.7	73.4	0.0	20.3	26.9
Total Data Loss Duration (Days)	2.1	1.5	1.5	1.3	4.7	2.9	3.7	2.0	8.8	7.3

Short-term Missing Data

During the period of operation, when traffic data was collected from 10 detectors over a 25-mile stretch of I-70, the system recorded the timestamps when each detector collected each available piece of data and timestramps when the data arrived at the database for all 3,268,287 data records collected between February 9th and June 4th, 2006. Table 6-2 shows the distribution of data delays, defined as the difference in the timestamps between when the detectors collect a traffic data item and when that data item enters the database. The table shows that most data records experienced a delay of less than 2 minutes. About 3% of the data were delayed longer than 30 minutes. This occurred mainly because the communication device on one detector failed, preventing the collected traffic data in the detector buffer memory from transmitting to the database until the failure was fixed.

Table 6-2 Distribution of the communication delays of all available data

Category (second)	Count	Percentage
0-30	203	0.01%
30-60	2867504	87.74%
60-90	248928	7.62%
90-120	16521	0.51%
120-150	8290	0.25%
150-180	7333	0.22%
180-210	7064	0.22%
210-240	3073	0.09%
240-300	2624	0.08%
300-360	1111	0.03%
360-420	696	0.02%
420-480	639	0.02%
480-540	606	0.02%
540-600	591	0.02%
600-1200	3436	0.11%
1200-1800	2490	0.08%
>1800	97178	2.97%

Figure 6-1 illustrates the distribution of timestamps of the traffic data versus the arrival timestamp in the database for data at Detector 5 collected between 14:21 and 15:42 on June 2nd, 2006. It clearly shows that the order of data recovery was "first missing first recovered" for all 3 short periods of data loss.

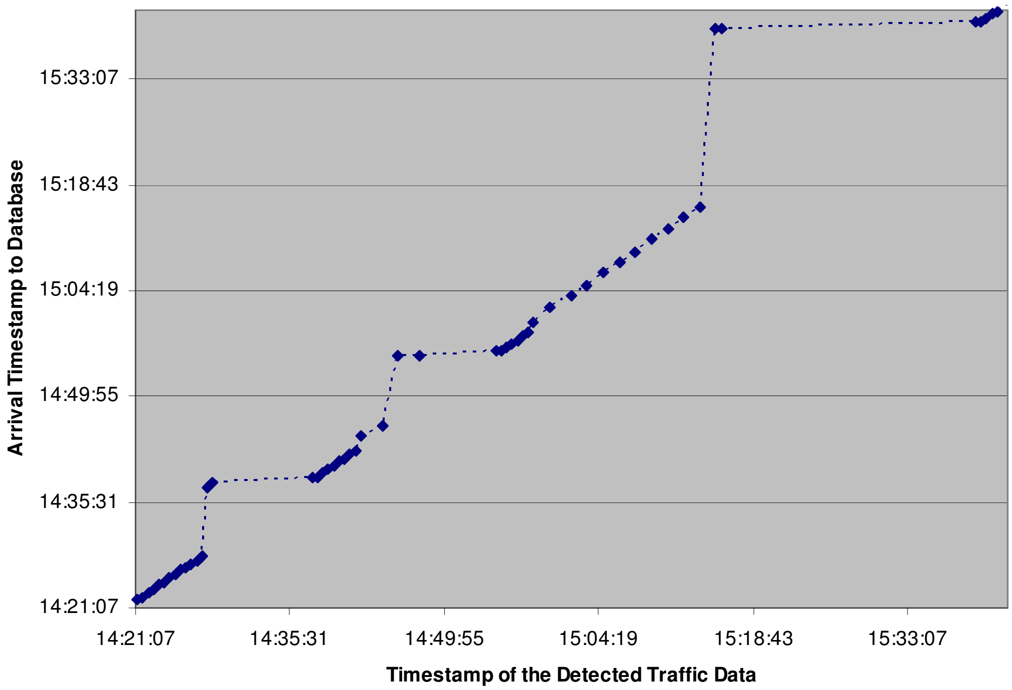


Figure 6-1 Timestamp of the detected traffic data vs. timestamp of data arrival at the database

6.2.2 Impacts of the Missing Data on Travel Time Predictions

Table 6-3 shows the example impacts of the missing data on the travel time predictions when Detector 10 experienced missing data rates of 20% and 100%. The target segment was from Detector 2 to Detector 10 with a free flow travel time of 694 seconds. The example shows the distribution of prediction errors for the travel time prediction model developed in Chapter 5 with the missing data being imputed using two methods: mean substitution (MS) and multiple imputation (MI). It is notable that for a 20% missing data rate, both imputation methods can estimate the missing values, allowing the travel time prediction model to maintain reasonable reliability. However, the average prediction error increased drastically when the missing data rate reached 100% on one detector. Although multiple imputation method can outperform the mean substitution method, the prediction system still experienced errors of more than 50% in some intervals, compared with less than 12% when no data was missing.

Missing Rate	20%		100%	
	MS	MI	MS	MI
16:50:00	5.37%	1.74%	22.06%	17.27%
16:51:00	7.40%	3.77%	48.91%	12.19%
16:52:00	7.06%	7.06%	63.83%	22.77%
16:53:00	9.27%	9.27%	27.81%	16.85%
16:54:00	11.59%	11.59%	26.12%	22.91%
16:55:00	15.69%	15.69%	42.02%	21.15%
16:56:00	15.85%	15.85%	19.78%	28.47%
16:57:00	18.56%	18.56%	24.36%	35.55%
16:58:00	19.60%	19.60%	79.97%	68.59%
16:59:00	20.87%	22.75%	102.46%	73.19%
17:00:00	22.93%	22.93%	135.27%	73.73%
17:01:00	20.12%	20.12%	102.32%	64.69%
17:02:00	16.98%	16.98%	25.39%	29.81%
17:03:00	13.92%	13.92%	31.79%	25.32%
17:04:00	9.10%	9.10%	21.47%	16.58%
17:05:00	3.49%	3.49%	5.81%	6.33%
17:06:00	1.77%	1.64%	3.91%	10.35%
17:07:00	0.86%	0.86%	15.04%	2.47%
17:08:00	11.33%	5.14%	19.51%	19.51%
17:09:00	9.43%	8.61%	19.79%	19.79%
17:10:00	9.62%	9.62%	20.16%	17.50%

MS: mean substitution.
MI: multiple imputation.

Table 6-3 Absolute relative errors of travel time predictions at missing data rate of 20% and 100% at Detector 10 between 16:50 and 17:10 on June 20th, 2006

6.3 Literature Review

Over the past few decades, researchers in different technical fields — including econometrics, social sciences, biostatistics and transportation — have devoted significant effort to solving the missing data issue. Some early studies contending with the missing data simply employed primitive approaches, such as case deletion and mean substitution (Little and Rubin, 1987; Schafer and Grapham, 2002). Since the 1970s (Rubin, 1976; Dempster et al, 1977; Little and Rubin, 1987; Rubin, 1987), more researchers have recognized the complexity of the missing data's nature and its impacts on the resulting performance of any model employed. Effective methods for dealing with the missing data issue may also vary with its pattern and the target applications (Little and Rubin, 1987).

This study has categorized existing methods for handling the missing data into three groups: data discard, single imputation, and multiple imputation models. Most recent studies in the literature indicated that multiple imputation approaches can outperform the single-imputation methods that are widely used due to their implementation convenience. A brief description of key studies in those three categories is presented in sequence below:

Data Discard

In practice, the method of case deletion, which discards the data unit with missing values, is the most popular data discard method and remains the default method for dealing with ignorable nonresponses in many statistical software packages (Shafer and Graham, 2002). Case deletion is generally valid only for the missingness mechanism of missing completely at random (MCAR) (Little and Rubin, 1987), in which the missing values are related neither to observed dependent values nor to independent variables. Graham and Donaldson (1993) found that case deletion may be valid and efficient in some scenarios of missing at random (MAR), in which the missing values are related only to the independent variables.

With the weighting factors estimated from the computed probabilities of nonmissing data, the use of reweighting may improve the performance of the case deletion method by eliminating the potential bias due to the differential responses that arise when modeling the response probability. However, this method cannot correct the bias related to the unused or unmeasured variables (Little and Rubin, 1987).

Single Imputation Methods

Some researchers have made great efforts to develope techniques that can estimate missing data based on partially available information so that the traditional statistical approaches can still be applied. Their proposed single imputation (SI) methods mainly impute the missing data from the means and distributions of the observable dataset.

As reported by many studies (Little and Rubin, 1987; Schafer and Graham, 2002), an unconditional mean substitution (such as using the mean of all available values in place of the missing value) may be effective for the types of study focusing on the mean of the data (Little and Rubin, 1987). However, it may underestimate the variances and distort covariance and intercorelations between variables. Schafer and Schenker (2000) presented an improved method that imputes the data from predictive means.

For studies that need both the means and the distributions of the data, some researchers have developed a set of different methods (Madow et al, 1983). One of the most widely used methods is the hot-deck method, which replaces the missing value with a value of the same variable estimated from one or several matching complete data records using certain searching criteria (e.g., the similarity).

Another commonly-used single imputation method that focuses on both the mean and the variance is the expectation-maximization (EM) algorithm (Dempster et al, 1977; Little and Rubin, 1987). EM is an iterative estimation method for missing data. It draws a random value from the probability distribution and iteratively add the new draw to the distribution until the probability distribution converges.

In summary, the core logic of single imputation methods is to find a way to estimate the missing value from the available data. However, such methods do not measure the imputed data quality, which varies with the nature of the data and other factors associated with the original dataset.

Multiple Imputation Methods

In view of the deficiencies of the single imputation methods, some researchers have developed the Multiple Imputation (MI) techniques to improve the imputation quality by incorporating the uncertainty of the missing data (Rubin, 1987; Little and Rubin, 1987). The common logic of multiple imputation methods is to estimate the same missing value *m* times ($m > 1$) with a simulated process (e.g., a Markov chain Monte Carlo [MCMC] simulation) to generate *m* complete datasets, and then analyzes the mean and variance of the estimators in these datasets to produce the final model output. MI has been widely applied to the social sciences (Saunders et al, 2006), biostatistics (Souverien et al, 2005; Gan et al, 2006), and transportation. Reportedly, the MI methods generally outperform the single imputation methods in most MAR cases. Another advantage of the MI methods is their ability to estimate the variance of the final model output for analysis.

Applications of the Data Imputation Methods in the Transportation

The use of missing data techniques has received an increasing amount of attention in the field of transportation since the turn of the century. However, several missing data treatment methods have already been used by practitioners and researchers in transportation applications. These methods include conditional mean substitution, regression models (such as interpolation), and time-series models (Nguyen, 2003). These transportation studies have focused mainly on replacing the missing values in the detected traffic variables (flow, occupancy and/or speed) with imputed values so as to construct a complete set of traffic data (Haj-Salem and Lebacque, 2002; Chen et al, 2003; Smith et al, 2003; Nguyen, 2003; Zhong et al, 2004, Al-Deek and Chandra, 2004, Al-Deek et al, 2004; Kwon, 2004; and Ni et al, 2005). Most studies applied single imputation techniques (such as EM); and only a few has employed the multiple imputation methods. (Ni et al, 2005). Some research also incorporated advanced prediction models, such as ARIMA, local weighted regression, and Neural Network models — for missing data estimation in order to capture the temporal and spatial distributions of the detector data. (Zhong et al, 2004; Al-Deek and Chandra, 2004). However, as reported in the literature on missing data theories (Little and Rubin, 1987; Schafer and Graham, 2002), the best imputation model may vary with the type of the model used for the application.

6.4 Model Structure

Grounded on the existing theories for missing data estimation, this study proposes two imputation models, named M-1 and M-2, to supplement the travel time prediction model in a sparsely-distributed detection environment based on the logic of the multiple imputation methods. Model M-1 integrates the missing data estimation with the travel time prediction to achieve a better overall prediction performance when data is missing. Model M-2 focuses on restoring the missing data used by the prediction models developed in Chapter 5. To facilitate the selection of the most effective imputation method, this study first analyzes the patterns of missing data from the field demonstration project.

6.4.1 Patterns of Missing Data

In the operating travel time prediction system, it takes the following four steps for the detected data to reach the traffic database: 1) collection by the traffic detector; 2) temporary storage in memory at the site; 3) transfer from the temporary storage to the data acquisition server; and 4) storing into the traffic database server. Of these four steps, most failures taken place at Steps 1, 2 and 4 are randomly distributed throughout the day of operations. The communication delays and failures of the communications may depend on available network bandwidth.

Figure 6-2 shows the average daily distribution of data delays from the dataset collected on I-70 in February, March, April and May of 2006. A data delay is defined as an interval of more than 3 minutes from the time data is detected to the time it enters the traffic database. The figure clearly shows that the data delays occur much more frequently in the daytime than at night, except in April. The delay patterns exhibited a high morning peak in March and May 2006 and a high evening peak in March 2006. Hence, one can categorize the data-missing pattern of these detectors as missing at random (MAR).

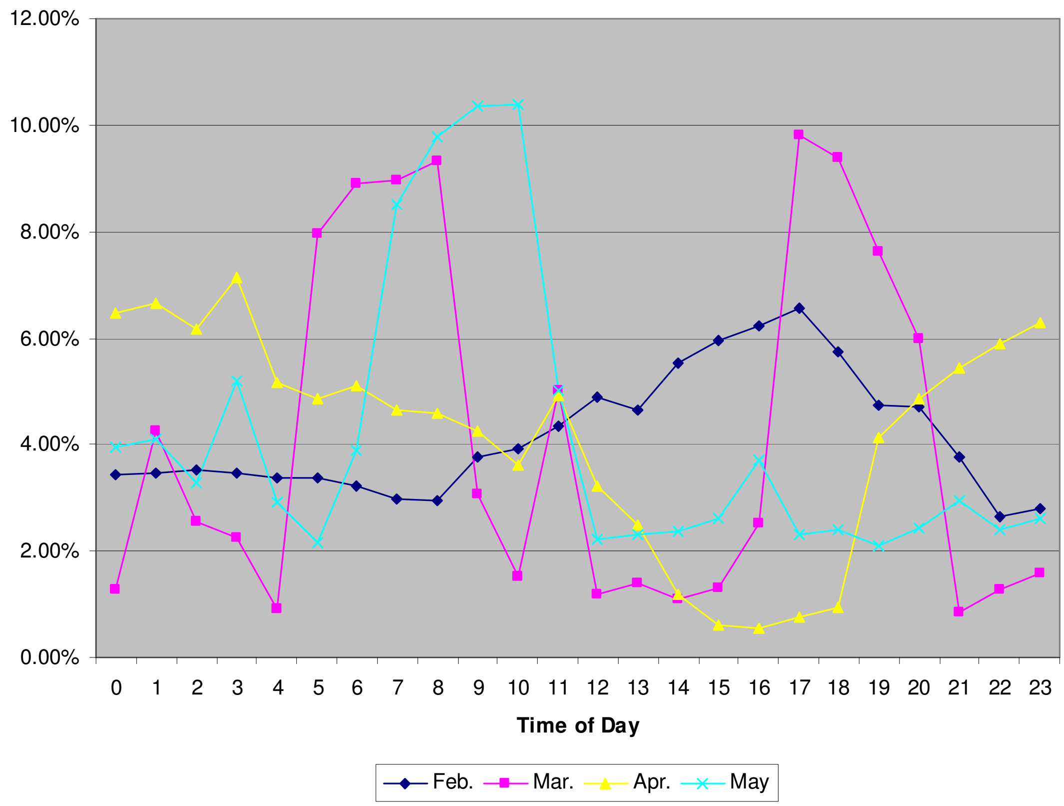


Figure 6-2 Distribution of data delay patterns over time in February, March, April and May 2006

6.4.2 Model Flowchart

Although traffic data may exhibit a pattern of high fluctuation, researchers are more interested in its trend rather than its variance in most transportation studies. Figure 6-3 illustrates the framework of the proposed system, which combines two missing data imputation models: Model M-1 and Model M-2. When the system detects that data is missing from the input dataset of the travel time prediction module during real-time operations, it will first apply Model M-1, an integrated multiple imputation and travel time prediction model, to perform the prediction despite the missing data to obtain $TT_{M1}(t)$. If the variance of the imputed result from Model M-1 is larger than the time-dependent threshold $TH_{M1}(t)$, the system will then switch to Model M-2 to impute the missing values (traffic flow and/or occupancy by lane) in the input dataset only. The system will apply the prediction models developed in Chapter 5 if the imputed values are reliable when compared with the time- and location-dependent flow threshold, $TH^{v}_{M2}(d,la,t)$, and/or the occupancy threshold, $TH^{o}_{M2}(d,la,t)$. The travel time prediction system will not display its predicted results if the imputed missing data has been detected to exceed the acceptable range.

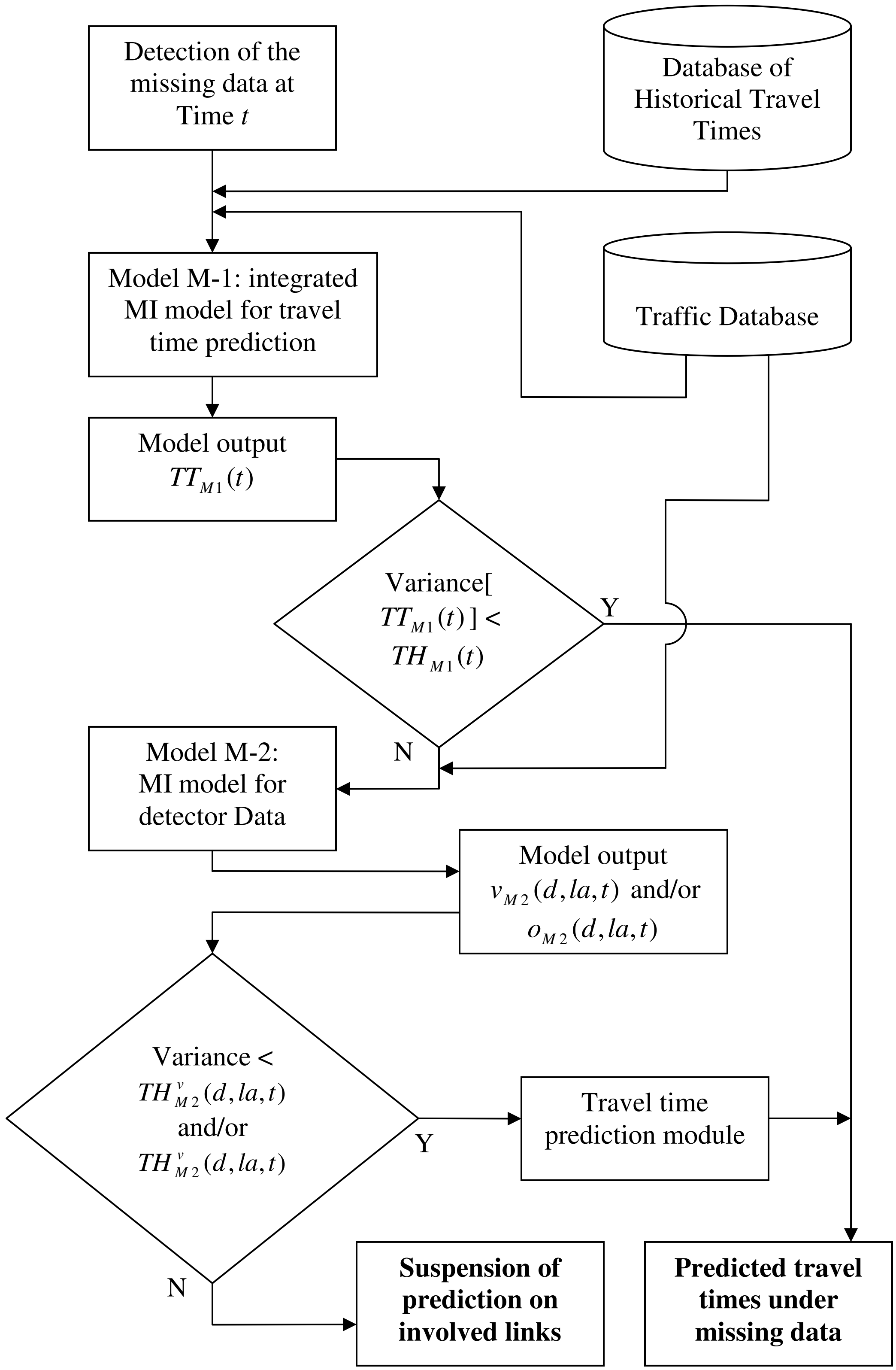


Figure 6-3 Flowchart of the missing data estimation Module

6.4.3 Model M-1: An Integrated Model for Travel Time Prediction under Missing Data

As reported in the literature, the performance of multiple imputation methods largely depends on the target applications. Hence, this study has developed an integrated model for travel time prediction using the incomplete dataset by taking advantage of the historical travel time information, geometry information and traffic patterns. A step-by-step description of the procedures for implementing the proposed integrated Model M-1 is presented below:

Step 1: Construct a dataset with the information from critical lanes that do not encounter missing data at current time t. The dataset shall also include data prior to the current time t from each critical lane. The guidelines for determining the time window can be found in Section 5.4.

Step 2: Search for h complete historical cases that have similar traffic conditions to conditions in the critical lanes at the current time. One can use the same search algorithm as the one used for the enhanced k -Nearest Neighbors model in Chapter 5. Note that the value of h needs to be sufficiently large to reliably estimate the distributions of the missing values. If historical cases are not adequate, the model will report that no reliable prediction can be made under the current missing patterns.

Step 3: Set imputation index i=1.

Step 4: Construct a set VAR_{COM} with variables in critical lanes in all h complete historical data records.

Step 5: Determine the probability distribution of missing variables, given the available complete data records $p(VAR_{MIS} \mid VAR_{COM})$.

Step 6: Impute all missing values and the travel time prediction based on $p(VAR_{MIS} \mid VAR_{COM})$.

Step 7: Integrate the newly obtained values from Step 6 with VAR_{COM} to form VAR'_{COM}.

Step 8: Test whether the i^{th} imputation converges based on the differences in both the mean and the variance between $p(VAR_{MIS} \mid VAR_{COM})$ and $p(VAR_{MIS} \mid VAR'_{COM})$. If it converges, then go to Step 9. Otherwise, let $VAR_{COM} = VAR'_{COM}$, and then go to Step 6.

Step 9: Record the imputation results, then let i=i+1. If $i \leq m$, go to Step 5.

Step 10: Determine the mean and variance of each variable in the m imputed data records. If all the variances are less than the assigned thresholds, then Model M-1 will output the average value of m imputed travel times as a reliable prediction despite the current missing data impact. Otherwise, the model will inform the system that no reliable result can be produced.

Note that one can execute the above procedures m times to generate a set of m imputed values. Prior to implementing the model, it is essential to determine four important parameters: the number of imputations m, the number of similar historical cases h, the criteria to determine the convergence of each imputation, and location- and time-dependent thresholds of the variances of missing values.

As reported in the literature (Little and Rubin, 1987), the efficiency of multiple imputation techniques can be assessed using Eq. 6.1. Therefore, Rubin (1987) suggested that m can lie between 3 and 10. However, due to the highly fluctuating nature of traffic data, one may need to perform an extensive sensitivity analysis to determine the optimal value for m in transportation applications.

$$E_{MI} = (1 + \frac{\gamma}{m})^{-1} \tag{6.1}$$

where E_{MI} is the efficiency of the multiple imputation method;

γ is the missing rate; and

m is the number of imputations.

Although the use of only a small number of similar historical cases h may help the system shorten its data collection and training stage, it might result in an unreliable estimation of the distributions of the missing variables. Therefore, this study requires the system to have identified at least 20 similar historical cases if Model M-1 to be used to impute the missing data. The system may take about 4 weeks to collect enough similar traffic conditions for recurrent congestion periods.

The convergence criteria and thresholds are available from the literature (Little and Rubin, 1987; Rubin, 1987). The former are usually defined as penalty terms that equal the covariance of two imputed values given the nonmissing data and the estimated covariance matrix.

6.4.4 Model M-2: Multiple Imputation of the Missing Detector Data

In addition to the integrated multiple imputation model, this study also develops a traditional multiple imputation model that imputes the missing values only for the scenarios in which Model M-1 cannot provide a reliable travel time prediction. Under a similar framework, this model groups related variables into one set of search indicators to generate imputations from similar cases. In the development of Model M-1 described in the previous section, variables in all critical lanes that may contribute to the predicted travel time are included in the set of search indicators. However, Model M-2 only takes into account traffic patterns in critical lanes in the same identified subsegment as the detector experiencing the missing data.

Most existing models in the literature employed parametric models to estimate the missing values, including temporal relations of values at the missing location prior to the current time and those spatial relations with other lanes at the same detector station and at neighboring detectors. However, due to the variation of traffic conditions across lanes at the same detector, data available from the neighboring lanes at the same detector will only be used in the search for similar cases.

To apply the proposed Model M-2, one needs to divide the target freeway segment into several subsegments. The dividing criteria presented below shall be time-dependent so as to fit the characteristics of daily traffic patterns (e.g., various dividing criteria for morning peak hours, evening peak hours and non-peak hours).

<u>Step 1:</u> Identify traffic scenarios based on the recurrent congestion patterns, and then perform the following steps for each traffic scenario.

Step 2: Group adjacent detectors into one subsegment if there are no ramps between the detectors.

Step 3: Combine adjacent subsegments if the detector at the interface point has a very low volume in the current traffic scenario and all ramps in the newly combined subsegment are covered by detector stations.

Step 4: Repeat Step 3 until no further combination is possible.

With the predefined subsegments for the current traffic scenario, one can further apply the following step-by-step procedures to estimate the missing values:

Step 1: Divide missing values into groups based on their locations in the predefined subsegments for the current traffic scenario.

Step 2: Search for *h* similar historical cases with complete data in the subsegment. If historical cases are not adequate, the model will report that no reliable prediction can be made for this group of missing data.

Step 3: Set the imputation index i=1.

Step 4: Construct VAR_{COM} with variables in the critical lanes within the subsegment from those *h* historical cases.

Step 5: Go through the same Steps 5 to 10 as Model M-1 to generate the final imputation results for the current subsegment.

Step 6: Repeat Steps 2 to Step 5 for all subsegments that experience missing data.

The system will then place the imputation results into the missing detector data at the current time *t* to construct a complete input dataset for use by the travel time prediction model. By taking into account the geometric features and traffic congestion patterns, Model M-2 can supplement Model M-1 when a direct estimate of travel time is not available.

6.4.5 Properties and Advantages of the Developed MI Models

The missing data issues in transportation-related applications require customized solutions due to the unique characteristics of the traffic data. In order to produce more accurate and robust estimation results, the two proposed multiple imputation models take into account both temporal and spatial relations of the traffic data and fit them into a multiple imputation framework, along with other factors, such as geometry impacts and congestion patterns. The essential logic of the multiple imputation technique is to treat the parameters as random variables rather than as fixed values (Rubin, 1987). The MI method first estimates the posterior distribution of the variables to be imputed. Denoting $Q = Q(X, Y)$ as the quality of the imputation, where X is the set of complete variables and Y contains the variables with missing data, the posterior distribution of Y_{mis} can be determined by Eq. 6.2 (Rubin, 1987):

$$\Pr(Q \mid X, Y_{obs}, R) \tag{6.2}$$

where Y_{mis} is missing values;

R is a $N \times p$ matrix with binary values indicating missing of Y; and

$Y = (Y_{obs}, Y_{mis})$

Rubin (1987) showed that Eq. 6.3 and Eq. 6.4 can properly estimate the mean, $\hat{Q}$, and the variance, U, of the posterior distribution of completed datay. Therefore, the simulation procedure incorporated in the multiple imputation framework is valid to estimate the posterior mean and variance of the missing values.

$$\hat{Q} = E(Q \mid X, Y_{obs}, Y_{mis}, R) \tag{6.3}$$

$$U = V(Q \mid X, Y_{obs}, Y_{mis}, R) \tag{6.4}$$

An important issue for multiple imputation is the proper estimation of the posterior distribution of the missing variables from their observed values. The solution to this issue varies with the application, due to the varied nature of data and the interactions of different factors. To fit the characteristics of a travel time prediction model, this study proposed searching mechanisms that consider geometric features of the roadway segment, historical traffic patterns and the temporal trends of all variables, which can reliably estimate the posterior distribution under most recurrent congestion.

The following section will discuss the properties and the reliability of the proposed multiple imputation models and compare them to commonly used single imputation approaches in three scenarios: low missing rate, high missing rate with stable traffic conditions and high missing rate with unstable traffic conditions in sequence.

Scenario 1: A Low Missing Rate

Several studies (Little and Rubin, 1987; Shafer and Schenker, 1999) have reported that many single imputation methods, for example mean substitution, can work well in some applications when the missing rate is low (i.e., less than 5%). Some single imputation approaches do not require rich historical data and, hence, can be implemented very quickly. On the other hand, the proposed multiple imputation models require a number of similar historical cases in the same cluster determined by the searching mechanism and therefore need a relatively rich historical database. Note that both two multiple imputation models developed in this study can share the historical traffic database with the travel time prediction module. As mentioned in Chapter 3, the developed travel time prediction system has a model training stage to collect traffic data and calibrate its model parameters. Therefore, one can rely on the data collected during this period being available for estimating the posterior distribution of the missing variable by the time the system is ready to operate. As reported in the literature and demonstrated in the numerical examples in this study, multiple imputation models have similar performance to single imputation methods under a low missing rate (e.g., less than 5%).

Scenario 2: A High Missing Rate with Stable Traffic Conditions

With a high missing rate, both proposed multiple imputation models can still estimate the distribution of the missing variable using information obtained from available temporal and historical data of the same variable and/or data from other critical lanes over the entire segment (Model M-1) or the determined subsegment (Model M-2).

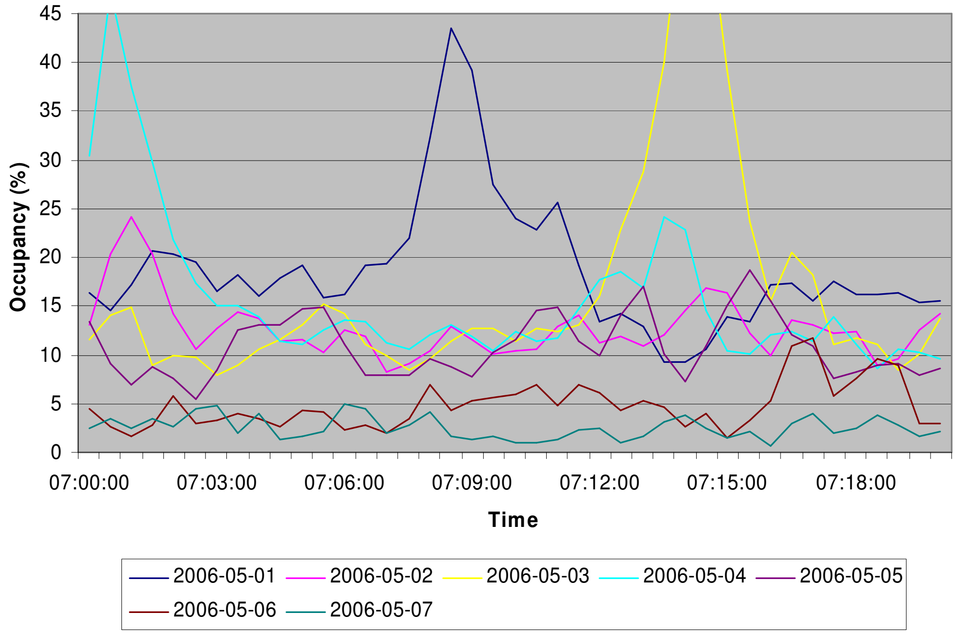
When the current traffic conditions are stable (e.g., free-flow traffic condition), with all similar historical cases showing no potential condition change at the missing variables (e.g., in late evening), the small variance of the posterior distribution of the missing values will not cause large errors under recurrent traffic patterns. Under a similar scenario, a common single imputation method, such as mean substitution or time series forecasting, can still provide acceptable results because the variability of the variable is low. However, common single imputation models cannot account for a possible sudden change in the traffic conditions, which may occur during the data-missing period. Schafer and Gahram (2002) approximated the coverage probability after mean substitution with Eq. 6.5. With the missing rate raising from 30% to 70%, the coverage probability decreased from 89.5% to 18.9%, which caused a 2 to 18 times increase in the possible error rate over the case with no missing data.

$$2\Phi[1.96(1-r)] \tag{6.5}$$

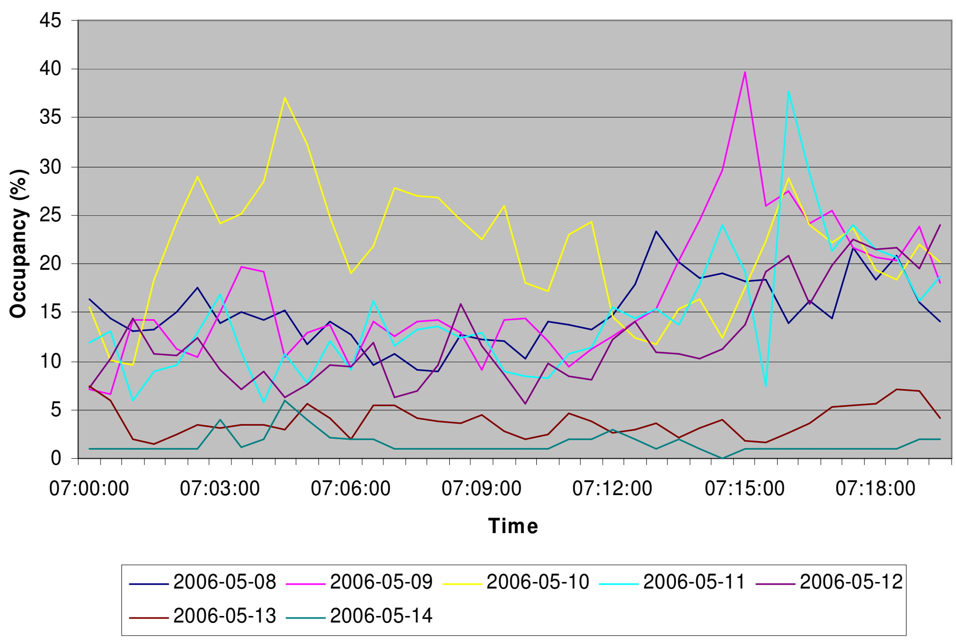
Where Φ is the standard normal cumulative distribution function, and

r is the missing rate.

Figure 6-4 shows an example of traffic in Lane 1 on Detector 5 between 7:00 and 7:20 AM, which is a potential transition period, in the first 2 weeks of May 2006. The figures show that traffic may stably maintain a light condition, incur a sudden change from uncongested to heavy congestion, or fluctuate between the two conditions. Without a carefully designed model to account for these possible situations, imputing each missing value with a commonly used single imputation method may significantly underestimate the data's variability.



(a)



(b)

Figure 6-4 Occupancy distributions in Lane 1 at Detector 5 between (a) May 1st and May 7th, and (b) May 8th and May 14th, 2006

Scenario 3: High Missing Rate with Unstable Traffic Conditions

Estimating missing data under unstable traffic conditions is a hard task due to the complex interactions between many factors which cause travel times to fluctuate widely. The proposed multiple imputation models first categorize similar traffic conditions with a customized clustering function, designed for the travel time prediction application, that takes of geometry features, traffic patterns and other factors into consideration, and then imputes the missing values multiple times according to the distribution determined by similar historical cases. With its integration of the travel time prediction, the proposed Model M-1 has its unique ability to of estimate the reliability of the predicted travel time under the impact of missing data. This estimate of output reliability is an essential function for a real-time travel time prediction system, because it prevents the system from displaying unreliable travel times. None of the commonly used single imputation methods take this issue into account, as most of them focus only on restoring the dataset, without analyzing potential errors for a specific application.

6.5 Numerical Examples

This section presents some numerical results from using these two proposed missing data imputation models. The numerical examples were based on the same dataset this study has been using, collected from 10 roadside detectors on a 25-mile stretch of I-70 eastbound between MD27 and I-695. The illustrative example is for the subsegment between Detector 2 and Detector 10, which is about 13.81 miles in distance with a free flow travel time of 694 seconds. The evaluation periods were between 15:00 and 19:00 on 4 of 5 consecutive weekdays from June 20th, 2006

(Tuesday) to June 26th, 2006 (Monday), excluding June 23rd, 2006 (Friday). The numerical examples intend to highlight the following issues:

- The missing rate;
- The type of imputation model used; and
- The number of multiple imputations executed.

Figure 6-4 shows the distribution of the estimated travel times between 15:00 to 19:00 on these four weekdays, which have different starting times for their peak hours but approximately the same ending times. The estimated travel times will serve as the true values for the performance evaluation of each missing data imputation model.

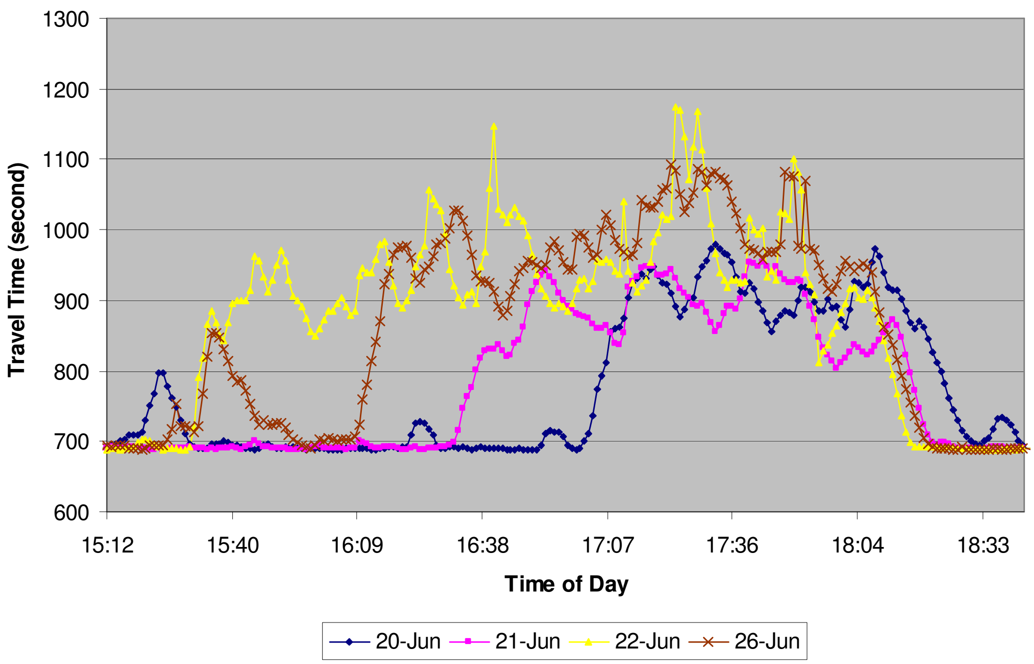


Figure 6-5 Distributions of Travel Times between 15:00PM and 19:00PM on 4 days in 2006

The numerical examples compare the performances of five types of models for missing data estimation: mean substitution (MS), Bayesian forecast (BF), multiple imputation of missing detector data (Model M-2), and the integrated multiple imputation approach (Model M-1). The numerical tests also include a sensitivity analysis of the number of imputations (m = 5, 10, 20 and 50) for all multiple imputation models. The experimental scenarios for evaluation include data missing rates of 20%, 40%, 60%, and 100% at Detector 10, which is a critical detector for use by both the estimation and the prediction modules of the system in all traffic scenarios.

Overall Performance over All Four Days

Figure 6-6 shows the distributions of average absolute relative errors (Eq. 4.16) from each of those four methods on all four days. In what follows, M-2-m and M-1-m denote Model M-2 and Model M-1 with m imputations, respectively. The results showed that Model M-1-50 has the best performance compared to all other models when the data is missing at 20%, 40% and 60%, and its performance is very similar to Model M-2-50 when the data is missing at a rate of 100%. Model M-2-50 provided a similar performance to MS and BF at the missing rate of 20%, but exhibited better accuracy than all three other methods at the missing dta rate of 100%.

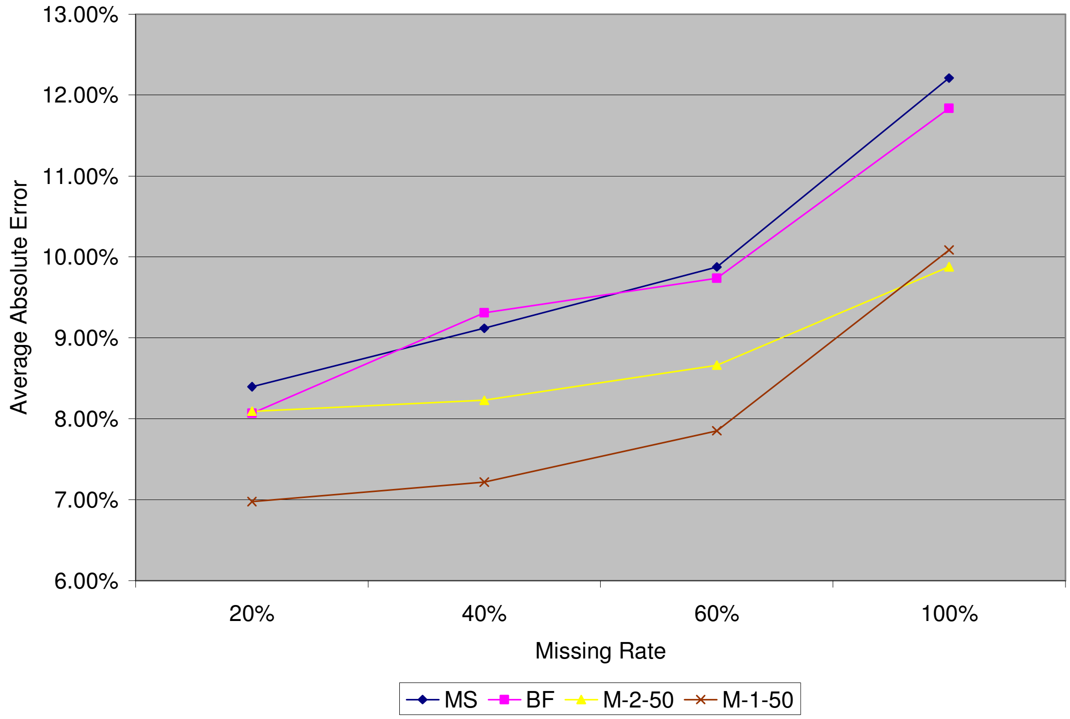


Figure 6-6 Average Absolute Relative Errors of All 4 Days under Different Missing Rates

Table 6-4 compares the performance of all methods in different travel time categories, which include congestion-free conditions (travel time less than or equal to 700 seconds), moderate congestions (travel time between 700 and 900 seconds), and heavily congested conditions (travel time exceeds 900 seconds). Model M-1-50 was the best among all models at the missing data rates of 20%, 40% and 60%, while Model M-2-50 outperformed the other three methods when Detector 10 could not function. Model-M1-50 and Model-M2-50 exhibited the same level of performance at the missing data rate of 100% in all three categories. Note that all sample cases contain sufficient information for executing Model M-1-50. There were a total of 22 cases on June 20th, 21st and 22nd, 2006 that did not pass the variance test by Model M-2-50. These cases have been removed from the evaluation.

Table 6-4 Performance of All Imputation Models in Different Traffic Conditions

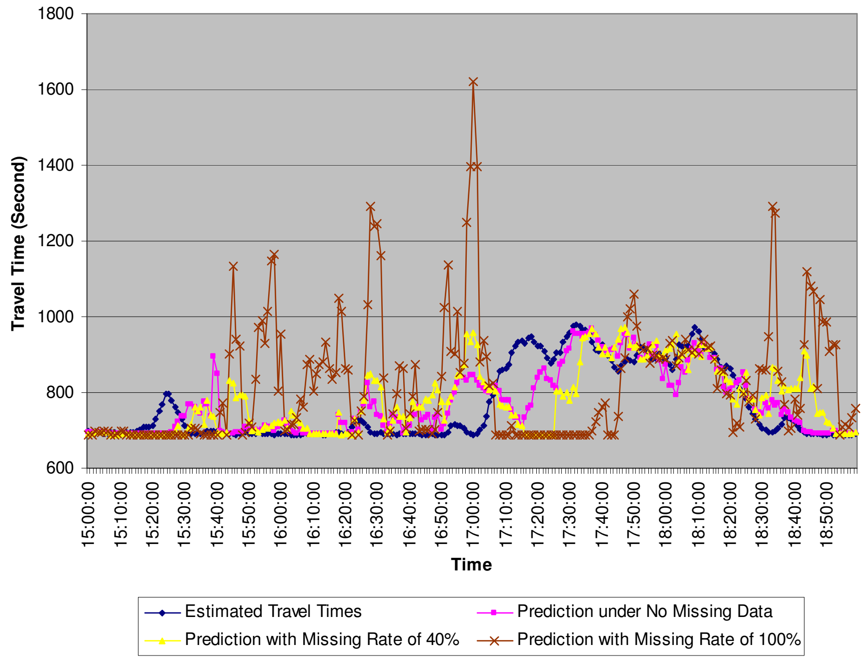
TT≤700		MS	BF	M-2-50	M-1-50
Missing Rate	20%	3.10%	2.78%	3.11%	2.54%
	40%	4.10%	3.63%	3.26%	2.80%
	60%	5.05%	4.47%	3.73%	3.07%
	100%	8.53%	7.47%	5.42%	6.37%
700<TT≤900		MS	BF	M-2-50	M-1-50
Missing Rate	20%	8.23%	7.35%	7.43%	6.65%
	40%	8.76%	8.56%	7.29%	6.43%
	60%	9.33%	8.64%	7.71%	6.95%
	100%	10.48%	10.26%	8.58%	8.66%
TT>900		MS	BF	M-2-50	M-1-50
Missing Rate	20%	13.46%	12.80%	12.36%	10.96%
	40%	13.82%	15.05%	12.57%	11.86%
	60%	14.29%	15.28%	12.99%	12.76%
	100%	16.12%	15.86%	13.55%	14.07%

TT: Travel time
MS: Mean substitute
BF: Bayesian forecast
M-2-50: Model M-2 with the number of imputation m=50
M-1-50: Model M-1 with the number of imputation m=50

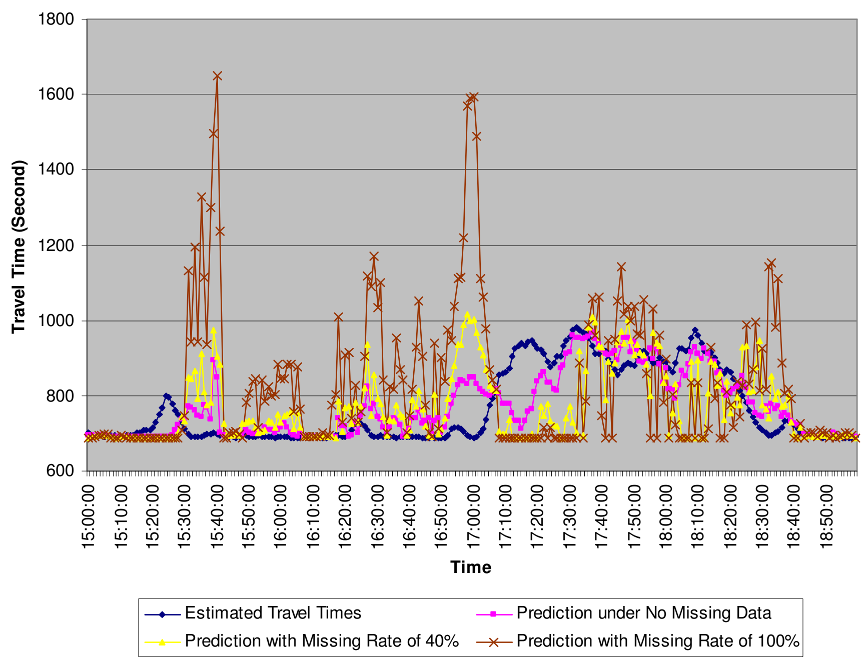
Performance Comparison with Individual Day Data

This study further explores the performance of each of the four tested models on a single day to evaluate the potential errors due to various congestion patterns. Among the four analyzed weekdays, prediction results for June 2006 exhibited larger errors due to missing data. As shown in Figures 6-6a and 6-6b, both MS and BF models, which are widely used in existing traffic data warehouse systems, provided satisfactory results when the missing data rate was 40% in the evening peak hours, except during the transition periods between uncongested and congested conditions. However, when the missing rate was 100%, both models yielded unacceptable prediction results.

Figures 6-7a and 6-7b show the prediction results from Model M-1-50 and Model M-2-50 under the same missing rates of 40% and 100% on June 20th, 2006. It is clear that travel time predictions with both multiple imputation models are more reliable and robust, especially during the transition periods. Model M-1-50 is much more robust than MS, BF, and MI-2-50; its largest prediction error was less than 4 minutes when detector 10 is not functioning at all. Model M-1-50 and Model M-2-50 have similar average absolute relative errors of 12.15% and 12.06%, respectively, over the entire evening peak on June 20th, 2006, compared to the prediction errors of 18.41% and 20.78%, respectively, for MS and BF.

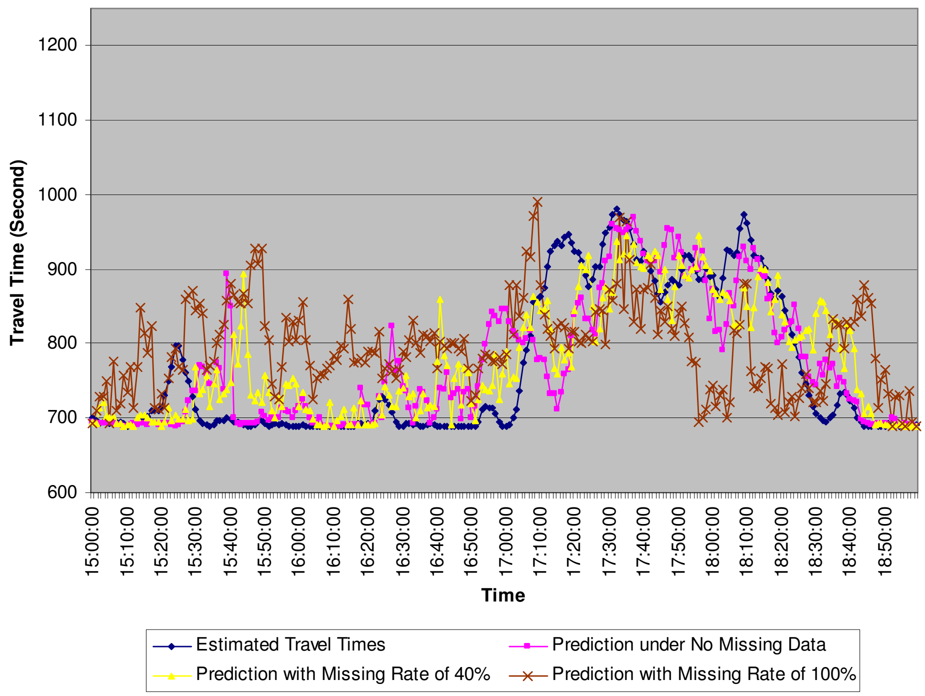


a) Mean substitution (MS)

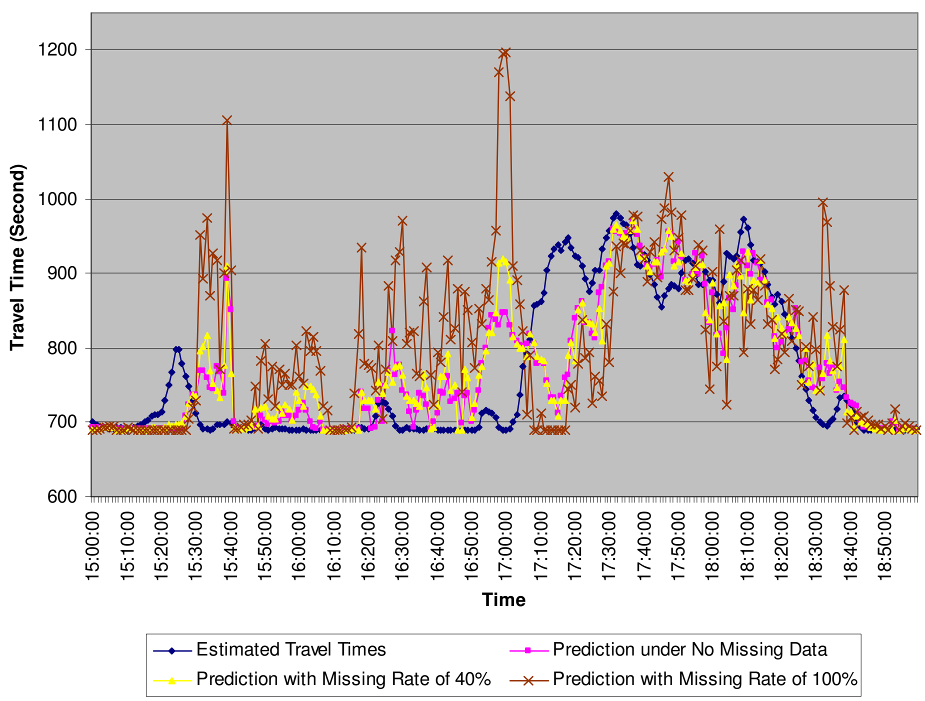


b) Bayesian forecast (BF)

Figure 6-7 Performance comparisons of MS and BF at missing data rates of 40% and 100% on June 20th, 2006



a) Multiple Imputation Model M-1 with *m*=50 (MI-1-50)



b) Multiple Imputation Model M-2 with *m*=50 (MI-2-50)

Figure 6-8 Performance comparisons of MI-1-50 and MI-2-50 at missing data rates of 40% and 100% on June 20th, 2006

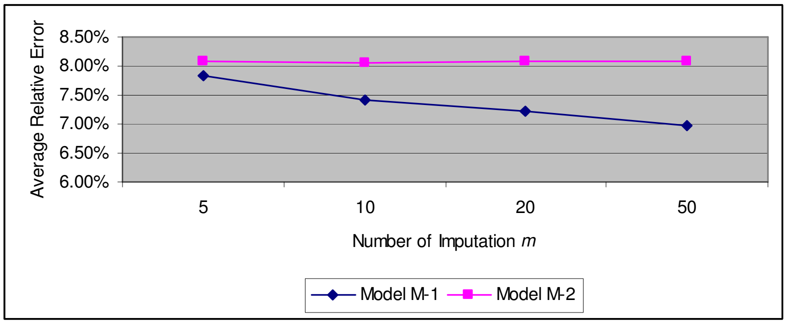
Comparison of Multiple Imputation Models

Since both Model M-1-50 and M-2-50 show better accuracy and reliability than other models on the data in these sample days, this section will further investigate their performance qualities under different numbers of imputations.

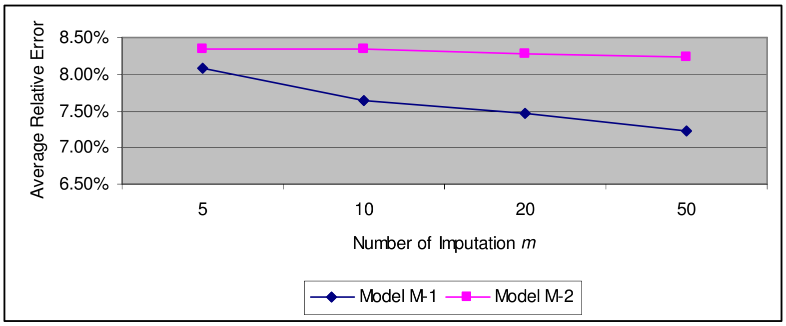
Figures 6-8a to 6-8d illustrate the average absolute relative errors from Model M-1 and Model M-2 on all four sample days with different numbers of imputations *m* and different missing data rates. Generally, a larger *m* improved the predicted travel time with the integrated multiple imputation method, Model M-1. Its performance increased more than 10% when the number of imputation *m* arguments went from 5 to 50. However, the increase of *m* has no impact on the performance of Model M-2. Its performance improvements are all less than 3% when *m* is increased from 5 to 50. The comparison results are consistent with Eq. 6.1 by Little and Rubin (1987), which suggested that *m* should between 3 and 10.

Table 6-5 Average Relative Errors of Model M-2

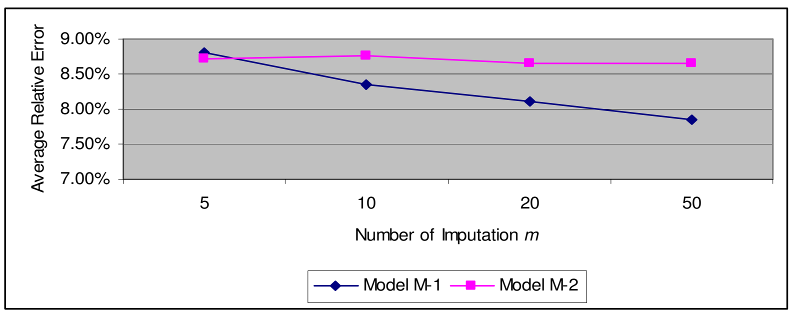
Missing Rate	*m* =5	*m*=10	*m*=20	*m*=50
20%	8.07%	8.06%	8.08%	8.09%
40%	8.34%	8.35%	8.28%	8.23%
60%	8.72%	8.76%	8.64%	8.66%
100%	10.16%	10.13%	9.85%	9.88%



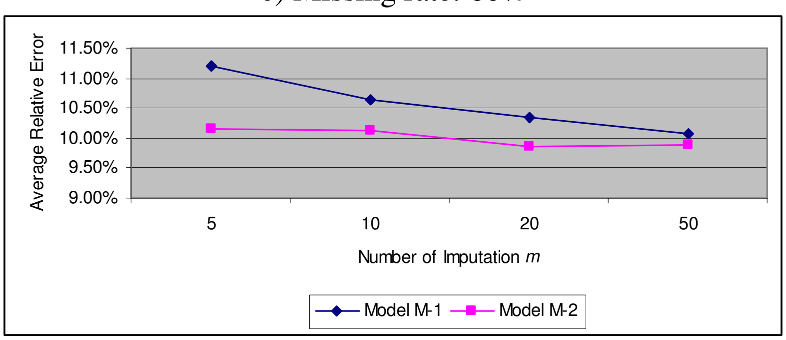
a) Missing rate: 20%



b) Missing rate: 40%



c) Missing rate: 60%



d) Missing rate: 100%

Figure 6-9 Average relative errors of models M-1 and M-2 on all four days with different *m* and missing data rates

Because the unique structure of Model M-1 was developed specifically for the proposed travel time prediction model, Little and Rubin's estimation of an efficient *m* does not fit this model. As shown in Figure 6-8 and Table 6-6, on average, the prediction error may increase about 5% when *m* decreases from 20 to 10 and increase about 3% when *m* decreases from 50 to 20. For a prediction error of 4 minutes, an increase of 3% is 7.2 seconds and an increase of 10% is 24 seconds. Hence, one can determine the number of *m* based on the required accuracy of the application. For example, an increased accuracy of 7 seconds may not be critical for a travel time prediction system that displays predicted travel times for commuters.

Table 6-6 Performance Improvements of Model M-1 with the Increase of *m*

Increase of *m*	Missing Rate			
	20%	40%	60%	100%
From 5 to 10	5.38%	5.32%	5.32%	5.12%
From 10 to 20	2.60%	2.82%	2.82%	2.82%
From 20 to 50	3.31%	3.12%	3.12%	2.39%

Overall, the developed missing data estimation module, which consists of an integrated multiple imputation model for predicting the travel time directly and a multiple imputation model for estimating the missing detector data, demonstrated its potential for practical use, based on the experimental results with the field data (June 20th, 21st, 22nd and 26th, 2006). Both models provided better travel time predictions than other widely-used methods. With the number of imputations set at 50, the integrated model provided acceptable accuracy and robustness over those sample days of the field study.

6.6 Conclusions

This chapter has developed two multiple imputation models, one integrated imputation model for the travel time prediction (Model M-1) and one multiple imputation model for estimating the missing detector data (Model M-2). Both models take into account of geometric features and traffic patterns for better accuracy and robustness, and both models incorporate the ability to estimate the reliability of the output. In the evaluation based on data collected from 10 roadside detectors on I-70 eastbound, both Models M-1 and M-2 outperformed commonly used methods (mean substitution and Bayesian forecast) when missing data rates of 20%, 40%, 60% and 100% occurred at a critical detector. A sensitivity test showed that the performance of Model M-1 may increase more than 10% when the number of imputation (m) increases from 5 to 50. The sensitivity analysis results for Model M-2 are consistent with the number of imputations reported in the literature.

Chapter 7: Research Summary and On-going Tasks

7.1 Research Summary and Contributions

This research focuses on the development of a real-time travel time prediction system with sparsely distributed detectors. By taking into account real-world constraints, such as detector reliability, traffic variability, and operating cost, this study has developed a system that can provide reliable prediction of travel time under recurrent traffic patterns with much less number of traffic detectors than the state-of-practice by the traffic community. With its embedded missing data estimation module, the prediction system is able to extend its operations under certain missing data scenarios, and turns off the function if the error caused by the missing data cannot be accommodated with existing theoretical methods. The key research issues associated with developing such a system are presented in Chapter 1.

Chapter 2 has provided a comprehensive literature review that covered the following topics: travel time estimation, travel time prediction, existing simulated and real-world application systems. It has been found from the review that most existing studies for travel time estimation and travel time prediction are for short links with densely distributed detectors (e.g., one detector every 0.5 miles). Nonparametric models, such as *k*-Nearest Neighbor Model, are reported to outperform the parametric models. Previous studies also indicated that a proper combination of different models may improve the system reliability. In contrast to the use of sophisticated algorithms presented in most research studies, all existing real-world systems for traveler information employ simple algorithms that perform the prediction of travel time based on that assumption that the traffic condition within the predicted time horizon will be identical to the current detected traffic states.

In response to the identified needs and constraints, Chapter 3 proposed the framework of a travel time prediction system for use on most 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, the travel time estimation module continuously estimate travel times for those completed trips and store them in a database. The prediction module takes the real-time input from traffic detectors and then performs the prediction with its hybrid model structure. The missing data estimation module is responsible for imputing the missing variable during real-time operations, and then for estimating the potential impacts on the predicted travel time. The proposed operating architecture ensures that the travel times of newly completed trips can be added to the database in real time, and immediately available for the system to perform the prediction during the next time interval.

Chapter 4 has developed a hybrid travel time estimation model by combining a clustered linear regression model and an enhanced trajectory-based model. A clustering function will first categorize traffic patterns in a link, based on its congestion levels in critical lanes. Then, the travel time estimation module will further calibrate a linear regression model in each cluster that has sufficient samples of field data. For clusters without adequate sample data, this study has developed an enhanced trajectory-based model as a supplemental component, which integrates the traffic propagation relations with a piecewise-linear-speed-based model. With such a component, one can estimate the time-varying in-segment speed of a vehicle in a long link based on the distance from its location to detectors, and then approximate the link travel time. The results of extensive numerical experiments have showed that the developed hybrid model is able to provide acceptable accuracy with only 10 detectors on a 25-mile stretch of I-70 eastbound. This is far less than the number of detectors needed by state-of-art and state-of-practice studies.

Chapter 5 has detailed a hybrid model structure for the travel time prediction on freeways with sparsely distributed detectors. A multi-topology Neural Network model serves as the main model that uses a customized rule-based clustering function to take into account the impacts of geometry features and daily congestion patterns. The proposed Neural Network model does not rely on a large historical traffic database, thus can start to operate after a short period of system training (e.g., four weeks). In the hybrid model structure, an enhanced *k*-Nearest Neighbor model serves as the supplemental model for taking advantage of historical traffic conditions and travel times. With a customized searching function and criteria for traffic characteristics, the supplemental model can efficiently improve the system's reliability for less-frequently observed traffic scenarios with a rich historical database. The numerical examples have demonstrated that the proposed travel time prediction model provided satisfactory results and outperformed other models found in the literature under all types of traffic conditions in a detection environment with the detector spacing exceeding 1 mile.

To contend with commonly incurred missing data issue, Chapter 6 developed two missing data estimation models based on the multiple imputation technique. Model M-1 integrates the missing data estimation and the travel time prediction with a searching function similar to that developed for the travel time prediction module to ensure the reliability on travel time prediction under some missing data scenarios. Model M-1 is able to concurrently estimate the distributions of both missing data and the impacted travel times. If this model cannot produce an acceptable level of accuracy under the missing data scenario, the system will then switch to the secondary model, M-2, which will divide the target segment into subsegments based on the geometry features and the daily traffic patterns, and take available information from all critical lanes in the subsegment to estimate the missing data with the multiple imputation technique. The system will then execute the travel time prediction module with the completed dataset if the model can confirm the reliability of the imputed values. The developed missing data estimation module have proved its performance under various data-missing scenarios, and outperformed widely-used single imputation methods.

In summary, this research has made the following key contributions:

- Design a system framework of a travel time prediction system that takes traffic data from most common traffic detectors to provide reliable prediction of travel times on a freeway segment with large detector spacing under various geometry features and complex traffic congestion patterns. Such a system does not require concurrent measurement of travel times during its real-time operations. None of the systems found in the literature or implemented in the real-word applications can function properly with the number of detectors far less than the standard of practices in the traffic community.

- Develop a hybrid model, which combines a clustered linear regression model and an enhanced trajectory-based model, for estimating travel times on a freeway segment by categorizing traffic scenarios with identified critical information and applying the best model structure. The estimated travel times produced by such a module are sufficiently accurate for incorporating into the historical travel time database.
- Construct a hybrid travel time prediction model with a multi-topology Neural Network model, which uses a rule-based clustering function as the main model and an enhanced *k*-Nearest Neighbor model as the supplemental model for predicting travel times under recurrent traffic congestions and sparsely distributed detectors.
- Develop an integrated missing data estimation model with the multiple imputation technique contend with both short- and long-term data missing, which occurs frequently in a real-world system. The proposed models are capable of estimating the reliability of the imputed missing values and predicted travel times under the missing data impact so as to avoid the potentially large system errors.

7.2 Future Research

Future studies related to a real-time travel time prediction system with large detector spacing are listed below:

- Development of a Model to Determine Optimal Detector Locations

Both travel time estimation and travel time prediction models developed in this study can function reliably on a freeway segment consists of long links. However, some segments with highly fluctuating traffic conditions may experience a prediction error if it is far away from a detector station. The locations of 10 traffic detectors used in this study were predetermined based on the knowledge of local traffic patterns and geometry features. A model for determining the optimal detector locations may help improve the performance of a travel time prediction system, and reduce the impact of the missing data at key locations.

- Detection of Incidents and Other Special Events to Minimize the Potential Prediction Errors

This study focuses on the travel time prediction under recurrent traffic patterns. However, the reliability of such a system may be reduced under the impacts of an on-going accident and/or a special event, due to the fact that the actual travel time during such scenarios will be dependent on the duration of the incident that varies with its nature, efficiency of the response team or local management strategies, and availability of equipment or other supports. An incident detection model, which has quick and accurate detection with sparsely distributed detectors, can certainly improve the reliability of a travel time prediction system by recognizing the abnormal traffic conditions. The system can then inform the control center to switch the entire operations to the incident or special event management mode.

- Monitoring the Change in Traffic Patterns and Estimating the Potential Impacts

All models developed in this study take into account historical traffic congestion patterns to improve their reliability. Although recurrent traffic patterns may remain steady in a fairly long period in most freeway segments under normal operations, it is still likely that the traffic conditions may differ significantly from their typical patterns, such as having long-term work-zones on the target freeway segment or nearby area. Commuters may also change their driving patterns based on the knowledge of the information provided by responsible agencies. Therefore, a model is needed to monitor the change in traffic patterns and to alert the control center for recalibrating all key system parameters if a potential large impact is detected.

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.

Al-Deek, H., and Chandra, C., (2004) "New Algorithms for Filtering and Imputation of Real-Time and Archived Dual-Loop Detector Data in I-4 Data Warehouse," *Transportation Research Record* 1867, pp. 116–126.

Al-Deek, H., Kerr, P., Ramachandran, B., Pooley, J., Chehab, A., Emam, E., Chandra, R., Zuo, Yo, Petty, K., and Swinson, I., (2004) "The Central Florida Data Warehouse (CFDW), Phase-2-, The Central ITS Office Funding," *Research Report to Florida Department of Transportation*.

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.

Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P., (2003) "Detecting Errors and Imputing Missing Data for Single-Loop Surveillance Systems," *Transportation Research Record* 1855, pp. 160-167

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.

Dempster, A., Laird, N., and Rubin, D., (1977) "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (with Discussion)," *Journal of the Royal Statistical Society*, Series B, 39, pp. 1-38

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.

Gan, X., Liew, A. and Yan, H., (2006) “Microarray missing data imputation based on a set theoretic framework and biological knowledge,” *Nucleic Acids Research*, Vol. 34(5), pp. 1608-1619.

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

Graham, J. and Donaldson, S., (1993) “Evaluating Interventions with Differential Attrition: the Importance of Nonresponse Mechanisms and Use of Followup Data,” *Journal of Applied Psychology*, Vol. 78, pp. 119-128.

Haj-Salem, H., and Lebacque, J., (2002) “Reconstruction of False and Missing Data with First-Order Traffic Flow Model,” *Transportation Research Record* 1802, pp. 155-165

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, T., (2004) "TMC Traffic Data Automation for Mn/DOT's Traffic Monitoring Program," *Research Report to Minnesota Department of Transportation*, Report No. MN/RC-2004-29.

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.

Little, R., and Rubin, D., (1987) “Statistical Analysis with Missing Data”, *Wiley*, New York.

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.

Nguyen, L., and Scherer, W., (2003) "Imputation Techniques to Account for Missing Data in Support of Intelligent Transportation Systems Applications," *Research Project Report for the Virginia Department of Transportation*, Research Report No. UVACTS-13-0-78.

Ni, D., Leonard II, J., Guin, A., and Feng C., (2005) "Multiple Imputation Scheme for Overcoming the Missing Values and Variability Issues in ITS Data," *Journal of Transportation Engineering*, Vol. 131 (12), pp. 931-938.

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.

Rubin, D., (1976) “Inference and Missing Data,” *Biometrika*, 63, pp. 581-592.

Rubin, D., (1987) “Multiple Imputation for Nonresponse in Surveys,” *Wiley*, New York.

Saunders, J., Morrow-Howell, N., Spitznagel, E., Dork, P., Proctor, E., and Pescarino, R., (2006), “Imputing Missing Data: A Comparison of Methods for Social Work Researchers,” *Social Work Research*, Vol. 30(1), pp. 19-31.

Schafer, J., and Schenker, N., (2000) “Inference with Imputed Conditional Means,” *Joruanl of the American Statistical Association*, Vol 95, pp. 144-154.

Schafer, J., and Graham, W., (2002) “Missing Data: Our View of the State of the Art,” *Psychological Methods*, Vol. 7(2), pp. 147-177.

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.

Smith, B., Scherer, W., and Conklin, H., (2003) “Exploring Imputation Techniques for Missing Data in Transportation Management Systems,” *Transportation Research Record* 1836, pp. 132-142

Souverein, O., Zwinderman, A., and Tanck, M., (2006), “Multiple Imputation of Missing Genotype Data for Unrelated Individuals,” *Annals of Human Genetics*, Vol. 70, pp. 372-381.

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.htm. 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

Wen, Y., Lee, T., and Cho, H., (2005), "Missing Data Treatment and Data Fusion Toward Travel Time Estimation for ATIS," *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, pp.2546-2560.

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.

Zhong, M., Sharma, S., and Lingras, P., (2004) "Genetically Designed Models for Accurate Imputation of Missing Traffic Counts", *Transportation Research Record* 1879, pp. 71–79.

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.