ABSTRACT

Title of Document:	REAL-TIME SHORT-TERM TRAVEL TIME PREDICTION
	WENXIN QIAO, DOCTOR OF PHILOSOPHY, 2012
Directed By:	ALI HAGHANI, PROFESSOR DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

Real-time short-term travel time prediction is a critical component of the Intelligent

Transportation System (ITS) and an important element of the Advanced Traveler

Information System (ATIS). Accurate and reliable travel time prediction enables both

user and system controller to be well informed of the likely future conditions on

roadways, so that pre-trip plans and traffic control strategies can be made accordingly

in order to reduce travel time and relieve traffic congestion. With these travel time

predictions, roads may be used more efficiently with better overall network

performance. This research will study short-term travel time prediction for freeway

applications using various sources of real time travel time data. The integrated

prediction model proposed here will put emphasis on travel time prediction under

various traffic and weather scenarios and especially inclement weather conditions.

REAL-TIME SHORT-TERM TRAVEL TIME PREDICTION

By

Wenxin Qiao

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
2012

Advisory Committee:
Professor Ali Haghani, Chair
Professor Paul Schonfeld
Assistant Professor Cinzia Cirillo
Assistant Professor Lei Zhang
Professor Christopher Davis

Dedication

To my parents

For their love and support

Acknowledgements

Throughout the time I was writing this dissertation, I received help from so many people who deserve my special acknowledgment.

To my great advisor, Professor Ali Haghani, whose strong and continuous support guaranteed the completion of my dissertation. His thoughtful and wise advice always directed me on the right track for my study and I cherish everything I learned from him as I face obstacles and pursue opportunities. I would especially like to thank him for giving me the courage to embrace the challenges and to defend my beliefs. I will always appreciate the help and care he gave to me.

I would also like to give my deep appreciation to Professor Paul Schonfeld, Professor David Chris, Professor Cinzia Cirillo, and Professor Lei Zhang who served on my dissertation committee. I want to thank them for contributing their precious time and expertise to better my work.

I feel grateful as well to my supportive colleagues and loving friends for being there to offer help to me whenever needed. I would like to take this opportunity to give my special thanks to Masoud, who always gave me sound suggestions and advice on my research without hesitation. I also want to thank Kaveh, Yang and other members in our group. I appreciate their help and effort whenever I met with difficulties as well as their great encouragements. I also want to express my gratitude to my dear colleagues at the Maryland State Highway Administration for the practical expertise they offered, the experiences we shared, as well as their emotional support. Although I have not named everyone who contributed to my studies at the University of Maryland, they will always be on my mind and my deepest appreciation goes to those who are not named here but played a positive role in my life during these years.

Last, but not the least, I would like to dedicate this dissertation to my parents, whose love and support have kept me moving forward, at every step, throughout my life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Study Background

      1.1.1 Short term travel time prediction and ITS

Promoting energy efficiency and environmental quality; ensuring safe and efficient

travel choices and improving mobility are the strategic transportation goals of the

nation. According to 2010 TTI Urban Mobility report, congestion caused urban

Americans to travel 4.8 billion hours more and to purchase an extra 3.9 billion gallons

of fuel for a congestion cost of $115 billion in 2009. According to the 2008 Condition

and Performance report, the average daily percent of vehicles miles traveled under

congested conditions has increased from 24.9% in 1997 to 28.6% in 2006 for all

urbanized areas combined. As the network congestion level rises rapidly, various

strategies are applied to relieve traffic congestion. An efficient and reliable

transportation network is required to relieve urban congestion and reduce air pollution

and traffic accidents. One of the approaches is to expand the road capacity however,

adding capacity to existing networks is not sometimes practical due to the restrictions

of land area and project costs. A more viable approach would be using the existing

network resources more effectively to provide better road service level. Intelligent

Transportation Systems (ITS) technologies provide solutions to congestion problems.

According to the AASHTO report (2007), the advanced ITS technologies along with

better system management techniques needs to be utilized to reduce congestion,

improve the throughput and increase system reliability. A variety of intelligent

transportation systems have already been developed and applied in transportation

networks. To name a few, we have Advanced Traveler Information Systems (ATIS),

Advanced Traffic Management Systems (ATMS) and Routing Guidance Systems (RGS).

One critical component of ITS is traffic data. Accurate and reliable traffic data serve as the foundation of all ITS and guarantee acquisition of usable output. With the rapid development in electronics and internet technologies, a variety of traffic data such as traffic volume, travel speed, detector occupancy, and travel time are now available. Among these, travel time is one of the most important data since it provides the users the most direct conception of the current traffic condition that is easy to be perceived.

Short-term travel time prediction has long been serving as a critical element of the ITS and an important base of the ATIS. A robust ITS is not just providing reactive services, it is also moving towards a more proactive system and travel time prediction is an essential input element for such a system. As congestion increases rapidly in most urban networks, providing reliable travel times can help road users to choose an optimal route in order to shorten their travel time, relieve traffic congestion, reduce air pollution and save energy. Travel time information can be delivered to road users for either pre-trip planning or during the trip. Pre-trip travel time information enables the user to make decisions on the best route to take and travel time provided during the trip gives user the option to take an alternative route with less travel time or at least relieve the anxiety resulting from being unaware of the situation.

1.1.2 Modern technology and Challenge

The long term efforts to develop, demonstrate and deploy ITS tools have shown their benefits. Modern technology makes high quality automatic vehicle identification devices available to be installed on the roads, which makes it possible to perform short-term traffic flow analysis and develop forecasting techniques. However, predicting travel time is very challenging since the accuracy of results varies with many variables such as: day-to-day traffic demands, individual driver behavior, weather condition, incident occurrence, detectors' accuracy and reliability and so on.

The factors contributing to the unpredictability of traffic systems include among others: accidents, erratic driver behavior and various weather conditions. Given that the nature of transportation networks is dynamic, unstable and complex, it is critical for the prediction model to be able to fully capture the stochastic nature of the travel time and to exhibit robust performance under various traffic conditions: free flow, recurrent congestion, and non-recurrent congestion caused by accidents or inclement weather or other externalities.

The complicated interrelations between detectors, historical data, and traffic flow characters have made travel time prediction challenging. This is one reason why most real world systems provide travel times to the public based solely on the estimation of current traffic conditions, instead of a prediction. However, obtaining predicted travel time information is a necessity for both en-route trips and pre-trip planning. To contend with these issues, researchers have proposed and implemented a variety of

approaches for providing predicted travel times in the past decade. The research

objective in this study is to develop an efficient and reliable travel time prediction

model which will generate great benefits both on the road user's (the traveler) side

and the control decision maker's (the traffic management center) side, providing

better network performance.

1.2 Definition

Travel time is the time that it takes for an individual vehicle to traverse a unit length

of roadway. Short term travel time prediction is the process of estimating the

anticipated travel time at a future time, given the historical data and continuous

feedback of current travel time information. Travel time prediction can be short-term

(5 minutes to 15 minutes into the future) and/or long-term (1 hour, a day).

Figure 1 shows the concept of travel time prediction through time and space:
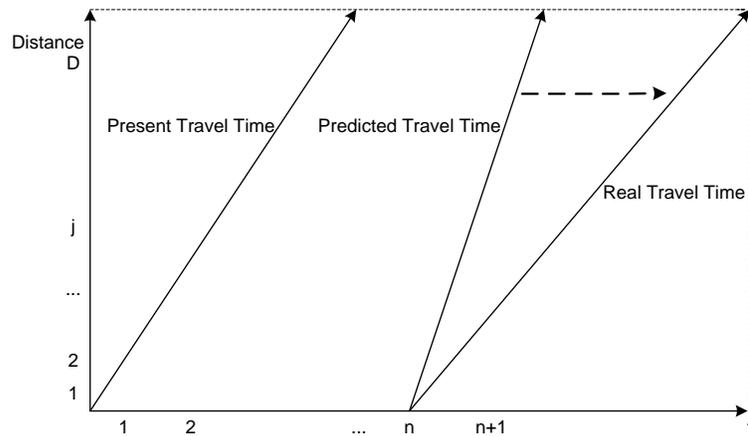
Figure 1 Travel time prediction through time and space

4

1.3 Research Motivation and Contribution

1.3.1 Research Motivation

For the past three decades, traffic flow prediction has been explored immensely. Numerous prediction models and algorithms have been proposed and applied by researchers. With modern technology developments, travel time prediction has become most popular in the research area due to its essential role in the intelligent transportation systems. A variety of models including regression models, time series models, Kalman filter, neural network, nearest neighbor models, support vector regression models, as well as simulation approaches have been developed for travel time prediction. As one of the most fundamental inputs in the ITS, travel time information in the recent future is very much needed for both travelers to make trip decision and traffic management center for developing strategies for operation control. As a result, a reliable short term travel time prediction model is needed.

The United States State Departments of Transportation provide both historical and real-time traffic data, which make it possible to develop the data driven models. At the same time, federal agencies have put great effort in developing reliable road weather management system to save lives, time, and money from the inclement weather. The Office of Operations of Federal Highway Administration (FHWA) in conjunction with the Intelligent Transportation Systems office of the Research and Innovative Technology Administration (RITA) developed the Road Weather Management Program (RWMP) to address road weather challenges through research, technology development, community outreach and promotion of strategies and tools.

This research intends to develop an integrated travel time prediction model that is based on historical dataset and available real time traffic information and perform consistently accurately under various traffic scenarios, especially inclement weather conditions.

1.3.2 Research Contribution

Many travel time prediction models have already been proposed and applied in traffic systems, and they perform pretty well in case studies. However, incorporating the weather information into the prediction model has only recently been studied. Some researchers have studied the impact of rain or snow on the traffic flow from both supply and demand sides. Recently, a few researchers included weather information as part of the prediction models. Most of these research studies are based on either regression models to include weather information as an explanatory variable or are simulation based approaches. The research in considering weather effect in travel time prediction is sparse and very limited. However, studying traffic conditions under various weather conditions, especially for inclement weather conditions is very important and necessary. Accurate information regarding changes in weather conditions is critical for the transportation system to remain safe and efficient. Addressing weather impacts on traffic congestion has a significant potential in mitigating congestion and ensuring safety. Accurate and timely weather information helps users make better decisions and respond to the driving requirements under adverse weather condition well. This research will focus on two parts: one is to develop a reliable short term travel time prediction model; the other part is to study in

depth the impact of weather conditions on travel time prediction and to incorporate

the weather information in travel time prediction under various weather conditions.

1.4 Organization of the Dissertation

In this research several existing prediction models that have proved to work

efficiently are applied to traffic data to perform prediction. A modified non-

parametric model K nearest neighbor model KNN-T is proposed that will enhance the

traditional KNN model with trend adjustment. The prediction results obtained from

each model are compared and discussed. Then, an integrated travel time prediction

model which incorporates various sources of traffic and weather data is proposed and

its prediction efficiency is investigated through several case studies. Last, an

extension of the integrated model is proposed adding the features of path travel time

prediction and multi-step ahead travel time prediction.

This dissertation is organized as follows. First, the study background introduces the

importance of short term travel time prediction from the perspective of the user and

network controllers. The challenge of prediction is also illustrated. A comprehensive

literature review is conducted, which includes both traditional prediction models and

recent studies of weather impacts on traffic stream. In the third part the research

problem is described, a set of widely used prediction models that are proved to work

efficiently are implemented to datasets from freeway segment for travel time

prediction. These include Historical Average, ARIMA, Kalman filter and K-nearest

neighbors. A modified non-parametric model KNN-T is proposed that enhances the

traditional KNN model with trend adjustment. Then a small case study is conducted applying Bluetooth travel time data from a freeway segment. Performances of each model from case studies are investigated and reported. Bluetooth travel time data collected from the sensors deployed at freeway roadsides is used for model calibration and validation. In the fourth part, a new integrated prediction model incorporating weather impacts is proposed. The results from this model are compared with the results generated from previous ones. Case studies are designed to examine the performance and the efficiency of the proposed integrated model on selected freeway segments. In the fifth part of this dissertation, the proposed integrated prediction model is further enhanced while adding the features to perform prediction on longer freeway path composed of several continuous segments as well as multi-step ahead prediction. Prediction performance from 5 minutes up to 30 minutes ahead of time are investigated and discussed. Finally, a summary is given and directions for future research are discussed.

Chapter 1 provides the introduction and the importance of short term travel time prediction. Chapter 2 includes a comprehensive review of the literature in parametric and non-parametric models as well as their application in existing systems, weather impact on traffic stream and related prediction models. Chapter 3 introduces the statement of the problem and the travel time prediction models that are tested in this research. It also presents the comparison of the results obtained from these models in the first case study. Chapter 4 discusses the proposed integrated model that incorporates the impact of weather conditions. It also discusses the results of

comparison of different models and their performance using combined data sources in the second case study. Chapter 5 discusses the extension of the integrated model that incorporates the features of path travel time and multi-step ahead travel time prediction. Results of comparison of different models and their performance in the third case study are provided as well as the sensitivity analysis on the efficient size of historical dataset. Finally, Chapter 6 presents the conclusions and directions for future research.

Chapter 2: Literature Review

2.1 General Prediction models

Prediction is a statement about the way things will happen in the future. Prediction models have been applied in a variety of areas such as the stock market, natural disasters, pandemics, demography, climate and meteorology. Mulhern and Caprara (1994) forecast market response and provided an empirical demonstration using store scanner data for consumer packaged goods. They introduced a multivariate methodology that uses a nearest neighbor technique to represent time series behavior that is complex and non-stationary. Karlsson and Yakowitz (1987) forecast the rainfall runoff through a rainfall-runoff model with an eye toward the advantageous use of the massive data sets being accumulated and the modern computers capable of dealing effectively with such sets. Prediction models are developed and used widely in a variety of both research and practice areas and they are promoted with modern technology developments. The prediction results can help people be prepared for future conditions and facilitate making rational decisions about future plans.

2.2 Travel time prediction models

During the past three decades, a variety of traffic prediction approaches have been developed and explored in the literature. Numerous models have been proposed for the prediction of traffic volumes, speeds, and travel times. There is no uniform way to categorize the variety of existing traffic prediction models. Generally we can categorize the existing models into two types: Parametric models and Non-Parametric models. The main techniques applied in these two categories are discussed below.

2.2.1 Parametric models

Parametric models can be divided into statistical parametric techniques and state space models. Statistical parametric techniques are categorized as follows.

(1) Historical Average models use an average of past traffic streams to predict future traffic streams. These are linear based models that are easy to understand, simple to apply but can't deal with real-time, stochastic and unstable traffic data. These models have applications in the early urban traffic control system (UTCS) (Stephanedes et al. 1981) and traveler information systems AUTOGUIDE (Jeffrey et al. 1987) and LISB (Kaysi et al. 1993) in Europe. These models are also commonly used as the naïve models for model accuracy performance comparison.

(2) Linear Regression models predict the expected value of a dependent variable in response to changes in one or more independent variables. These models are developed using the least squares method. The objective is to minimize the sums of squared residuals to obtain the best fit. Kwon et al. (2000) presented a linear regression approach with stepwise-variable selection method and tree-based methods to estimate future travel times on freeways. Both methods performed satisfactorily. Rice and Zwet (2004) proposed a prediction method which arises from empirical observation that there exists a linear relationship between any future travel time and the current status travel time. The prediction scheme is by means of linear regression with time-varying coefficients. Kwon and Petty (2005) proposed a travel time prediction algorithm with a time-varying coefficient (TVC) linear regression model as

the component predictor, which is scalable to large freeway networks with many nodes with arbitrary travel routes. The algorithm improves the baseline historical travel time predictor with a 40% to 60% reduction in the prediction error. van Hinsbergen and van Lint (2008) proposed a Bayesian combination framework with the use of two simple linear regression models as a showcase, and showed that this Bayesian combination improved prediction accuracy for real-time applications, but a Bayesian combined model is sensitive to bias. If all models have a bias with the same sign, then the Bayesian framework will have a larger prediction error than the best of the single models. It is recommended to increase the number and diversity of the models inside the model layer of the framework to decrease the chance of all models having the same bias.

(3) Time Series models assume that the knowledge of the past values in a time series is the best predictor of the variable in the future. These models includes: Autoregressive model (AR); Moving Average model (MA); Autoregressive Moving Average model (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA). 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). The seasonal autoregressive integrated moving average model (SARIMA) has also been applied to travel time prediction to cope with the seasonal pattern exhibited in the traffic flow (Williams et al. 1998, Williams 1999). Recently, Cetin and Comert (2006)

used ARIMA models for developing forecasting models while the process

mean is monitored by two detection algorithms to account for occasional

regime changes. The intercept of ARIMA is updated on the basis of the

detected shifts in the mean level to adapt to any potential new regimes.

Results showed significant improvements in accuracy compared with

traditional ARIMA models with fixed parameters. Farokhi et al. (2010)

evaluated the performance of three moving average techniques (one simple

moving average method with constant weights and two adaptive moving

average methods) in predicting average travel speeds up to 10 minutes ahead

of time. Results indicated that the method using optimized weights produced

slightly better predictions at a higher computational cost.

The most widely used State Space model is applying the Kalman Filter (KF)

technique, which was first applied in traffic volume prediction by Okutani and

Stephanedes (1984). It is based on the Kalman Filtering theory proposed by Kalman

(1960). This model describes the dynamic system in modern control theory. KF

provides a computational scheme to adapt the parameters of a model to observed

system states, trying to minimize the state estimation error conditioned on the

acquired measurements. This model is generally composed of two basic equations, a

state equation and an observed equation and has been successfully applied in

prediction techniques with a high degree of accuracy. Recently, Yang et al. (2004)

proposed a recursive least square (RLS) approach for speed prediction based on a KF

model to adapt to changed pattern quickly. Results showed that most of the true value

fell in the 95% confidence interval of less than 10 mph. Xie et al. (2006) combined

the wavelet decomposition with KF model for speed prediction and study results indicated the wavelet KF model performed better than the KF model.

2.2.2 Non-parametric models

K Nearest Neighbor (KNN) and Neural Networks are two of the mainly used techniques in Non-Parametric models. Most non-parametric models share a common feature of searching a collection of historical observations for one or more records that are similar to the system's current state and use such records to perform the prediction, for example, the K-nearest Neighbor model. The non-parametric models exhibit advantages especially under stochastic conditions (Disbro and Frame 1989; Mulhern and Caprara 1994). The first KNN model to forecast traffic volume was developed by Davis and Nihan (1991). More recently, Bajwa et al. (2003, 2004) proposed a KNN prediction model along with a genetic algorithm to generate adaptive parameters. Zou et al. (2009) proposed a hybrid travel time prediction model that combined the KNN and neural network models.

Neural network models find the complex relationships between inputs and outputs through learning processes and generalize to new examples (Zhang et al. 1998). Neural Network models have the capability of pattern recognition and the feature of robustness. These models require large sampling, and the training process is usually very long. They also suffer from an over-fitting problem. Neural network models hold the assumption that nonlinear relationships exists in the traffic data. They have some drawbacks such as: being trapped in local optima, over fitting and large computation burden, however, they still draw research interests with their ability to perform self-

learning and deal with non-linear problems. Clark et al. (1993), and Smith and Demetsky (1994), applied such topology for prediction: a basic and fully connected back propagation multilayer perceptron (MLP) consists of one input layer, one hidden layer and one output layer. Park and Rilett (1998) compared a neural network with KF and exponential smoothing model and showed that the NN model has better performance. Yun et al. (1998), and Lingras and Mountford (2001) applied the time-delay neural networks (TDNN) for prediction by incorporating one tapped delay line in the input layer to better fit the nature of the time-series data, so input time-series data items will travel through the tapped delay line to provide TDNN with a better short-term memory. Vlahogianni et al. (2005) extended past research by providing a genetic algorithm based, multilayered structural optimization strategy that assists both in the proper representation of traffic flow with temporal and spatial characteristics as well as the selection of appropriate neural network structure. Satisfactory results were indicated.

2.2.3 Other models

Besides the above models, there are also some other models that have been proposed in this research area including: Wavelet Analysis based models (Xiao et al., 2003, Jiang et al., 2005); Chaos Theory based model (Wang, 2005); Catastrophe Theory based models (Navin, 1986, Forbes and Hall, 1990); Support Vector Regression Models (Wu et al., 2003, Lam and Toan, 2008); Traffic simulation based model (Liu et al. 2006); Cell transmission based model (Juri et al., 2007) and Dynamic Traffic Assignment (DTA) based model (Ben-Akiva et al., 1992).

Considering the complex and dynamic nature of traffic flows in the system, using one

model to perform prediction usually cannot capture the complete characteristics of the

stochastic traffic data, thus may not predict the traffic under various conditions with

high accuracy. As a result, many hybrid models are being developed and proposed in

the recent years. Hybrid methods usually use two or more models together along with

a clustering approach and then assign one model structure to each cluster with locally

fitted parameters. Relevant research was conducted by Chen et al. (2001), Lingras

and Mountford (2001), Yin et al. (2002), Zheng et al. (2006), and Zou et al. (2009).

2.2.4 Prediction systems

Several traffic prediction systems currently are being used across the world. In the

United States, TrEPS (Traffic Estimation and Prediction System) developed by

FHWA (Federal Highway Administration) is the central traffic information provider

serving as the supporting component for ATIS and ITS. In Europe, the system

CAPITALS provides traffic information in capital cities including Berlin, Paris,

Brussels, Madrid and Rome. In England, Traffic England is used for traffic

management on the freeway network. BayernInfo provides traffic prediction

information in Germany. In China, www.BJJT.cn is a website developed by the

Beijing municipal transportation information center to provide comprehensive real

time road information both for user's trip planning and management center's control

decision making.

2.3 Weather impacts on traffic stream

2.3.1 Research on weather impacts on traffic stream

Weather significantly affects the capacity and safety of the highways. Poor road conditions (i.e. wet pavement, low visibility) lead to slow down for the drivers, causing significant roadway capacity reductions that will considerably increase travel time and may contribute to accidents. Providing updated weather conditions to drivers is critical for the transportation system to remain safe and efficient. Maze et al. (2006) showed that weather condition does matter to the traffic demand, safety, operations and flow. The majority of research related to the impacts of weather conditions on traffic flow fall into three categories: demand, operation and safety.

To analyze the impacts of weather on traffic demand, Chung et al. (2005) investigated the effects of rainfall on travel demand and travel time and found travel time is longer for high density traffic and not significant at low density traffic during rainy periods. Cools et al. (2008) used the linear regression approach to identify and quantify weather effects on traffic volume, which is considered to be closely related to road safety. Results indicated that snowfall, rainfall and wind speed delimit traffic volume, while high temperature would significantly increase traffic volume. Samba and Park (2010) proposed a probabilistic approach to determine the average reduction of traffic volume under rain and snow. They found that inclement weather has a probabilistic impact on demand. Reduction varies with respect to time of day and snow has a larger impact on volume than rain. Datla and Sharma (2010) conducted investigations of traffic variations with severity of cold, amount of snow and various combinations of

17

the two. Results indicate that association of highway traffic flow with cold and snow varies with day of week, hour of day and severity of weather conditions. 1% to 2% reduction in volume for each centimeter snowfall is observed with mean temperature above zero. Watkins and Hallenbeck (2010) analyzed the impact of rain on freeway travel times in greater Seattle and found that rainy weather causes congestion only in the presence of sufficient volume. Significant differences in average travel time due to the rain are seen only during peak hours and rain has much greater influence as the peak is building than towards the end of the peak period. They used sensor volume and NOAA weather data.

In operations, Lamm et al. (1990) concluded that speeds are not affected by wet pavement until visibility is reduced. Other than light rain, heavy rain affects operating speeds and has a noticeable effect on traffic flow behavior. Also, Ibrahim and Hall (1994) found minimal reductions (2.0 km/h) in operating speeds in light rain, but significant reductions in heavy rain (5.0 to 10.0 km/h) during free-flow conditions. Similar results were found in the case of snow where light snow has minimal effects (0.96 km/h) and heavy snow resulted in a (38.0 to 50.0 km/h) in free-flow speed reduction. Kyte et al. (2000) found that the effects of light rain or snow and heavy rain may be 50% higher and the effect of heavy snow may be about 20% lower than stated in the Highway Capacity Manual 2000. Also, the effect of high wind should be included in the assessment of free-flow speed. Chung et al. (2006) found that highway capacity is reduced between 4%-7% for light rain and 14% for heavy rain. Tu et al. (2007) found that on average, adverse weather results in twice the travel time

variability compared with normal weather conditions. Rain has little or no effect on travel time variability below a certain critical inflow, but progressively impacts on travel time variability above it. Byun et al. (2010) provided a procedure for estimating the average speed using data collected under rain and congested conditions. They used regression analysis to develop a speed-flow model to describe conditions under clear weather, rain and congested conditions and found that as flow increases, speed decreases under clear and rainy conditions.

To consider safety, Edwards (1998), Keay and Simmonds (2006), Qiu and Nixon (2008) studied the effect of weather on road accidents and found that rain increases the accident frequency. Edwards (1998) also found that rain decreased the severity of the accidents.

2.3.2 Weather impact on travel time prediction

Previous research concluded that weather has significant impact on traffic flow and quantified the impact on both traffic demand and supply. However, the area of incorporating weather impact in short term travel time prediction is less explored. Huang and Ran (2003) proposed a neural network model, including weather conditions as explanatory variables, for predicting the traffic speed under adverse weather conditions. Similarly, Butler et al. (2007) examined the effect of including rainfall inputs in forecasting of daily traffic volumes through neural networks and suggested a smaller sampling interval like 15 minutes and a more rain rich data set. Hranac et al. (2006) proposed a weather adjustment factor (WAF). Let v be the visibility, r be the precipitation intensity of rain and s be the precipitation intensity of

snow, if the weather impacted link is characterized as (v, r, s), a set of WAFs can be calculated for this link using the following equation where WAFs vary as a function of the precipitation type, precipitation intensity, and visibility level:

$$F_i = \beta_0 + \beta_1 * v + \beta_2 * r + \beta_3 * s + \beta_4 * v * r + \beta_5 * v * s \tag{1}$$

F_i is the weather adjustment factor for parameter i;

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are coefficients;

Let x_{normal} be the travel time under normal weather condition (clear weather or base condition), the estimated travel time under inclement weather $x_{inclement}$ can be represented as: $x_{inclement} = F_x * x_{normal}$

To calibrate the equation, these steps need to be followed:

1: Collect weather data and define weather conditions in the form of (v, r, s);

2: Collect, associate and classify weather conditions with traffic observations (travel time observations);

3: Calibrate parameter of traffic flow model for each weather condition and calculate WAF for each parameter: $F_x = \frac{x_{inclement}}{x_{normal}}$;

4: Establish the relation of WAF and weather condition through a linear regression model.

Rakha et al. (2008) quantified the impact of inclement weather on traffic stream behavior and key traffic stream parameters using weather and detector data obtained from Baltimore, Minneapolis–Saint Paul and Seattle. They demonstrated that jam density is not affected by weather conditions. Reductions in free-flow speed and speed at capacity increase as the rain and snow intensities increase and snow has larger impacts compared with rain. Reductions in capacity are not affected by the

precipitation intensity except in the case of snow. This paper also developed weather

adjustment factors for key traffic parameters: free-flow speed, speed-at-capacity, and

capacity. These factors are multiplied by the base clear-condition variables for use in

HCM to compute inclement weather parameters. Dong et al. (2010a, b) addressed

both supply and demand aspects of users' response to adverse weather and proposed a

framework for evaluation and implementation of weather responsive advisory and

control strategies using real time traffic estimation and prediction systems based on

simulation results. They incorporated the weather adjustment factor proposed by

Hranac et al. (2006) to represent the inclement weather impact on traffic operational

parameters. Tsirigotis et al. (2011) incorporated weather and traffic mix (speed and

volume) as exogenous variables in short-term freeway speed forecasting models and

investigated their effects on the predictability of traffic speed using several ARIMA

models. Results indicated that including exogenous variables only marginally

improves prediction performance, while modeling innovations such as Vector and

Bayesian estimation improve the models significantly.

2.4 Conclusion and Contribution

This chapter reviewed the related research in the traffic prediction models including

regression models, time series models, Kalman filter, neural network and nearest

neighbor non-parametric models. Linear regression models require that the

explanatory variables are statistically independent while it is common that many

variables in traffic network are highly correlated (volume, speed, density). ARIMA

models employ the internal relationships obtained from historical data, however, large

variations in the historical dataset would generate significant prediction error. In addition, most traffic systems exhibit nonlinear relationships, which make it difficult for linear models to capture the stochastic characteristics. In contrast, Kalman filter and neural network models do not require a predefined traffic pattern. However, for the parametric models such as ARIMA and Kalman filter models, as well as some non-parametric model such as neural network models, there remain some limitations in real time prediction. These include: complicated process of estimating parameters, long training time, over fitting problems, and difficulty to transfer. Also, for both time series, linear regression or Kalman filter models, when sudden changes in the traffic stream occurs, the time lag problem is generated. All regression models demonstrate some degree of lagging effects between the observed and predicted values which means the regression models are not able to adapt quickly to the changes in the real-time traffic condition without a point of reference. In general, statistical models have good performance on recurring traffic but may not be satisfactory for non-recurrent traffic.

Non-parametric models are adapted to work under stochastic conditions under the assumptions that a large and sufficient historical database is available. But the computation burden always has to be considered when using these models for real time applications. The input variables as well as the weighing factors need to be selected for optimal performance, which also adds to the computation burden. Research also concludes that weather has significant impact on traffic streams and many researchers quantified the impact of weather on both traffic demand and supply.

However, the area of incorporating weather impact in short term travel time prediction is less studied.

Accurate travel time prediction under various weather conditions would help users to make better decisions and respond to the trip requirements under adverse weather condition well. This research would focus on two parts: one is to develop an accurate and reliable prediction model; and the other is to study the weather impact on travel time and incorporating the weather information in predicting travel time under various weather conditions.

Chapter 3: Problem Statement and Prediction Models

3.1 Problem statement

3.1.1 Problem description

This research will focus on short-term travel time prediction for freeway segments. In this research, both traffic data and weather information from multiple sources will be utilized. An integrated non-parametric model is proposed to predict short term travel times based on a large historical traffic and weather information dataset, along with the available real time traffic information. Multiple freeway segments are selected covering various dynamic traffic characteristics that exhibit both recurrent and non-recurrent congestions. The performance of the model is tested under various weather scenarios. Travel time prediction up to 30 minutes ahead of time is also added to the integrated model.

In this research one source of traffic data is those obtained using Bluetooth sensors deployed on freeway segments to sample travel time of vehicles on freeways. To process Bluetooth data, four steps are carried out: a) Bluetooth Data Collection; b) Data Selection: study segments, time periods, traffic variables; c) Data Repair: recognition and repair of missing data and abnormal data; outlier filtering and fixing missing data by data extrapolation; d) Data Aggregation: generating aggregate travel time data at every 5 minutes interval level.

Consumer electronics are finding an ever-increasing role in our everyday lives. A majority of these devices in recent years are equipped with a point-to-point networking protocol commonly referred to as Bluetooth. Bluetooth technology is the primary means that enables hands-free use of cell phones. Bluetooth enabled devices can communicate with other Bluetooth enabled devices anywhere from one meter to about 100 meters. This variability in the communications capability depends on the power rating of the Bluetooth sub-systems in the devices. The Bluetooth protocol uses an electronic identifier, or tag, in each device called a Media Access Control (MAC) address. The MAC address serves as an electronic nickname so that electronic devices can keep track of who's who during data communications. In principle, the Bluetooth traffic monitoring system calculates travel times by matching Bluetooth MAC addresses at successive detection stations. Bluetooth data provides travel time and space mean speed directly with a relative high accuracy compared with most existing conventional detection techniques, and is also able to derive OD measurements. More details on using Bluetooth sensors for freeway travel time data collection is discussed in Haghani et al. (2010).

Other traffic data sources used in this study are from INRIX speed data for the average space mean speed and traffic counts on freeway segments from Maryland State Highway Administration (SHA) sensors. Weather data is also obtained from SHA, including the precipitation type, precipitation intensity, visibility and wind speed.

3.1.2 Research Objectives

The objective of this research is to utilize traffic and weather data from multiple sources and to develop an integrated traffic prediction model to predict travel times under various weather conditions, and especially severe weather conditions.

In the first part of the study, several prediction models that are proved to work efficiently in the area are selected and implemented using Bluetooth travel time data to perform prediction. Then, a modified non-parametric model KNN-T is proposed that incorporates the pattern feature from the traffic data in order to enhance the traditional KNN model with trend adjustment. The prediction results obtained from each model are compared and discussed. In the second part, an integrated prediction model is proposed that incorporates the impact of weather condition on the traffic stream. Then, this integrated model is enhanced, adding the features to perform travel time prediction on longer freeway path, as well as the multi-step ahead of time predictions. The research target is to predict travel time under different, especially adverse weather conditions through various sources of both traffic and weather data.

3.2 Prediction Models

3.2.1 Historical Average model

The historical average method uses an average of the past travel times to forecast future travel time for each time interval. This naive model is formulated by finding the historical average travel time for each time interval on each segment. At time interval (t), the predicted travel time at time interval (t+1) is estimated from the average of the previous historical travel times at the corresponding time intervals.

This model can be very easily applied by just computing an average from the segment's historical travel times and refined continuously by updating the historical average when new data become available and added. This model depends heavily on the repeatable nature of the traffic flow and thus is unable to capture the sudden changes in the system such as incident occurrence, severe weather and special events.

3.2.2 ARIMA (Autoregressive Integrated Moving Average) Model

Time series models have been applied to predict the future data points based on the trends and variations from the previous data points observed. ARIMA model is a generalization of ARMA model and applied under the condition where data points exhibit non-stationary characteristics (upward or downward trends). The ARIMA model combines the autoregressive model and moving average model which is generally represented as ARIMA (p, d, q) where p, d, and q are integers referring to the order of the autoregressive, integrated (the number of times the time series is differentiated), and moving average parts of the model respectively. The model is written as:

$$X_t = c + \sum_{j=1}^{p} b_j X_{t-j} + \sum_{j=1}^{q} a_j e_{t-j} + e_t \tag{2}$$

For the selection of the best p, q combination, Akaike's Information Criterion (AIC) is applied. The approach to estimate the value of p and q is by testing different combinations of p and q and check for the lowest AIC where AIC is defined as:

$$AIC(p, q) = -2 \ln L\left(\hat{\varphi}, \hat{\theta}, \hat{\sigma}^2\right) + 2(p + q + 1) \tag{3}$$

ARIMA model development is conducted in three steps: (1) model identification: where p, d, and q are estimated from the autocorrelation function and partial autocorrelation function of the time series, together with the Akaike's Information Criterion (AIC) applied for the selection of the best p, q combination; (2) parameter estimation: where the coefficients can be estimated from least square estimation (LSE); and, (3) model analysis and validation through prediction results.

These linear based time series models (AR, MA, ARMA, ARIMA) mainly predict the mean values and often fail to deal with large variations under some congested patterns or incidents. ARIMA models usually display a lagging effect between the predicted and observed travel time because they cannot promptly adapt to the change in recent time interval without reference to the recent traffic conditions. As missing data is always expected due to the failure of detector, interruption in communication or other types of malfunction, ARIMA are usually not quite suitable for wide application in the real traffic system since it requires continuous and stationary data series, which is not practical to obtain especially for online travel time prediction when data are updated every five minutes. Also, models have not been successfully applied to trips consisting of several links due to the complexity in multiple time-series datasets.

3.2.3 Kalman Filtering model

The Kalman filter is composed of a set of mathematical equations providing an efficient recursive approach to estimate the state of a process while minimizing the mean of the squared error. The Kalman filter became famous for its featured power to

support estimations of the past, present, and future states without knowing the precise nature of the system. Kalman filter algorithm is applied in travel time prediction since it allows the prediction of state variable (travel time) to be continuously updated when new observations become available.

When applying Kalman filter in travel time prediction, the equations turn into:

State equation: $\qquad\qquad\qquad x_t = \varphi_{t-1} x_{t-1} + w_{t-1}$ $\qquad\qquad\qquad\qquad$ (4)

Observation equation: $\qquad z_t = x_t + v_t$ $\qquad\qquad\qquad\qquad\qquad\qquad$ (5)

Where x_t is the predicted average travel time in time interval t; φ_{t-1} is the state transition parameter matrix describing the relationship between travel time of the current and previous time interval; z_t is the observed average travel time in time interval t; w_{t-1} and v_t are white noise terms indicating the process noise and measurement noise respectively.

The Kalman filter estimates a process through a feedback control where the filter estimates the state at some time and obtains feedback in terms of noise measurements. The procedure is conducted iteratively in two steps: the prediction step and the correction step:

Prediction Step:

$\hat{x}_t^- = \varphi_{t-1} \hat{x}_{t-1}$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (6)

$P_t^- = \varphi_{t-1} P_{t-1} \varphi_{t-1} + Q_{t-1}$ $\qquad\qquad\qquad\qquad\qquad\qquad$ (7)

In these equations, P_t^- is the a priori prediction error covariance and P_t is the posteriori prediction error covariance.

Correction Step:

$$K_t = P_t^-(P_t^- + R_t)^{-1} \tag{8}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - \hat{x}_t^-) \tag{9}$$

$$P_t = (I - K_t)P_t^- \tag{10}$$

K_t is defined as the gain which minimize the posteriori error covariance P_t.

The predictor equations projecting forward the current state and error covariance estimates to obtain the a priori estimates for the next time step. The corrector equations give feedback by incorporating a new measurement into the a priori estimate to obtain an improved posteriori estimate.

One potential issue arises when applying KF to a long segment with large variations in travel times. Since actual travel times will be available only after the trip completion, KF may not have the actual data to update its parameters to contend with dramatic changes in travel time.

3.2.4 Non-Parametric model

Rooted from pattern recognition, the non-parametric regression has been rapidly developed and used over the past 30 years to contend with the limitations in parametric models. Its early application is in forecasting the rainfall runoff (Karlsson and Yakowitz, 1987) and it also has application in market prediction (Mulhern and Caprara, 1994). According to Altman (1992), non-parametric regression is a set of techniques for fitting a curve without making strong assumption about the true shape

of the regression function. The techniques are useful when there is not much prior knowledge about the true curve.

The essential of the non-parametric approach is to locate the current system state in a past time neighborhood with similar status and use the past situations in this neighborhood to estimate current state. The assumption for this model is the recurrent nature of traffic streams. The objective is to find the hidden relationship from the large database instead of from the model developer. It is stated that the nearest neighbor approach will result in an asymptotically optimal forecaster (Karlsson and Yakowitz, 1987). It means that for an input state containing m values, the nearest neighbor will asymptotically be at least as good as any m^{th} order parametric model. The non-parametric model is not searching for an optimal result, but instead a sub-optimal or near optimal result for a satisfactory solution. This data driven heuristic technique can predict travel time through a large historical traffic dataset. The problem with a non-parametric model is that when sufficient good matches are not available in the historical database, the model may fail to generate a reliable prediction.

Parametric models mostly use statistical methods which involve the estimation of parameters. Non-parametric models work without reference to specific parameters and display several advantages over parametric models.

First, non-parametric models are less demanding for data. Parametric models usually involve certain underlying data distribution assumptions such as stationary data series or normal distributions which are not required for non-parametric models to achieve valid results. Second, for many cases, one need to select the key variables for parametric models and non-parametric models are sometimes able to provide a quick view with less calculation. Third, there are usually complicated interactions among data and certain types of patterns may exist. It is very difficult to model these interactions when the number of variables grows large. Fourth, non-parametric models provide results with practical means, which is easier to understand and would be useful when there are questions with the results from parametric model of certain artificial matrix and those results are not recognized as reliable.

However, non-parametric models need a large data size to draw a conclusion with confidence and since no parameters are estimated from the non-parametric model, it is hard to reflect the differences between samples quantitatively. Moreover, it is not easy to associate the non-parametric model results with confidence intervals.

This prediction problem consists of two types of variables and two sets of data. The first type of variable is the response variable or decision variable, which is the predicted travel time for the next time interval; the second type of variable is the explanatory variables, which are variables that are closely related or have potential relations to the response variable. More than one explanatory variable exist. The first dataset is the historical dataset and includes a continuous time period traffic data of traffic volume, travel time, weather condition and so on. The second dataset is the test

dataset or validation dataset which is used as a validation for the accuracy of the prediction model results.

To use the non-parametric model in travel time prediction, first we need to define the state vector and this definition should be appropriate both in sufficiency and simplicity. Some possible variables to define the state vector include the previous time interval travel times, traffic volumes, occupancies and speeds. The general methodology for the prediction can be concluded in the following steps:

Step 1: Build historical database: A representative and sufficient historical database is required for using the non-parametric model.

Step 2: Define Neighborhood: Quality of the neighborhood reflects directly on the accuracy of the prediction. There are two basic approaches for defining neighborhood: kernel and nearest neighbor. The kernel neighborhood has a fixed bandwidth (or radius) which indicates a fixed space. While the nearest neighbor neighborhood has a fixed sample size K, which indicates that each neighborhood has the same number of samples.

Step 3: Calculate distance (for nearest neighbor): Several distance calculation methods may be applied such as: absolute value distance, Euclidean distance and weighted Euclidean distance.

Step 4: Finding K (for nearest neighbor): Tests need to be conducted to find the best value of K.

Step 5: Define prediction function: several functions exist such as taking the average of the neighborhoods or the weighted average.

3.3 K Nearest Neighbor model

3.3.1 Basic K Nearest Neighbor model

The basic KNN prediction model studied in this dissertation follows the general

concept in traditional KNN models. The variables included to define the state vector

are the previous continuous time interval travel times. In this model, it is assumed that

the predicted time interval travel time is related to the previous time interval travel

times which are considered as a combined group and their nearest neighbors in the

historical records are found to predict the travel time for the targeted next time

interval. The total length of these grouped previous time intervals should be long

enough to represent the evolution of the traffic status but not exceeding that to avoid

inclusion of unnecessary or misleading information. The algorithm is described as

follows:

Step 1. Build a historical database with previous time interval travel times;

Step 2. Select T continuous previous intervals as a combined group, $t = 1...T$;

Step 3. Calculate and rank the neighborhood similarity to find nearest neighbors for

next interval (with smallest Euclidean differences) where:

$$\text{Dist}(x^h) = \sqrt{\sum_{t=1}^{T}[x_t - x_t^h]^2} \qquad (11)$$

where h is the sequence number of the historical data and x^h is the corresponding

travel time record in the historical data;

Step 4. Find a set of K nearest neighbors;

Step 5. Predict the targeted next interval travel time by taking the average of the

nearest neighbors:

$$\hat{x}_{T+1} = \frac{1}{K}\sum_{h=1}^{K}(x_{T+1}^{h}) \tag{12}$$

3.3.2 K Nearest Neighbor-Trend Adjustment Model (KNN-T)

In this research, a modified KNN model with trend adjustment is developed to include the traffic trends effects into the prediction model. Kim et al. (2005) proposed a pattern recognition technique which considers the past sequences of traffic flow patterns to predict the future states, overcoming the memory-less property of previous nearest neighbor non-parametric regression. This algorithm recognizes the traffic flow pattern by defining the flow change directions qualitatively, which is solely based on the signs of changes in the traffic volumes and results indicate that it is superior to the previous nearest neighbor non-parametric regression models.

In this model, the travel time trends will be considered both qualitatively and quantitatively to perform the travel time prediction. Compared to the previous research, in this proposed model, not only the signs of changes will be considered, but also the magnitudes of changes in travel times will be included. This KNN-T model considers the pattern recognition of the traffic streams that incorporates the trend adjustment feature into the traditional KNN model. It is designed to improve the prediction by capturing recurring traffic patterns. This model is composed of two parts: one part follows the traditional concept of a KNN model and the other part considers the trend effect of the travel times. The neighborhood similarity for the second part is calculated based on the square sum of the differences of each adjacent pair between the corresponding current and historical records. The prediction function

is also a combination of the two parts reflecting both the average of the nearest neighbor value and their differential values for the trend adjustments.

A simple example is used to explain the importance of considering the trend effect. Consider a series of travel times for current time intervals: (1, 3, 4) and two historical series of (3, 5, 6) and (4, 4, 3) respectively. Using both the Euclidean distance and trend (differential) distance to calculate the similarities for comparison, it can be seen from Table 1 that historical series 2 has a smaller Euclidean distance however the trend distance is larger than historical series 1. Figure 2 clearly indicates the necessity of including historical series 1 since it has the same trend with the current series which makes it a very good candidate in the neighborhood.

Table 1 Example for trend effect

Travel time	Interval 1	Interval 2	Interval 3	Euclidean distance	Trend distance
Current series	1	3	4	-	-
Historical series 1	3	5	6	12	0
Historical series 2	4	4	3	11	8
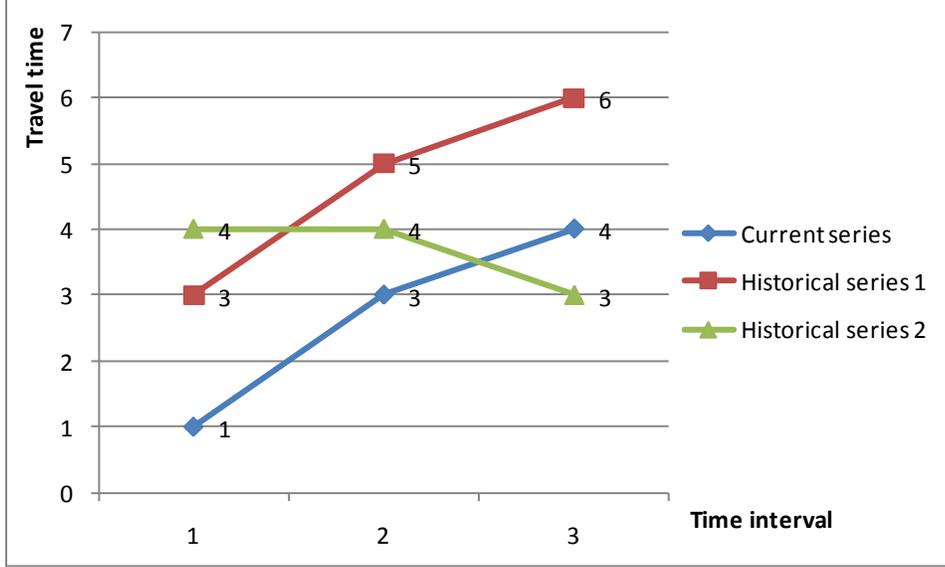
Figure 2 Example for trend effect

A weighted combination of both similarity schemes is used in finding the nearest neighbors and the optimal weight parameter α will be decided. The algorithm is described here:

KNN-T Model

Step 1. Build a historical database with previous time interval travel times;

Step 2. Select *T* continuous previous intervals as a combined group, t=1…T;

Step 3. Calculate and rank the neighborhood similarity to find nearest neighbors for next interval where:

$$Dist\ (x^h) = \alpha \sqrt{\sum_{t=1}^{T}[x_t - x_t^h]^2} + (1-\alpha)\sqrt{\sum_{t=2}^{T}[(x_t - x_{t-1}) - (x_t^h - x_{t-1}^h)]^2} \quad (13)$$

T is the total number of continuous time intervals included and h is the sequence number of the historical data;

Step 4. Find a set of K nearest neighbors with the weighted distance.

Step 5. Predict the targeted next interval travel time by taking the combined weighted

average of (1) next interval value of each nearest neighbor and (2) differential

value (between T and T+1) of each nearest neighbor where:

$$\hat{x}_{T+1} = \alpha \frac{1}{K} \sum_{h=1}^{K}(x_{T+1}^{h}) + (1-\alpha)[x_T + \frac{1}{K}\sum_{h=1}^{K}(x_{T+1}^{h} - x_T^h)] \qquad (14)$$

To find the best combination of parameters α, T and K to get prediction results with

higher accuracy, every combinations of these three parameters are tested for smallest

Mean Absolute Percentage Error (MAPE) within the reasonable range where

$T_{max} = 6$ and $K_{max} = 60$ and the optimal combination is generated for KNN-T

$(T, K, \alpha,)$.

3.4 Case study – Experimental Results for Model Comparison

3.4.1 Site description and Data

As discussed previously, Bluetooth data can provide travel time and space mean

speed directly with a relatively high accuracy compared with most existing

conventional detection techniques. The traffic data used in this case study is data

collected continuously using Bluetooth data collection devices.

The test location is one freeway segment selected from Virginia Route I-66 East

Bound ending at Exit 62 with 1.18 miles segment length and the available Bluetooth

data was collected from November 6th to November 13th, 2009. Raw data are filtered

and the aggregate Bluetooth average travel times are provided at every 5 minutes time

interval with outliers removed. For intervals missing travel time data (error or no

observations), the missing data is fixed through simple data interpolation. The Bluetooth data collected from Nov 6th through Nov 12th formed the dataset for model calibration. Data collected on Nov 13th are used for prediction validation.

3.4.2 Prediction error indices

Commonly used prediction error indices for validation of prediction results include: Root-mean-square error (RMSE), Root-relative-square error (RRSE), Mean absolute error (MAE), Mean absolute percentage error (MAPE), Mean absolute relative error (MARE). These error indices are calculated as follows:

(1) Root-mean-square error (RMSE):

$$\text{RMSE}(\%) = \sqrt{\frac{\sum_{t=1}^{N}(x_t - \hat{x}_t)^2}{N}} \qquad (15)$$

x_t is the real travel time observation at time interval t;

$\hat{x}_t$ is the predicted travel time at time interval t;

N is the total number of observations processed during the time interval provided; (number of samples)

(2) Root-relative-square error (RRSE): RSSE has the property of penalizing large prediction errors thus providing accurate assessment of model performance.

$$\text{RRSE} = \sqrt{\frac{\sum_{t=1}^{N}[\frac{(x_t - \hat{x}_t)}{x_t}]^2 x_t}{\sum_{t=1}^{N} x_t}} \qquad (16)$$

Or

$$\text{RRSE} = \sqrt{\frac{\sum_{t=1}^{N}[(x_t - \hat{x}_t)]^2}{\sum_{t=1}^{N}(x_t)^2}} \qquad (17)$$

(3) Mean absolute error (MAE):

$$\text{MAE} = \frac{\sum_{t=1}^{N}|x_t - \hat{x}_t|}{N} \tag{18}$$

(4) Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{100 \sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right|}{N} \tag{19}$$

(5) Mean absolute relative error (MARE):

$$\text{MARE} = \frac{\sum_{t=1}^{N}\frac{|x_t - \hat{x}_t|}{x_t}}{N} \tag{20}$$

The accuracy of the predictions results are usually expressed by RMSE and MAPE, which are the most widely used criteria. RMSE, which indicates the expected value of the error, is measured in the same units as the data and is representative of the size of a "typical" error. The mean absolute percentage error (MAPE) is also often useful for purposes of reporting, because it is expressed in error percentage terms and more understandable.

Every error index is calculated and listed for each prediction model and as in many other research studies the MAPE (given in equation 19) is used as the main error index for model performance comparison.

3.4.3 Experimental Results

For each model, travel time prediction for all day and morning peak hours (6:30am-9:00am) are conducted respectively. Figures 3 through 7 show the prediction results for each model.
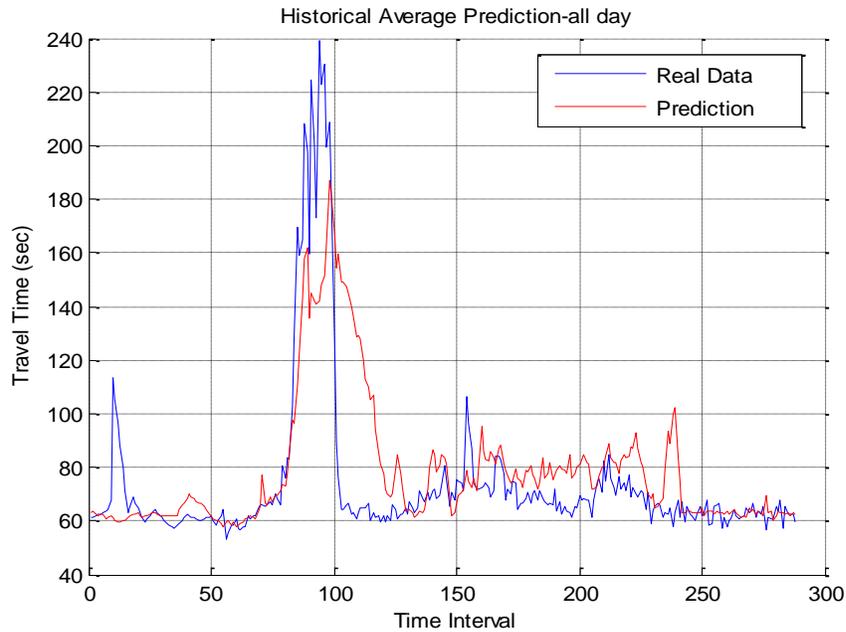
A. All Day prediction
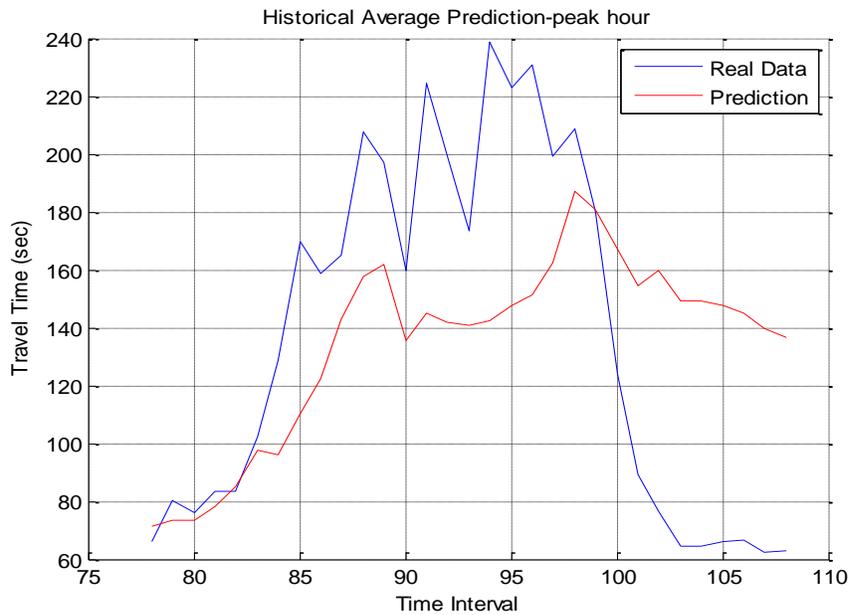
B. Peak Hour Prediction

Figure 3 Historical Average model - all day and peak hour prediction
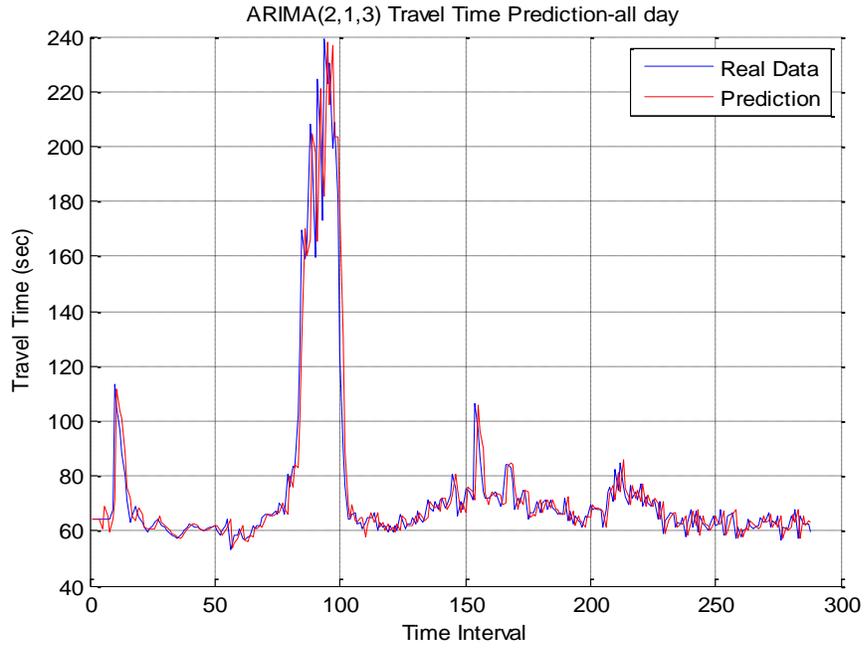
A. All day prediction
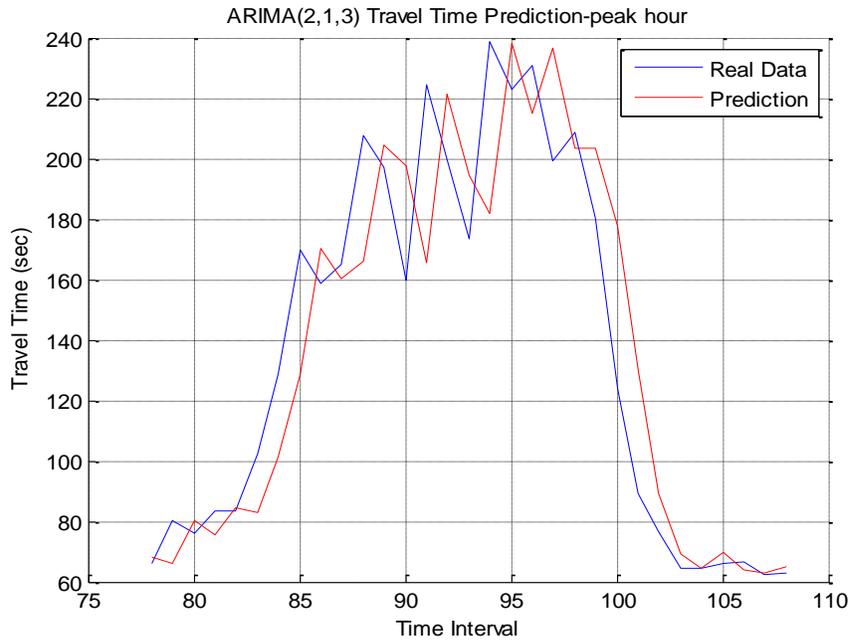
B. Peak Hour Prediction

Figure 4 ARIMA (2,1,3) model- all day and peak hour prediction
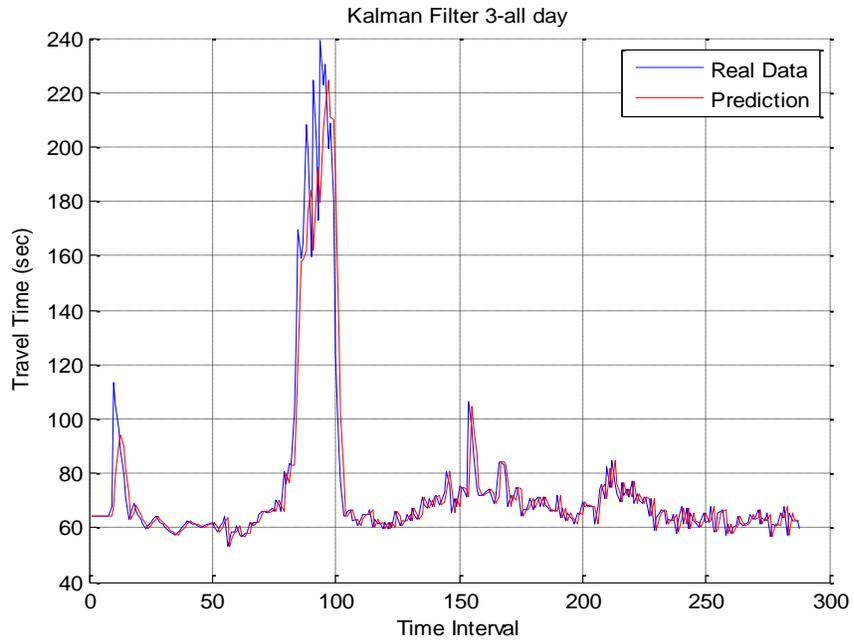
A. All day prediction
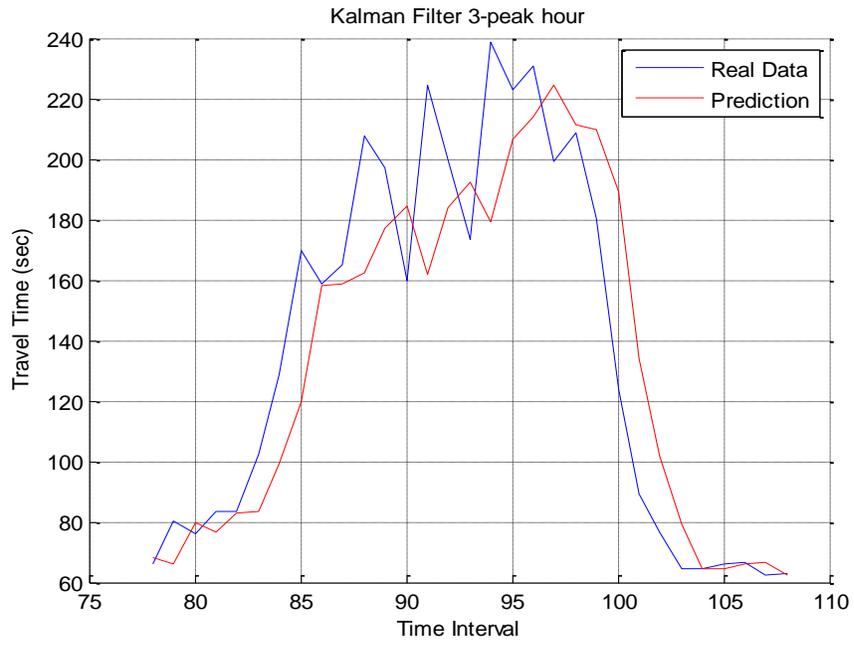
B. Peak Hour Prediction

Figure 5 Kalman Filter model-all day and peak hour prediction

43
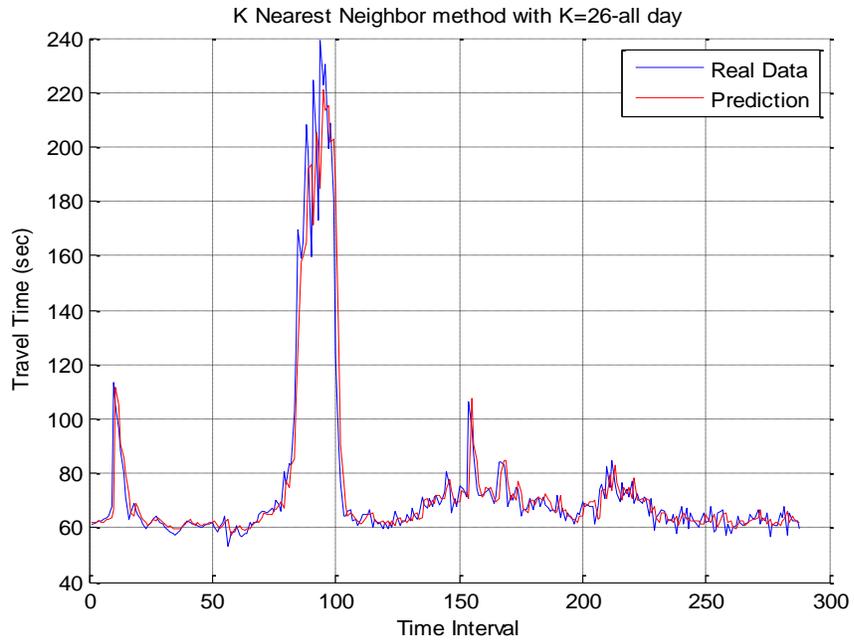
A. All day prediction
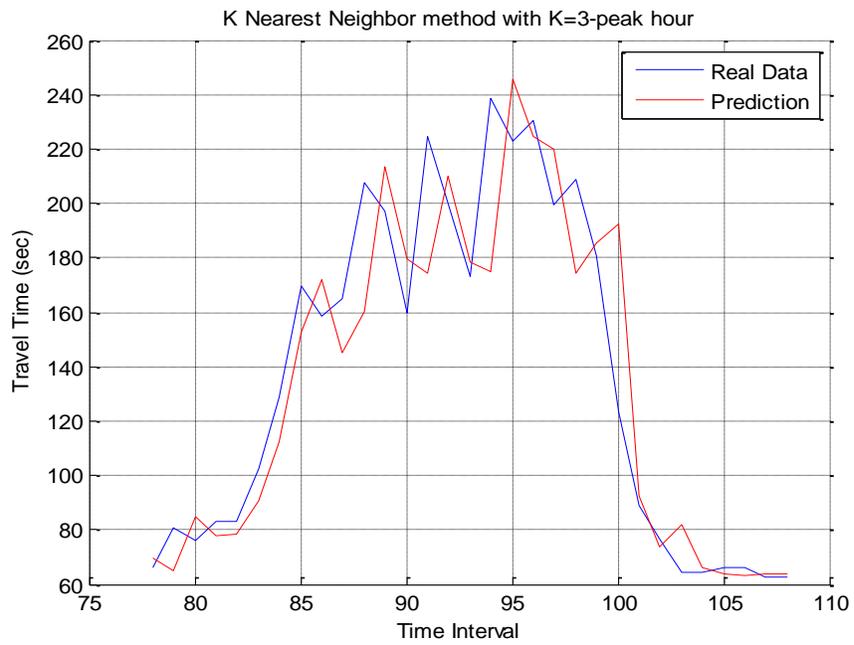
B. Peak Hour Prediction

Figure 6 KNN model – all day and peak hour prediction
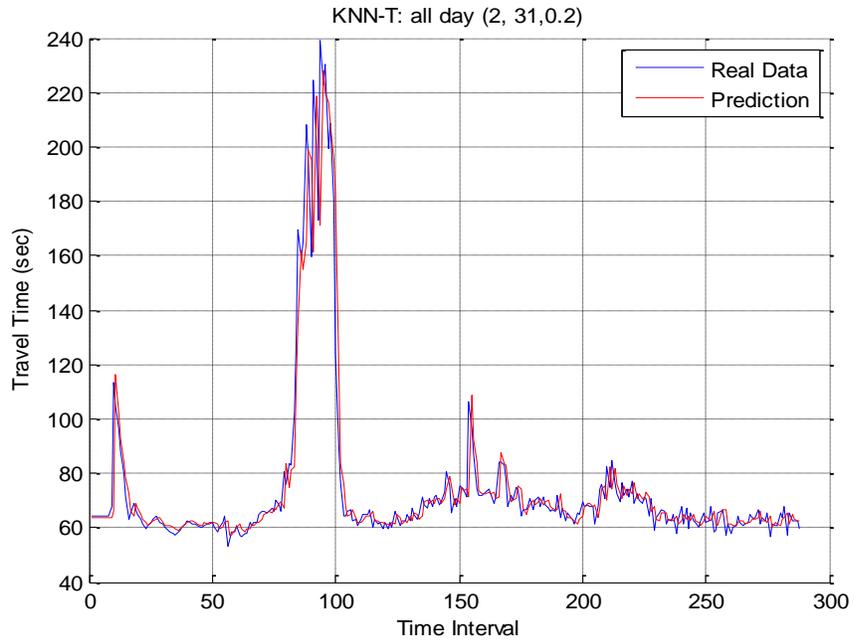
A. All day prediction
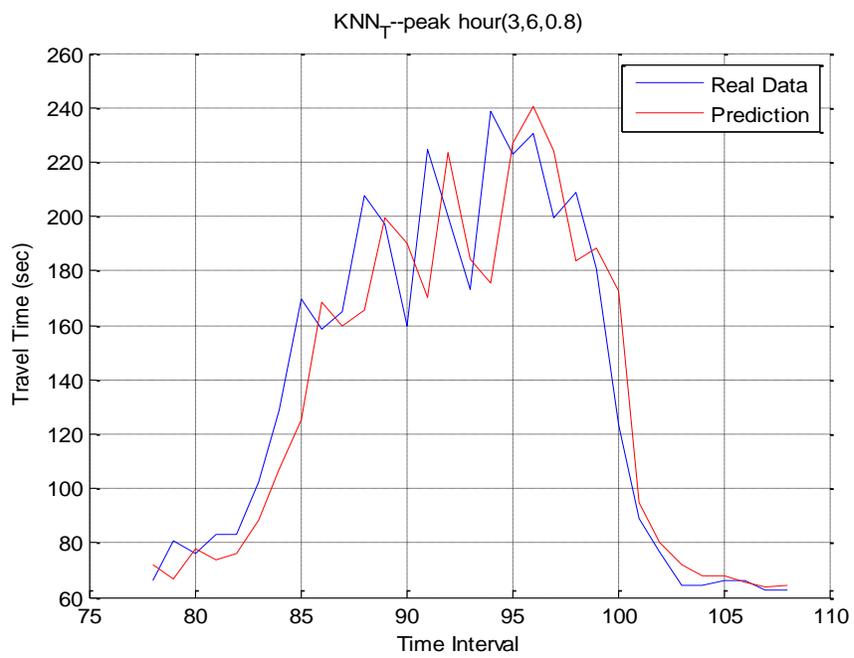
B. Peak Hour Prediction

Figure 7 KNN-T model – all day and peak hour prediction

Table 2 gives a complete comparison of the prediction error indices for each model.

Table 2 Prediction results for all models

Model prediction results for all day period					
Model	Historical Avg	ARIMA (2,1,3)	Kalman Filter	KNN (2, 26)	KNN-T (2,31,0.2)
RMSE	22.8573	10.0485	10.6593	9.5652	9.7404
RRSE	0.2854	0.1254	0.1331	0.1195	0.1216
MAE	12.7659	5.131	5.1777	4.5098	4.4564
MAPE	16.2619	5.8037	5.8011	5.081	**4.9568**
Model prediction results for peak hours period					
Model	Historical Avg	ARIMA (2,1,3)	Kalman Filter	KNN (2, 3)	KNN-T (3, 6, 0.8)
RMSE	55.7254	26.2357	28.0746	24.4595	23.6559
RRSE	0.3718	0.175	0.1873	0.1632	0.1578
MAE	46.4152	19.2323	20.2358	16.7779	16.3059
MAPE	44.2092	13.1143	14.2943	11.3184	**10.7796**

Table 2 indicates that the non-parametric models - KNN and KNN-T outperform the other models (Historical average, ARIMA and Kalman filter models) in this case study. Historical average model gave the least satisfactory performance especially for the peak hour period. This is as expected and is due to its dependency on a repeatable traffic pattern and inability to capture the dynamic nature of the traffic characteristics. ARIMA and Kalman filter models exhibit similar performance under both all day and peak hour periods. As can be observed from Figure 4, ARIMA model prediction results display a time lag between the predicted and observed travel time and large variations during peak hours since ARIMA model requires continuous and stationary series of data which is not obtainable from the dynamic and unstable traffic system. In Figure 5, Kalman filter could not provide satisfactory results for the peak hour period either. Since actual travel times are available only after the trip completion, the

actual data is not available to update the parameters of KF to contend with the dramatic change in travel time.

The non-parametric models, KNN and KNN-T both display better performance over the ARIMA and Kalman filter models by decreasing the MAPE over 10% for all day period and 20% for peak hour periods in prediction accuracy. This indicates when sufficient historical data are available, non-parametric models have the potential to provide much better prediction accuracy without going through the complicated model calibration and computations that are required for the ARIMA and KF models. The KNN-T model proposed in this study decreased the MAPE of traditional KNN model by approximately 2.5% for all day period and 4.8% for peak hour period. This indicates that studying the trend effects on travel time patterns has the potential to improve the prediction accuracy.

To compare the KNN-T model more clearly with the traditional KNN model, Table 3 lists the model performance under the same parameter of T and K with different values of α. Note that when α =1, the KNN-T model is equivalent to the traditional KNN model and when $\alpha = 0$, the KNN-T model only considers the trend effect. The results show that the best prediction accuracy comes from KNN-T model by using $\alpha = 0.2$ for all day period and $\alpha = 0.8$ for peak hour periods. These results are consistent with the traffic characteristics since peak hours usually do not have a clear trend or pattern from the more unstable traffic flows.

Table 3 KNN and KNN-T model performance comparisons

Model prediction results MAPE for all day period			
Model	KNN-T (2, 31, 0.2)	KNN-T (2, 31, 1)	KNN-T (2, 31, 0)
RMSE	9.7404	9.8787	10.2392
RRSE	0.1216	0.1233	0.1278
MAE	4.4564	4.6102	4.9672
MAPE	**4.9568**	5.1815	5.5943
Model prediction results MAPE for peak hours period			
Model	KNN-T (3, 6, 0.8)	KNN-T (3, 6, 1)	KNN-T (3, 6, 0)
RMSE	23.6559	24.5067	24.5003
RRSE	0.1578	0.1635	0.1634
MAE	16.3059	17.5702	19.5433
MAPE	**10.7796**	11.8195	13.8398

Chapter 4: Integrated prediction model incorporating weather impacts

The traffic stream on a road segment can either be in stationary or non-stationary status. Travel time is closely related to the flow and speed. In the stationary status, according to Daganzo (1997), there should be a relationship between speed and flow that will be a property of the road characteristics (number of lanes, geometry), weather conditions and population of vehicles. This is based on the hypotheses that one can reasonably expect drivers to do the same on average under the same average conditions. For the non-stationary status, we also need to consider the incident and roadwork impacts to determine the travel speed.

An accurate travel time prediction model should take into account both statuses. Based on the previously proposed KNN-T model with trend adjustment features, the traffic volume, weather condition especially severe weather conditions, and incidents occurrence will also be added to the integrated model.

4.1 Weather Impacts on Travel Time

4.1.1 Importance of studying weather impact on traffic

Accurate information regarding changes in the weather conditions is critical for the transportation system to remain safe and efficient. Weather is the second largest cause of non-recurrent traffic congestion, according to FHWA (2010). It accounts for 15 percent of all congestion in the United States and 25 percent of all non-recurrent delays. About one billion hours are lost each year due to delays caused by snow, rain,

ice, wind, and fog. 25 percent of total crashes involved weather. According to FHWA (2010), between 1995 and 2008, an average of 7,400 people were killed and over 629,000 were injured in weather-related crashes each year. The estimated annual economic cost of these deaths and injuries is $42 billion. Also, state and local agencies spend over $2 billion per year on snow and ice removal.

Congestion occurs when demand approaches or exceeds the road capacity. The affecting factors for the operational capacity of a roadway segment are most commonly summarized as "the seven sources of congestion", according to FHWA report (2005) "*Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation.*" These sources are: traffic incidents; weather; work zones; fluctuation in demand; special events; traffic control devices; and bottlenecks / inadequate base capacity. Among the above factors that would affect highway performance, incorporating the weather impacts into traffic prediction models is an important and challenging task. Both qualitative and quantitative effects of weather on traffic stream need to be understood.

Federal agencies have put great effort in developing reliable road weather management system to save lives, time, and money from inclement weather. The Office of Operations of Federal Highway Administration (FHWA) in conjunction with the Intelligent Transportation Systems office of the Research and Innovative Technology Administration (RITA) developed the Road Weather Management

Program (RWMP) to address road weather challenges through research, technology development, community outreach and promotion of strategies and tools.

4.1.2 Weather impacts on Demand and Supply

Back in the early 1950s, Tanner (1952) had recognized that a high negative correlation exists between rainfall and traffic. During the past decades, there has been continuing increased interest in research on weather impact on traffic stream. Research studies related to weather impact on highway traffic mainly focus on two sides of the transportation network: the demand side and supply side (Dong et al., 2010a). On the demand side, the inclement weather has impact both on changes in dynamic OD pattern (trip cancellation, departure time, and mode choice) and changes in traffic assignment in response to information and traffic controls (flow distribution on network, route choice decision). On the supply side, the inclement weather has impact on speed-density model on freeways and arterials; service rate; capacity; accidents and work zone related characteristics.

The Highway Capacity Manual (2000) states that in light rain, a 1.9 km/h reduction in speed during free-flow conditions is typical. In heavy rain, a 4.8 to 6.4 km/h reduction can be expected. Light snow has a statistically significant drop of 0.96 km/h in free-flow speeds and heavy snow results in a 37.0 to 41.8 km/h free-flow speed reduction.

4.1.3 Technologies implementation and benefits

State-of-the-art technologies and tools are used to mitigate the weather impact on roads such as: Environmental Senor Stations (ESS), freeway gate closure systems, wet pavement, fog, and high wind warning systems and integrated decision support systems. According to FHWA (2010), Utah DOT reported a saving of $2.2 Million in labor and materials for snow and ice removal from the statewide use of ESS with a benefit-to-cost ratio of 10:1. The automatic bridge anti-icing system reduced crashes by 64%. In Minnesota, the I-90 freeway gate closure system reduced the road clearance cost by 18%. In North Carolina, the wet pavement warning systems reduced crashes by 39%. In Denver Colorado, the anti-icing system reduced snow and ice related crashes by 14%. In northern Idaho, the anti-icing program reduced winter maintenance labor hours by 62% and decreased winter crash frequency by 83%. In Tennessee, the fog warning system reduced the fog-related crash number to only one between 1994 and 2003 on I-75 where over 200 fog-related crashes were recorded over a 20 year period. In Oregon, about 90% of motorist surveyed indicated they would slow down in response to messages displayed by automated high wind warning system. In Washington, 94% of travelers surveyed indicated a road weather information website made them better prepared and 56% agreed it helped them avoid delays. All of the above numbers and percentages are from the same report of FHWA (2010). All of these technologies and tools are efficiently applied to better respond to weather problems on roads.

4.1.4 Types of Weather Information

Weather data is usually updated every hour and does not fluctuate significantly by minutes. The weather related data can be obtained from the National Weather Services (NWS), National Climate Data Center (NCDC) and Clarus System. Federal funding has supported the Clarus (which is Latin for "clear") as a research initiative to develop and demonstrate an integrated surface transportation weather observing, forecasting and data management system, and to establish a partnership to create a Nationwide Surface Transportation Weather Observing and Forecasting System. The objective of Clarus is to provide information to all transportation managers and users to alleviate the effects of adverse weather (e.g., fatalities, injuries and delays). Clarus evaluates the benefit of real-time weather information and related pavement assessments (Pisano et al. (2008)). Clarus uses data from ESS that measure atmospheric conditions, pavement conditions, water levels and can include cameras, precipitation detectors, etc.

Various types of weather data are available, for example, Clarus gives the air temperature, dew point temperature, precipitation rate, precipitation intensity, humidity, visibility, surface temperature, wind and so on.

Based on the existing literature, there are several types of weather information that are used to study the weather impacts on traffic streams and the most frequently used are precipitation type, precipitation intensity, visibility and average wind speed. Their definitions are listed:

53

Precipitation type: Type of precipitation detected by a precipitation sensor, if one is available. It is indicated by Rain, Snow or other.

Precipitation intensity: Intensity of the precipitation as derived from the precipitation rate. The National Weather Service defines the following intensity classes: light, moderate, or heavy.

Visibility: Average distance that you can see, both day and night, computed every three minutes.

Wind speed: Average speed of the wind during a one minute period.

4.2 Model Description

4.2.1 Model Process

In this section, the proposed integrated model is presented with a description of each step in the modeling process. Each step describing the modeling approach is provided here.

Step 1: Historical Database Clustering/Classification:

1.1: The historical data records for each day are classified into three subsets: weekday (0), weekend (1) and holiday (2).

1.2: For each of the above three subsets, divide the records based on time intervals with incidents and without incidents.

1.3: Define 2 categories as: clear weather and severe weather (Rain or snow weather)

1.4: Under severe weather category, classify into rain and snow.

1.5: For each rain and snow group, divide into light, moderate, and heavy.

1.6: Under each weather type and intensity subgroup, classify into visibility>0.25 mile and visibility<=0.25 mile

1.7: Divide each subgroup into wind speed<=5 and wind speed>5.

Snow, fog or dark would cause visibility reduction, and according to OFCM (2002b) report, reduction in visibility under a quarter mile would decrease the driver's ability to see and be seen within a safe reaction distance. Winds speed greater than 25 mph can inhibit the maneuverability and stability of high profile vehicles (OFCM, 2002b). Stern et al. (2003) analyzed the weather impact on traffic flow in the metropolitan Washington DC area through regression methods and they used the criteria of 0.25 mile for visibility and 30mph for wind speed. This research will follow the criteria used in the above two references. Table 4 below summarized the features included.

Table 4 Category for weather features and incident

Category				
Precipitation type	Rain	Snow	Other	
	1	2	0	
Precipitation intensity	Light	Moderate	Heavy	Other
	1	2	3	0
Visibility	>0.25 mile	<=0.25 mile		
	0	1		
Wind speed	0-5	6-14		
	0	1		
Incident	YES	NO		
	1	0		

Figure 8 provides the framework for the data classification.

1st Layer	2nd Layer	3rd Layer	4th Layer	5th Layer	6th Layer
Day Features	Incident Features	Precip. Type	Precip. Intensity	Visibility	Wind Speed
Weekday	With incident	Rain	Light	>=0.25 miles	0-5
Weekend	Without incident	Snow	Moderate	<0.25 miles	6-14
Holidays		Other (Clear)	Heavy		
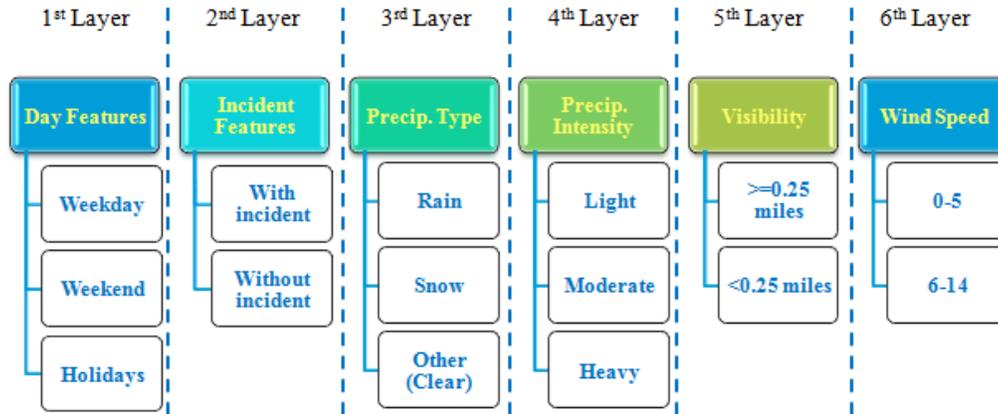
Figure 8 Framework for data classification

Step 2: Define Variables:

Notations:

Subscript:

h: index of historical records. Each record is composed of volume or travel time values over T continuous time intervals.

t: subscript for time intervals within the historical record dataset.

Historical records:

V_t^h: traffic volume in hth historical record at time interval t

TT_t^h: travel time in hth historical record at time interval t

Current measurement/estimation records:

V_t: current traffic volume at time interval t

TT_t: current travel time at time interval t

Parameters:

α, β, K: number of nearest neighbors

Variables:

x^h: binary variables that indicate the selection of hth historical record, $x^h = (0, or\ 1)$

Weather information notations:

Wtp: Precipitation type (rain, snow, other)

$Winten$: Precipitation Intensity (light, moderate, heavy, other)

$Wvis$: Visibility (miles)

$Wwind$: Average wind speed

Step 3: Calculate and rank the neighborhood similarity to find nearest neighbors

For the current estimation, the distance will be calculated in its neighborhood that is defined by the previous categories. The distance (similarity) between current estimation and hth historical record is calculated by:

$Dist^h =$

$$\alpha\sqrt{\sum_{t=1}^{T}\left(TT_t - TT_t^h\right)^2} + (1-\alpha)\sqrt{\sum_{t=2}^{T}\left[(TT_t - TT_{t-1}) - \left(TT_t^h - TT_{t-1}^h\right)\right]^2} +$$

$$\beta\sqrt{\sum_{t=1}^{T}\left(V_t - V_t^h\right)^2} + (1-\beta)\sqrt{\sum_{t=2}^{T}\left[(V_t - V_{t-1}) - \left(V_t^h - V_{t-1}^h\right)\right]^2} \qquad (21)$$

Note that both travel time and volumes are accounted for in calculating the neighbor distances. Normalization is used to balance the effect of each term. To normalize the travel time, the travel time distribution is studied and the value corresponding to the 95 percentile mark up is used. Then each travel time record is divided by that number. A similar approach is taken for normalizing the volume data. The case study in this

research uses 300 and 900 respectively for travel time and volume weights. The numbers will vary if other data sets are used but the approach will be the same. These numbers can be justified in the tests.

Step 4: Find an optimal set of K nearest neighbors

Objective function:

$$\min(\sum_{h=1}^{K} Dist^h x^h) \tag{22}$$

$$\text{s.t.:} \sum_h x^h = K \tag{23}$$

Step 5: Predict the targeted next interval travel time

Take the combined weighted average of (1) next interval value of each nearest neighbor and (2) differential value (between T and T+1) of each nearest neighbor:

$$TT_{T+1} = \alpha \frac{1}{K} \sum_h (TT_{T+1}^h) x^h + (1-\alpha)[TT_T + \frac{1}{K} \sum_h (TT_{T+1}^h - TT_T^h) x^h] \tag{24}$$

Step 6: calibrate K, T, α, β

4.3 Case Study for Integrated Prediction Model

In this section, a set of numerical experiments are conducted to evaluate the performance of the proposed model. One freeway segment is selected and the historical dataset covers a time period of 10 months including travel time, volume, weather and incident information. The model performance under different scenarios are tested and discussed.

4.3.1 Data Description

- Site description:

The study segment is the I-95 freeway southbound between MD-216 and MD-198. This freeway segment is about one mile long which covers one TMC and there are no on ramps or off ramps within this segment. This segment is from a main corridor and exhibits recurrent work day morning congestions. The travel time and traffic volume information are available for this target segment from a third party company INRIX and MDOT. There is a weather station located on this segment providing real time weather related information.

INRIX anonymously collects traffic speed data from personal trips, commercial delivery vehicle fleets and a range of other agencies and companies and compiles them into an average speed profile for most major roads. The speed data is collected from GPS devices installed in actual vehicles on the road, not from stationary detectors. Speed data comes in 1 minute intervals and is aggregated to 5 minutes intervals for this research. With the speed data, the readings are provided for actual segments of road instead of a single point, each measurement is labeled with a tmc_code, which identifies a specific stretch of the roadway. The file content includes: TMC code, measurement_tstamp (time stamp for each measurement), speed, average speed and reference speed. Travel time is calculated from speed which is space mean speed. The freeway segment selected here is:

ID: 110-04262   MD-198/EXIT 33 Southbound   Miles: 1.08919

Start Latitude, Longitude: 39.115394   -76.873651

End Latitude, Longitude: 39.104161   -76.887826

- Sensor Location (NAVTEQ detector, microwave radar detector)

The detector data file includes: detector ID, measurement_tstamp, volume and occupancy, the segment total volume is calculated by adding up all lanes' volumes.

Upstream Detector ID: 1121 (This Detector's volume is used since it is upstream and there is no outlet within this TMC segment).

Latitude, Longitude: 39.12402, -76.86595,

I-95 @ 0.23 Mile North of Stansfield Rd

Downstream Detector ID: 1134 (This detector's volume data is used to validate detector 1121.)

Latitude, Longitude: 39.10555, -76.88497,

I-95 @ 0.46 Mile North of Sandy Spring Rd/SR-198

- Weather station

The CHART weather station data files are created each night for the previous day for each weather station. All of the weather data files have the columns in the following order: Date / Time reported (GMT or UTC is 5 hours ahead of EST); Air Temperature; Relative Humidity; Dew Point Temperature; Barometric Pressure; Average Wind Speed; Wind Gust; Wind Direction; Precipitation Type; Precipitation Intensity; Precipitation Accumulation; Rate (rate per hour in inches); Visibility (miles); Surface Temperature; and Precipitation total for the past 24 hours.

Weather stations selected are:

US-29 @ Mid Paxutent   ID: 551053

Latitude, Longitude: 39.16754, -76.88445

I-95 @ Patuxent River   ID: 551006

Latitude, Longitude: 39.11469, -76.87425

- Federal Holidays (used as holidays in datasets)

In this study, the holidays identified in the database are federal holidays which are recognized by the United States federal government. There are 10 federal holidays in the year 2010.

01/01/2010	New Year's Day
01/18/2010	Martin Luther King, Jr. Day
02/15/2010	President Day
05/31/2010	Memorial Day
07/04/2010	Independence Day
09/06/2010	Labor Day
10/11/2010	Columbus Day
11/11/2010	Veterans Day
11/25/2010	Thanksgiving Day
12/25/2010	Christmas Day

- TMC and Detector location on Google map

  The study segment and the locations of the detectors are shown in Figure 9, as

  well as the location of the weather stations.
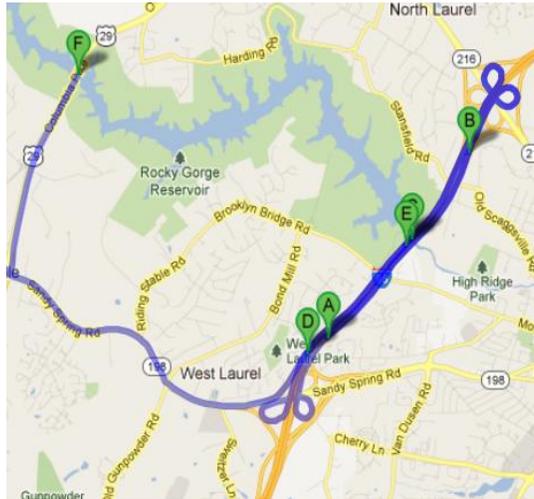
A: Downstream detector volume 1134

B: Upstream detector volume 1121

C: TMC starting point

D: TMC ending point

E: Weather station on I-95 (without visibility)

F: Weather station on US-29 (with visibility)

Figure 9 TMC and detector location on map

- Incident Info

Incidents occurred along the segment and close to the segment are selected by

location ID. The incident type, incident duration and other incident information are

provided. Figure 10 is an example of the 2010 incidents at this segment.

Type	Location	Time opened	Time closed	Duration	vehicles involved	max lanes closed
Collision, Personal injury	I-95 S at MD 198	05/28/2010 5:05	05/28/2010 5:58	53 minutes	3	0
Collision, Personal injury	I-95 S at MD 216	03/17/2010 9:26	03/17/2010 9:50	23 minutes	0	1
Collision, Personal injury	I-95 S at MD 198	06/07/2010 5:32	06/07/2010 6:21	49 minutes	2	3
Debris in roadway	I-95 S at MD 198	07/16/2010 9:10	07/16/2010 9:33	22 minutes	0	0
Disabled in roadway	I-95 S at MD 198	09/24/2010 6:24	09/24/2010 7:09	45 minutes	1	1
Vehicle fire	I-95 S at MD 198	03/23/2010 5:13	03/23/2010 5:28	15 minutes	1	0

Figure 10 2010 incident example

4.3.2 Historical Database

Historical database date range:

2010.01-2010.10

1. Speed data: in 5-minute interval of the average speed data. This is converted to travel time data by dividing distance by speed.

2. Volume data: in 5-minute interval.

3. Weather data: in 5-minute interval.

4. Incident data: location and duration. The capacity of a freeway segment is the total number of vehicles it can serve and may reduce due to accidents, work zone or other incidents occurrence. There are various forms of an incident such as: disabled vehicle, accidents, and planned road work. Incident duration is considered in this model.

4.3.3 Model Performance test – case 1- one week prediction

This section provides test results of case one, which is the prediction for a one week period. Five prediction models were used and compared including: Historical average model, ARIMA model, KNN model, KNN-T model and KNN-Integrated model. The performance measures used here are MAPE and the 5% error, which indicate the total number of time intervals that have accuracy of less than 5% MAPE in the predicted results. Table 5 lists the prediction results of the 5 models during one week period. Figures 11-15 show the difference of the real travel time data and the predicted travel times.

Table 5 Performance test – case 1 – one week prediction

Case 1	Historical database	01/01/2010-10/30/2010			
	Prediction Period	11/01/2010-11/07-2010			
Model	Historical Average	ARIMA	KNN	KNN-T	KNN-Integrated
Parameter	-	(3,1,2)	(3, 96)	(3, 96, 0.1)	(3, 96, 0.1, 0.2)
MAPE	8.3847	2.9462	2.9812	2.8377	2.9427
5% error range	1117/2016	1709/2016	1716/2016	1734/2016	1712/2016
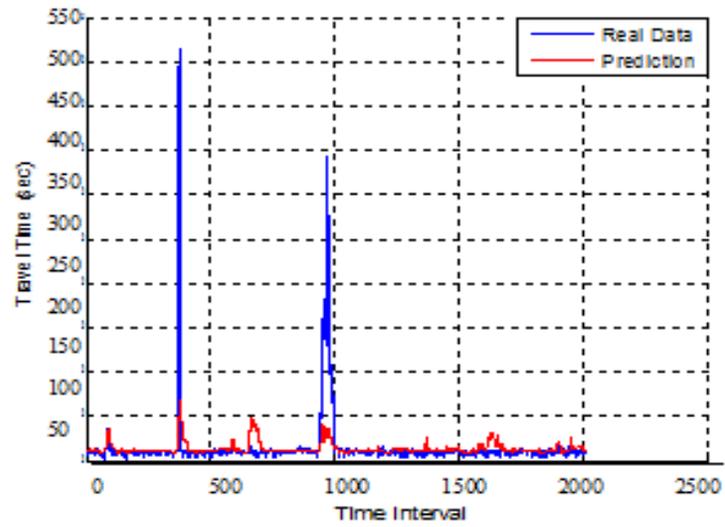
Figure 11 Historical average results - one week
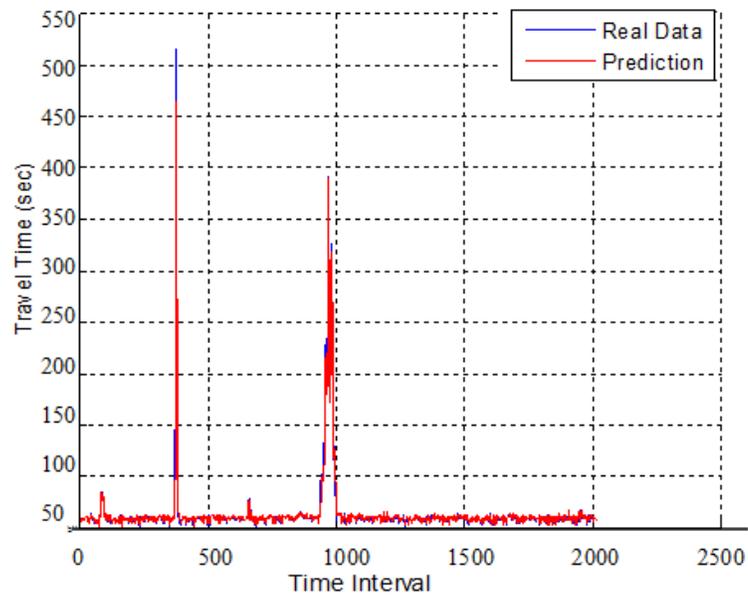
Figure 12 ARIMA results - one week

65
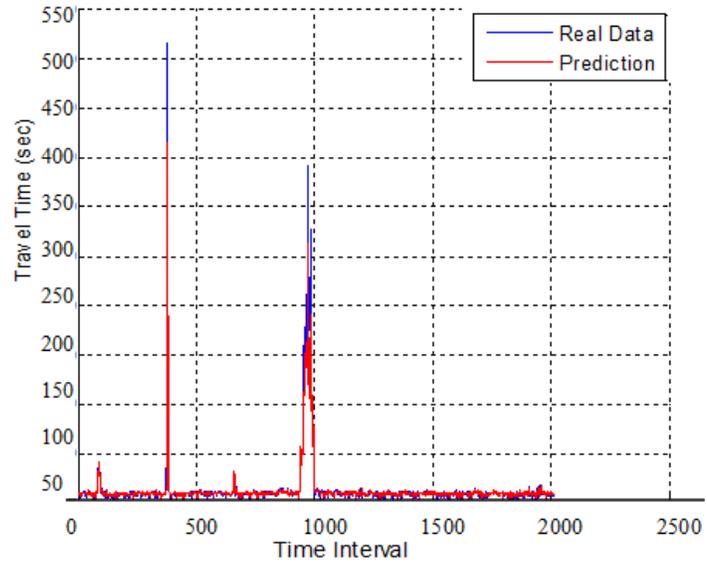
Figure 13 KNN results - one week
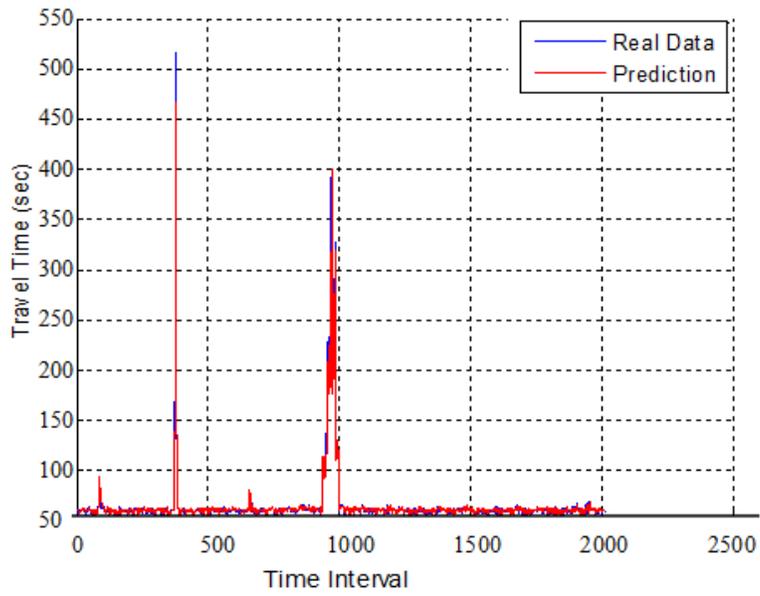
Figure 14 KNN-T results - one week
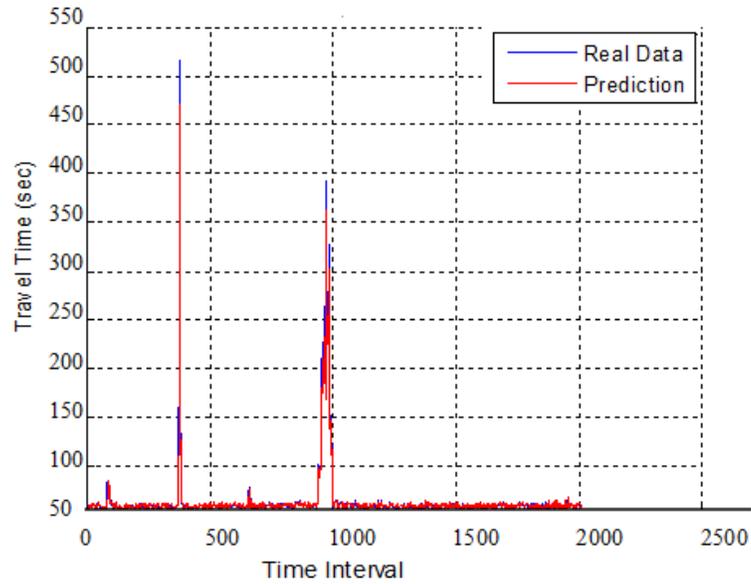
Figure 15 KNN-Integrated - one week

4.3.4 Model Performance test – case 2- one day prediction

This section provides test results of case two which is the prediction for a one day

period. Five prediction models were used and compared. The performance measures

used here are MAPE and the 5% error. Table 6 lists the prediction results of the 5

models during one day period. Figures 16-19 show the difference of the real travel

time data and the predicted travel times.

Table 6 Performance test – case 2 – one day

Case 2	Historical database	01/01/2010-10/30/2010			
	Prediction Period	11/01/2010			
Model	Historical Average	ARIMA	KNN	KNN-T	KNN-Integrated
Parameter	-	(3,1,2)	(3,47)	(3,47,0.1)	(3,47,0.1,0.9)
MAPE	7.3165	2.5667	2.6751	2.6597	2.6598
5% error range	150/288	248/288	248/288	245/288	251/288
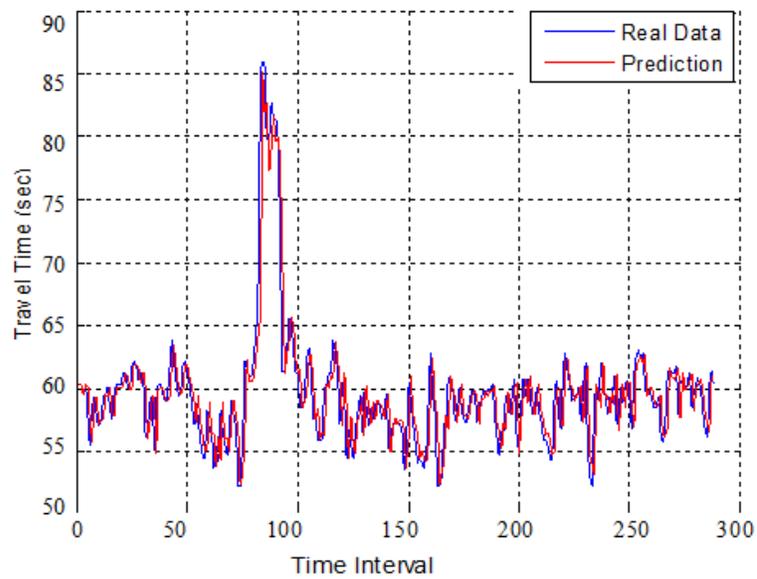
Figure 16 ARIMA results - one day
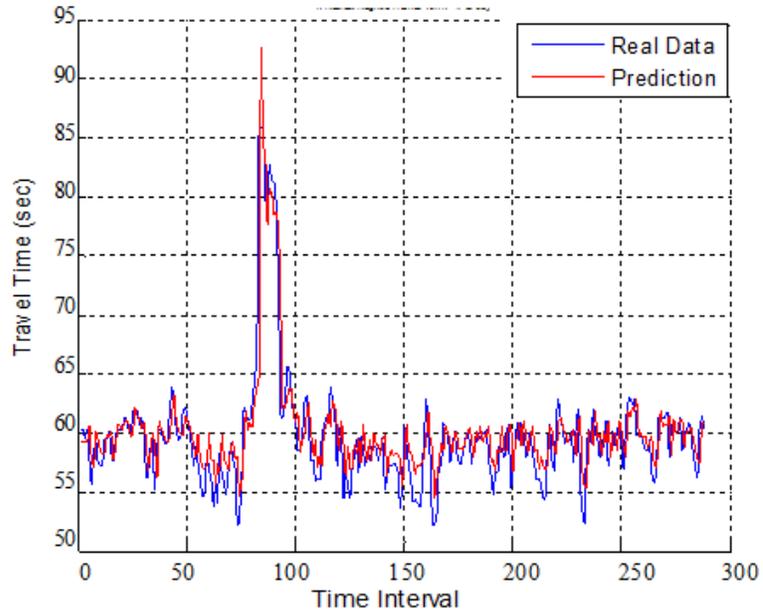
Figure 17 KNN results - one day
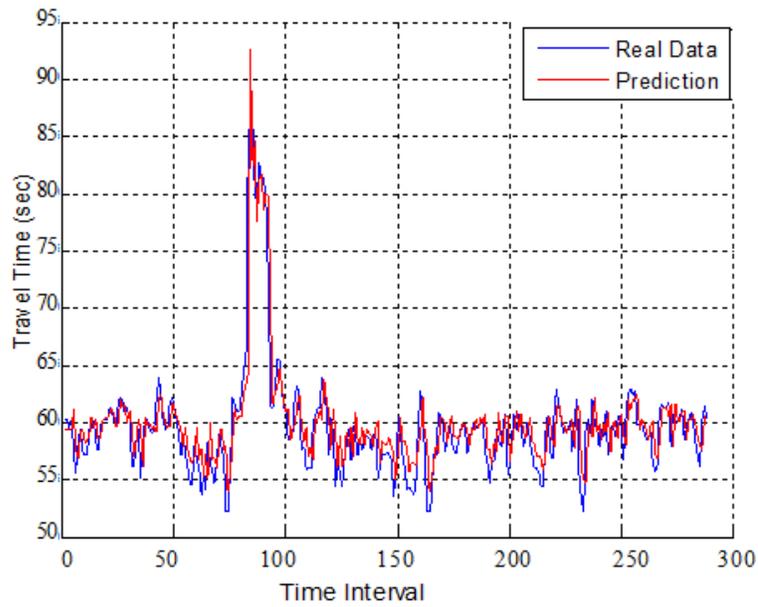
Figure 18 KNN-T results - one day
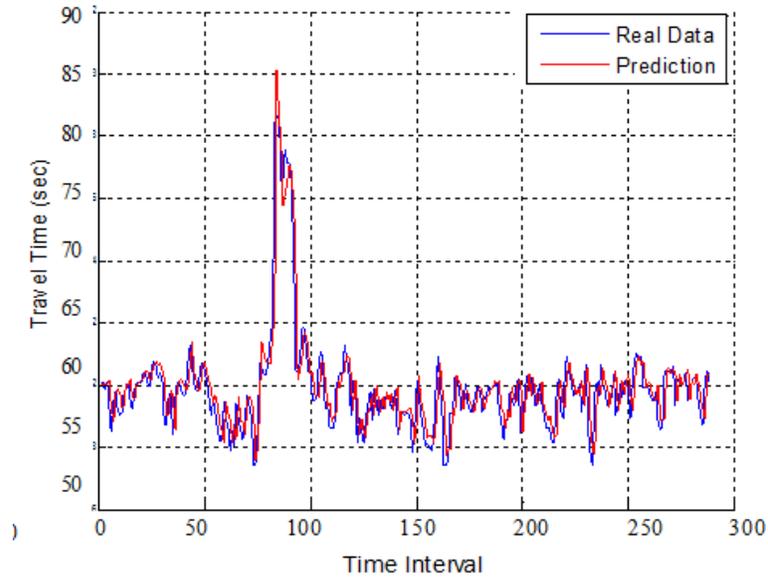
Figure 19 KNN-Integrated results - one day

4.3.5 Model Performance test – case 3- various weather conditions

To test the models' performance under various weather conditions, 19 days exhibiting

different weather features were selected and used. The days selected were all

weekdays without incidents with the purpose of focusing on the weather impact on

travel time. The four weather variables precipitation type, intensity, visibility and

wind speed were used to identify a specific weather condition. For example, in Table

7 below, for condition 17, 1311 means rain with heavy intensity, low visibility and

strong wind. The performance results are shown in Table 7 followed by the model

results comparison chart shown in Figure 20.

Table 7 Performance test – case 3- various weather conditions

Weather	Test Date	Precipi	Inten	Visib	Wind	KNN		KNN-T		KNN-Integrated		ARIMA (2,1,3)	
						MAPE	5% error	MAPE	5% error	MAPE	5% error	MAPE	5% error
1	11/01/2010	0	0	0	0	2.63	252	2.64	246	2.65	250	2.57	248
2	01/15/2010	0	0	0	1	5.87	230	5.78	230	4.38	224	5.87	229
3	03/03/2010	0	0	1	0	2.59	251	2.65	250	2.60	249	2.84	244
4	04/08/2010	0	0	1	1	3.52	220	3.50	223	3.33	229	3.59	228
5	02/03/2010	0	1	1	0	3.18	235	3.27	234	3.03	238	3.04	237
6	02/09/2010	0	3	0	0	4.69	204	4.70	205	4.40	213	4.33	217
7	03/26/2010	0	3	1	1	3.82	230	4.05	225	3.64	226	4.65	216
8	05/17/2010	1	1	0	0	3.29	237	3.35	235	3.22	234	3.65	225
9	03/02/2010	1	1	1	0	3.01	246	2.98	243	2.83	243	2.97	239
10	03/12/2010	1	1	1	1	3.00	238	3.05	233	2.77	247	2.79	243
11	06/07/2010	1	2	0	0	2.65	247	2.63	249	2.73	253	2.95	242
12	10/27/2010	1	2	0	1	4.40	233	4.29	232	4.66	225	5.62	219
13	09/30/2010	1	3	0	0	3.80	220	3.73	225	3.53	229	3.74	217
14	06/04/2010	1	3	1	0	2.17	260	2.15	262	2.34	255	2.30	258
15	05/03/2010	1	3	1	1	3.13	246	3.04	246	3.28	242	3.86	235
16	01/08/2010	2	1	0	0	3.11	236	3.03	238	3.03	239	2.96	239
17	02/01/2010	2	1	0	1	2.57	252	2.55	253	2.45	252	2.55	244
18	02/02/2010	2	2	0	0	4.08	219	4.14	215	3.84	224	4.16	214
19	02/05/2010	2	2	0	1	5.22	197	5.26	198	5.07	200	4.88	200
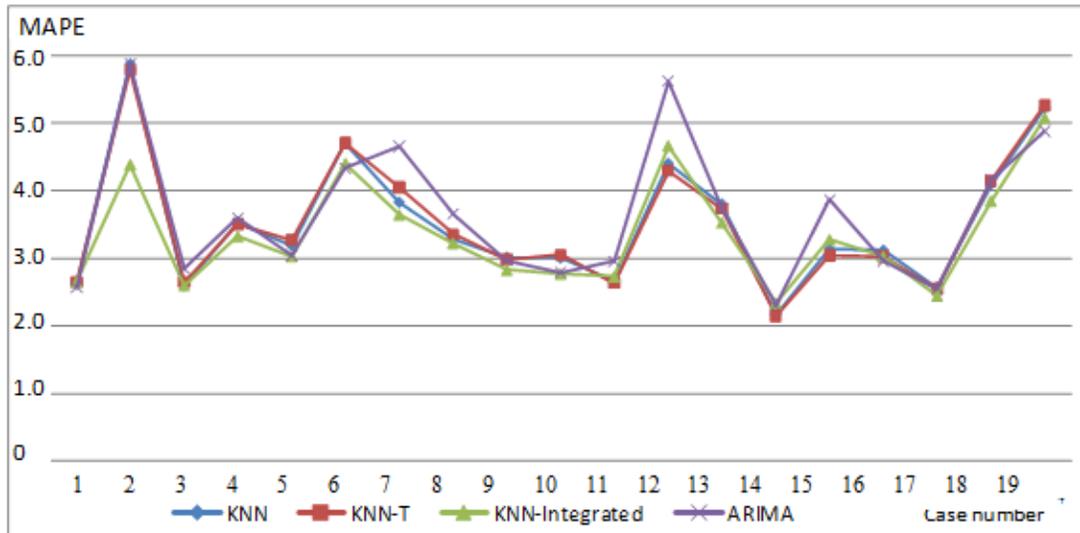
Figure 20 Model comparison chart

Four models were compared for their performances including: ARIMA model, KNN, KNN-T and KNN-Integrated model. From the test results in Figure 20, KNN-

71

Integrate model shows better accuracy for most cases. The average MAPE of the 19

days selected for KNN-Integrated model is 3.36% and for the ARIMA model, the

average MAPE is 3.65%. The KNN-Integrated model decreased the average MAPE

of ARIMA model by approximately 8.0%. To further investigate the model results, a

paired T-test was conducted for the model performance comparisons. 95% confidence

intervals were constructed comparing the KNN-Integrated model against the ARIMA,

KNN and KNN-T models. The results in Table 8 indicate that KNN-Integrated model

outperforms the others by passing every test.

Table 8 T-test results for model comparison

Model Comparison	P-Value	CI (α=0.05)
KNN-Integrated vs ARIMA	0.0048	-0.1169
KNN-Integrated vs KNN-T	0.0364	-0.0142
KNN-Integrated vs KNN	0.0380	-0.0122

4.3.6 Conclusion

The performance of the proposed model is evaluated through MAPE and the 5% error

range counts, which calculate the number of prediction values that fall into the range

of less than 5% error in accuracy. The model performance results indicate that

ARIMA and KNN models all performed well. Based on these preliminary results,

these models have similar performance under normal weather conditions. One week

and one day data under normal weather condition were used to test the models. For

the days under different weather conditions especially including inclement weather

conditions, the proposed KNN-Integrated model that incorporated weather impacts

performed better.

Chapter 5: Extension of the Integrated Model for Path and Multi-Step ahead Travel Time Prediction

In this chapter, the previously proposed travel time prediction model KNN-Integrated is extended with the ability to perform both path travel time prediction and multi-step ahead travel time prediction. Prediction performance from 5 minutes up to 30 minutes ahead of time are investigated and discussed. Results of the comparison of different models and their performance in the case study are provided as well as the sensitivity analysis on the efficient size of the historical dataset.

5.1 Model Description of the Extension of Integrated Prediction Model

In this section, the extension of the integrated model is presented with a description of the modeling approach. The extension of the integrated model is following the same modeling process as in the KNN-Integrated model while adding two features: path travel time prediction and multi-step ahead travel time prediction. The description of these two added features is provided in this section.

5.1.1 Path Travel Time Prediction

In the proposed KNN-Integrated model in Chapter 4, travel time prediction is calculated based on each individual freeway segment. To extend the model to a more general practice, in this section the path travel time is calculated for the continuous freeway segments based on each segment's individual travel times and backtracking computation procedures are used for this path travel time calculation.

Backward tracking procedure is applied in the path travel time calculation. For each time interval, the average travel times for each segment are calculated and stored in the database. Backtracking method is used based on these pre-computed travel times. The main concept is that starting from the last freeway segment, the travel time of each segment at its corresponding arrival time interval is used to calculate the path travel time. To be more specific, the following procedure is used for path travel time calculation.

For each freeway segment (link), the travel times are aggregated at 5 minute intervals. A matrix of these link travel times is then composed and denoted by $t_{i,k}$. The vertical dimension of this matrix represents the travel time at different time intervals and the horizontal dimension of this matrix represents the travel times of different links. Then, $t_{i,k}$ will be the average travel time of link i during time interval $[k\Delta, (k + 1)\Delta]$, where Δ is the aggregation interval of travel time (5 minutes).

Assuming the freeway corridor is composed of m links, and that $Y_i(x)$ denotes the average path travel time of vehicles that arrive at the end of link i at time x; then using the backtracking method, $Y_i(x)$ can be calculated from $t_{i,k}$ according to the following equation:

$$Y_i(x) = \begin{cases} t_{1,\delta(x)} & i = 1 \\ Y_{i-1}\big(x - t_{i,\delta(x)}\big) + t_{i,\delta(x)} & i > 1 \end{cases} \tag{25}$$

Where $\delta(x)$ is a function that maps continuous time variables x into the discrete time interval index. $\delta(x)$ is determined by the following equation:

$$\delta(x)\Delta \leq x \leq [\delta(x) + 1]\Delta, \quad \delta(x) \in Z^+ \tag{26}$$

74

According to the recursive definition given by Eq. (25), the path travel time is calculated based on the following algorithm written in pseudo code:

```
i=m; Ym=0;
while ( i>=1 )
do
        k=1;
        while (k<=T)
        do
                if kΔ ≤ t and t ≤ (k + 1)Δ
                break;
                k=k+1;
        end
        Ym = Ym + ti,k;
        t=t-ti,k;
        i=i-1;
    end
```

Following the above algorithm, the path travel times for each time interval are calculated and used as the input travel times for the extended KNN-Integrated model for the path travel time prediction.

5.1.2 Multi-Step Ahead Travel Time Prediction

In the KNN-Integrated model, only single step ahead travel time prediction (5 min) is provided for each freeway segment. To extend the model, multi-step ahead travel time prediction is provided which predicts travel time from 5 minutes up to 30 minutes ahead of time. Two methods are applied here: one-shot multi-step prediction and recursive multi-step prediction. Travel time prediction is performed at every five

minutes interval: 5 minutes, 10 minutes, 15 minutes, 20 minutes, 25 minutes and 30 minutes.

In the one-shot travel time prediction model, the predicted travel times for the next several time intervals are generated at the same time, which are obtained from the same set of nearest neighbors that were found in the model prediction for 5 minutes prediction interval. The predicted travel time is calculated from the weighted averages of this same set of neighbors through their subsequent time intervals' historical travel times.

In the recursive travel time prediction model, the predicted travel times for the next several time intervals are calculated sequentially, each from its individual nearest neighbor sets. The main concept is to use the predicted travel times of previous time intervals as input values for the neighbor search of the next time interval. When predicting the travel time of the next time interval, the predicted travel time from the previous time interval is considered as the real travel time input for the prediction of the next time interval. Then, a new set of nearest neighbors can be found for the prediction of each time interval in the future. Both of the one shot and recursive methods are used in the case studies and their performances are discussed in the next section. For the ARIMA model, similar recursive approach is used for the multi-step ahead prediction.

5.2 Case Study for the Extension of the Integrated Prediction Model

In this section, the results of a set of numerical experiments that are conducted to evaluate the performance of the extension of the integrated model are reported. One path from a freeway corridor is selected and the historical dataset covers a time period of 19 months including travel time, volume, weather and incident information. The model performance under different traffic and weather scenarios are tested and discussed.

5.2.1 Data Description

- Site description:

The study path is I-95 freeway northbound between MD-216 and I-895. This freeway path is about 9.6 miles long and covers eight TMC segments. This path is from a main corridor and exhibits recurrent work day afternoon congestion. The travel time and traffic volume information are available for this target path from INRIX and MDOT. There is a weather station located on this path providing real time weather related information.

Travel time is calculated from space mean speed in minutes. The eight consecutive TMC segments that are included in this path are listed in Table 9.

- Sensor Location (NAVTEQ detector, microwave radar detector)

The detector data file includes: detector ID, measurement_tstamp, volume and occupancy. The segment total volume is calculated by adding up all lanes' volume and the path volume is calculated by adding up all segments' volume and taking the

average volume. There are eight detectors that collect data along this freeway path and their location description is given in Table 10.

Table 9 TMC Location Description

	TMC ID	Location	Start_ latitude	Start_ longitude	End_ latitude	End_ longitude	Miles
1	110+04419	MD-32/EXIT 38	39.13	-76.85	39.16	-76.83	1.91
2	110P04419	MD-32/EXIT 38	39.16	-76.83	39.16	-76.82	0.80
3	110+04420	MD-175/EXIT 41	39.16	-76.82	39.17	-76.80	1.46
4	110P04420	MD-175/EXIT 41	39.17	-76.80	39.18	-76.78	0.86
5	110+04421	MD-100/EXIT 43	39.18	-76.78	39.19	-76.77	1.05
6	110P04421	MD-100/EXIT 43	39.19	-76.77	39.20	-76.76	0.91
7	110+04422	I-895/EXIT 46	39.20	-76.76	39.22	-76.72	2.33
8	110P04422	I-895/EXIT 46	39.22	-76.72	39.22	-76.72	0.27

Table 10 Detector Location Description

Location ID	Latitude	Longitude	Detector Location Description
1125	39.16	-76.82	I-95 @ 0.43 Mile North of SR-732
1124	39.18	-76.78	I-95 @ 0.42 Mile North of SR-175
1135	39.19	-76.77	I-95 @ 0.12 Mile North of SR-103
1143	39.21	-76.75	I-95 @ 0.86 Mile South of Montgomery Rd
53	39.21	-76.75	I-95 between MD 100 & Montgomery Rd
34	39.21	-76.74	I-95 @ Montgomery Rd
1114	39.22	-76.73	I-95 @ 0.92 Mile South of River Rd
1140	39.24	-76.70	I-95 @ 0.37 Mile South of Oakland Rd

- Weather station

The CHART weather station data covering the 19 months from January, 2010 to July, 2011 are used in this case study and the weather station located on this path is:

I-95 @ Howard/Baltimore County Line   ID: 551019

Latitude, Longitude: 39.23, -76.71

- Federal Holidays (used as holidays in datasets)

As the same in the previous case study, the holidays identified in the database are federal holidays recognized by the United States federal government. There are 20 federal holidays in the year 2010 and 2011, as listed below.

01/01/2010	New Year's Day
01/18/2010	Martin Luther King, Jr. Day
02/15/2010	President Day
05/31/2010	Memorial Day
07/04/2010	Independence Day
09/06/2010	Labor Day
10/11/2010	Columbus Day
11/11/2010	Veterans Day
11/25/2010	Thanksgiving Day
12/25/2010	Christmas Day
12/31/2010	New Year's Day
01/17/2011	Martin Luther King, Jr. Day
02/21/2011	Washington's Birthday
05/30/2011	Memorial Day
07/04/2011	Independence Day
09/06/2011	Labor Day
10/10/2011	Columbus Day
11/11/2011	Veterans Day
11/24/2011	Thanksgiving Day
12/26/2011	Christmas Day

- Study site location on map

The study segment and the locations of the TMCs are shown in Figure 21, as well as

the location of the weather station.
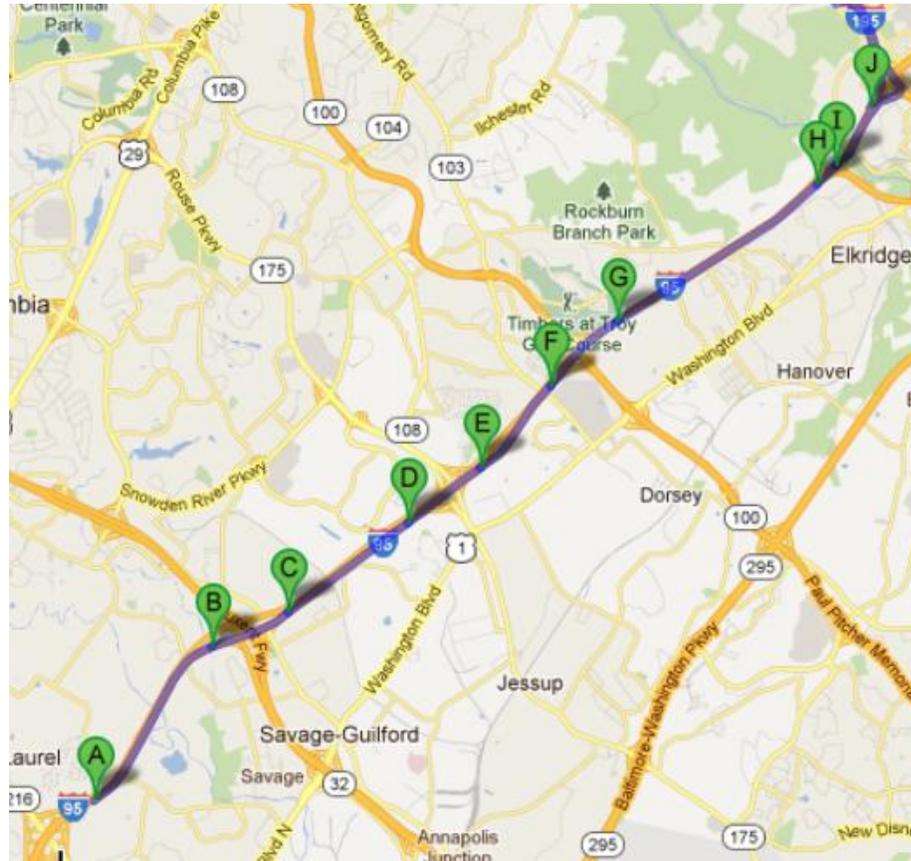
A – I: TMC starting and ending points   J: Weather station on I-95

Figure 21 Study site location on map

- Incident Info

Incidents occurred along the segment are selected and used as the data input for the

model. The incident type, incident duration and other incident information are

provided for this 19 months period by MDOT.

- Historical Database

Historical database date range: 2010.01.01-2011.07.26

1. Path travel time data: in 5-minute interval.

2. Path volume data: in 5-minute interval.

3. Weather data: in 5-minute interval.

4. Incident data: location and duration.

- Days Selected for Prediction

30 days from the database are selected for model performance test. Since the KNN-Integrated model is targeted to consider the weather impact on travel times, the 30 days selected cover all weather types as well as day and incident features. Weekday, weekends, holiday, incident, rain, snow, strong wind and low visibility conditions are all included in these selected days and these days are also the most representative days for each weather condition.

Travel time prediction is performed both for all day and peak hour periods. The freeway path selected is from a major corridor and it exhibits recurrent afternoon peak traffic conditions. The afternoon peak hour traffic pattern is studied for this path and based on the historical data, the peak hour period selected for this study is from 3:00 pm to 8:00 pm, for a 5 hour duration.

5.2.2 Model Performance test and comparison

This section provides the test prediction results for the 30 days selected. Four

prediction models were used and compared including: ARIMA model, KNN model,

KNN-T model and the extended KNN-Integrated model. The performance measure

used is Mean Absolute Percentage Error (MAPE).

(1) Model performance comparison between KNN-Integrated one shot and

KNN-Integrated recursive models

First, both one-shot and recursive multi-step travel time prediction are tested for the

KNN-Integrated model and the results are compared. 5 minutes to 30 minutes travel

time predictions are performed for each method. Tables 11 and 12 give the average

MAPE for 5 minutes up to 30 minutes prediction time interval for each day. Figure 22

shows the comparison results of the total average MAPE of all selected days over the

six time intervals between the KNN-Integrated one shot model and KNN-Integrated

recursive model.

Table 11 KNN-Integrated one shot prediction results for one day

Date	5 min	10 min	15 min	20 min	25 min	30 min
1/21/2010	2.983	3.822	4.598	5.572	5.864	6.584
1/25/2010	1.871	2.493	2.807	3.033	3.332	3.599
2/9/2010	2.490	3.954	4.531	5.289	6.061	6.730
2/10/2010	7.900	11.682	13.782	14.528	14.746	15.515
2/19/2010	7.210	12.053	16.087	19.141	21.371	21.872
3/12/2010	4.428	7.185	8.697	9.863	10.595	10.977
4/8/2010	2.024	3.047	3.550	4.136	4.831	5.650
5/31/2010	2.054	2.873	3.203	3.264	3.393	3.437
6/29/2010	2.139	3.186	4.002	4.711	5.550	6.035
7/3/2010	1.756	2.550	2.726	2.839	2.905	2.839
8/16/2010	1.872	2.888	3.696	4.257	4.799	5.149
8/18/2010	3.669	6.134	7.959	9.078	9.872	10.699
9/6/2010	2.009	3.027	3.481	3.517	3.615	3.568
9/27/2010	4.349	6.777	7.850	8.722	9.239	9.733
9/30/2010	4.122	7.490	10.628	13.248	14.464	15.984
10/1/2010	2.015	3.138	3.738	4.394	4.986	5.451
10/6/2010	1.750	2.524	3.089	3.481	3.921	4.093
10/11/2010	2.522	3.939	4.893	5.733	6.151	6.466
10/14/2010	1.938	3.109	4.022	4.790	5.471	6.259
11/4/2010	1.769	2.742	3.352	3.932	4.409	4.875
11/20/2010	1.657	2.482	2.782	2.876	2.886	2.941
12/16/2010	12.353	20.690	25.779	27.715	28.763	30.169
12/19/2010	1.349	2.040	2.299	2.435	2.371	2.412
12/31/2010	1.620	2.547	2.877	2.965	3.059	3.171
1/22/2011	1.654	2.412	2.636	2.675	2.715	2.819
1/26/2011	8.562	12.103	13.948	14.794	15.736	16.896
2/22/2011	6.614	10.487	12.926	14.064	15.055	15.901
5/24/2011	1.519	2.493	3.222	3.987	4.649	5.332
6/15/2011	1.586	2.269	2.787	3.085	3.404	3.712
7/4/2011	1.602	2.460	2.768	2.857	2.869	2.933
Average MAPE	3.313	5.153	6.290	7.033	7.569	8.060
Variance of MAPE	7.020	18.553	29.377	35.616	39.762	43.704

Table 12 KNN-Integrated recursive prediction results for one day

Date	5 min	10 min	15 min	20 min	25 min	30 min
1/21/2010	2.983	4.119	4.976	5.858	6.224	6.856
1/25/2010	1.871	2.521	2.863	3.042	3.090	3.049
2/9/2010	2.490	3.731	4.599	5.177	5.486	5.903
2/10/2010	7.900	11.406	13.887	15.360	15.430	15.281
2/19/2010	7.210	12.909	16.391	20.345	22.921	24.520
3/12/2010	4.428	7.176	9.018	10.726	12.261	13.584
4/8/2010	2.024	3.049	3.830	4.468	5.140	5.680
5/31/2010	2.054	3.040	3.645	3.873	3.953	3.938
6/29/2010	2.138	3.387	4.365	5.140	5.948	6.570
7/3/2010	1.756	2.608	3.036	3.176	3.317	3.313
8/16/2010	1.872	2.994	3.917	4.649	5.162	5.703
8/18/2010	3.669	6.460	8.330	9.725	10.986	11.934
9/6/2010	2.009	3.071	3.638	3.861	3.906	3.932
9/27/2010	4.349	7.218	9.250	11.143	12.485	13.647
9/30/2010	4.122	6.513	8.324	9.654	10.905	11.840
10/1/2010	2.015	3.014	3.689	4.284	4.841	5.343
10/6/2010	1.750	2.566	3.066	3.457	3.642	3.849
10/11/2010	2.522	3.981	5.146	6.267	7.048	7.591
10/14/2010	1.938	3.053	4.065	4.699	5.391	6.015
11/4/2010	1.769	2.794	3.566	4.198	4.704	5.229
11/20/2010	1.657	2.589	3.167	3.333	3.360	3.336
12/16/2010	12.353	21.162	27.538	33.274	39.190	44.048
12/19/2010	1.349	2.124	2.534	2.644	2.676	2.628
12/31/2010	1.620	2.630	3.071	3.331	3.471	3.402
1/22/2011	1.654	2.508	2.936	3.061	3.137	3.152
1/26/2011	8.562	12.192	15.337	16.597	18.395	18.983
2/22/2011	6.614	10.550	13.579	15.835	18.091	20.996
5/24/2011	1.518	2.457	3.209	3.760	4.233	4.689
6/15/2011	1.586	2.205	2.712	3.054	3.403	3.719
7/4/2011	1.602	2.558	3.000	3.073	3.064	3.066
Average MAPE	3.313	5.220	6.556	7.569	8.395	9.060
Variance of MAPE	7.020	19.238	32.219	46.576	62.821	77.843
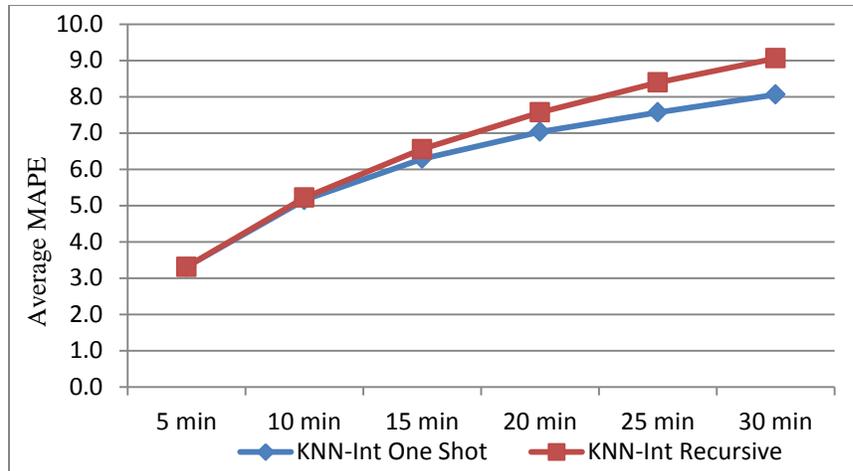
Figure 22 Comparison of Average MAPE of KNN-Integrated one shot and recursive models – one day

Similarly, Tables 13 and 14 give the average MAPE for 5 minutes up to 30 minutes prediction time interval for the peak hour period of each day. Figure 23 shows the comparison results of the total average MAPE of all selected days for their peak period over the six time intervals between the KNN-Integrated one shot model and KNN-Integrated recursive model.

Table 13 KNN-Integrated one shot prediction results for peak hour

Date	5 min	10 min	15 min	20 min	25 min	30 min
1/21/2010	6.816	8.410	11.021	17.262	16.135	16.789
1/25/2010	2.006	3.096	3.622	3.507	3.371	3.781
2/9/2010	2.416	3.930	5.013	5.986	7.023	8.195
2/10/2010	9.401	12.572	13.291	14.669	14.919	14.597
2/19/2010	19.922	34.326	47.583	57.526	55.332	44.140
3/12/2010	12.299	20.537	25.211	28.995	32.828	36.097
4/8/2010	3.039	5.527	7.545	9.634	11.669	14.795
5/31/2010	2.642	3.447	4.530	5.562	7.107	9.841
6/29/2010	3.751	6.620	8.576	11.291	14.753	17.523
7/3/2010	1.924	2.825	2.994	3.172	3.011	2.903
8/16/2010	2.559	4.541	6.823	9.143	10.761	10.927
8/18/2010	9.566	17.298	22.961	26.716	30.312	33.929
9/6/2010	1.810	2.737	3.271	3.471	3.584	3.497
9/27/2010	2.633	3.916	4.421	5.556	6.597	7.185
9/30/2010	8.564	15.355	23.689	27.915	30.251	32.796
10/1/2010	3.325	5.510	8.144	9.902	12.256	14.360
10/6/2010	2.096	3.478	4.687	6.003	7.472	8.699
10/11/2010	2.541	4.233	5.436	6.387	7.146	7.972
10/14/2010	2.940	4.822	6.848	9.274	11.821	14.075
11/4/2010	2.323	4.266	6.353	8.399	10.265	11.344
11/20/2010	1.515	2.503	2.919	3.325	3.373	3.686
12/16/2010	25.605	40.394	48.734	53.425	51.132	50.629
12/19/2010	1.159	1.634	1.865	2.010	2.157	2.248
12/31/2010	2.100	3.100	3.414	3.758	3.471	3.527
1/22/2011	1.720	2.521	2.605	2.694	2.787	2.782
1/26/2011	8.562	12.103	13.948	14.794	15.736	16.896
2/22/2011	1.269	1.991	2.210	2.242	2.539	2.563
5/24/2011	2.705	4.962	7.160	9.062	11.077	13.071
6/15/2011	2.371	3.995	5.753	6.275	6.931	7.360
7/4/2011	1.884	3.298	3.830	3.946	3.903	3.948
Average MAPE	5.049	8.132	10.482	12.397	13.324	14.005
Variance of MAPE	32.336	86.315	143.935	191.888	185.141	165.228

Table 14 KNN-Integrated recursive prediction results for peak hour

Date	5 min	10 min	15 min	20 min	25 min	30 min
1/21/2010	6.816	8.280	10.979	13.761	17.472	20.477
1/25/2010	2.006	3.058	3.558	3.767	3.736	3.917
2/9/2010	2.416	3.760	4.534	5.471	5.752	6.378
2/10/2010	9.401	12.684	14.009	15.877	16.275	16.353
2/19/2010	19.922	38.392	47.874	61.509	65.912	69.621
3/12/2010	12.299	21.211	26.470	34.626	41.675	48.475
4/8/2010	3.039	5.767	7.984	10.159	12.557	14.974
5/31/2010	2.642	3.913	4.593	4.949	4.907	4.870
6/29/2010	3.751	6.979	9.820	12.568	16.030	19.163
7/3/2010	1.924	2.797	3.233	3.709	3.520	3.605
8/16/2010	2.559	4.682	7.079	9.274	10.901	13.121
8/18/2010	9.566	19.919	27.128	34.225	38.460	41.999
9/6/2010	1.810	2.943	2.995	3.268	3.227	3.112
9/27/2010	2.633	4.204	5.331	6.185	6.733	7.408
9/30/2010	8.564	14.870	22.675	26.755	30.472	33.351
10/1/2010	3.325	5.278	7.328	9.521	12.171	14.209
10/6/2010	2.096	3.412	4.626	6.034	7.399	8.705
10/11/2010	2.541	4.193	5.712	7.095	8.529	9.730
10/14/2010	2.940	5.130	7.513	10.118	12.393	14.896
11/4/2010	2.323	4.443	5.592	7.601	8.903	10.355
11/20/2010	1.515	2.737	3.253	3.395	3.559	3.574
12/16/2010	25.605	46.732	53.630	65.927	63.457	72.197
12/19/2010	1.159	1.627	1.829	2.072	2.318	2.449
12/31/2010	2.100	3.201	3.809	4.020	3.694	3.561
1/22/2011	1.720	2.668	3.222	3.196	3.317	3.178
1/26/2011	8.562	12.192	15.337	16.597	18.395	18.983
2/22/2011	1.269	2.022	2.097	2.283	2.446	2.367
5/24/2011	2.705	5.387	7.582	9.868	12.472	14.785
6/15/2011	2.371	4.129	5.530	6.565	7.964	9.164
7/4/2011	1.884	3.275	3.994	4.229	4.247	4.104
Average MAPE	5.049	8.663	10.977	13.487	14.963	16.636
Variance of MAPE	32.336	111.057	162.451	258.521	281.869	345.478
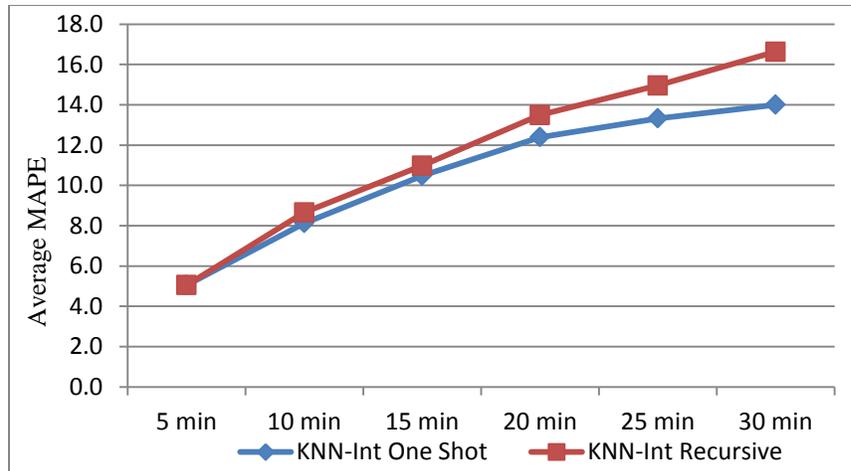
Figure 23 Comparison of Average MAPE of KNN-Integrated one shot and recursive models-peak hour

Table 15 lists the average MAPE and the variance of MAPE for the one shot and recursive model results from both one day and peak hours. The comparison results show that the two model one shot and recursive, exhibit similar travel time prediction performance. KNN-Integrated one shot model is better than the recursive model in terms of prediction accuracy, both for one day and peak hour periods. Also, for the computation time, the recursive model requires about five times more than the one shot model.

To further investigate the model's prediction results for each day other than the average MAPE of all days, a paired *t*-test was conducted for the model performance comparisons. 95% confidence intervals were constructed comparing the KNN-Integrated one shot against the KNN-Integrated recursive model for each time interval based on their MAPE. Only the 10 minutes prediction for one day did not

pass the *t*-test. The results in Table 16 indicate that KNN-Integrated one shot model

outperforms the recursive model.

Table 15 Comparisons of MAPE of the KNN-Integrated one shot and recursive models

		10 min	15 min	20 min	25 min	30 min
Average MAPE -one day	One shot	5.15	6.29	7.03	7.57	8.06
	Recursive	5.22	6.56	7.57	8.40	9.06
Variance of MAPE -one day	One shot	18.55	29.38	35.62	39.76	43.70
	Recursive	19.24	32.22	46.58	62.82	77.84
Average MAPE -peak hour	One shot	8.13	10.48	12.40	13.32	14.01
	Recursive	8.66	10.98	13.49	14.96	16.64
Variance of MAPE -peak hour	One shot	86.32	143.94	191.89	185.14	165.23
	Recursive	111.06	162.45	258.52	281.87	345.48

Table 16 T-test results of the KNN-Integrated one shot and recursive models

One shot and Recursive	10 min	15 min	20 min	25 min	30 min
P-Value for one day	0.1109	0.0162	0.0190	0.0235	0.0349
CI for one day	0.0240	-0.0647	-0.1170	-0.1494	-0.0974
P-Value for peak hour	0.0233	0.0187	0.0254	0.0079	0.0163
CI for peak hour	-0.0968	-0.1096	-0.1811	-0.5527	-0.6405

In conclusion, based on the prediction and *t*-test results, the comparisons of model

performance results indicate that the KNN-Integrated one shot model outperforms the

KNN-Integrated recursive model both in terms of accuracy and in computation time.

As a result, the one shot model is adopted and the results from the one shot model are

used as the results for the extended KNN-Integrated model in the following sections.

(2) Model performance comparisons between KNN-Integrated and other models.

In this part, the prediction results are presented and discussed for the selected 30 days, which exhibit various weather conditions including rain, snow, strong wind and low visibility scenarios. Model performance comparisons are made among four models: KNN-Integrated, KNN-T, KNN and ARIMA models, for both one day and peak hour periods' prediction. Tables 17 through 28 list the prediction results in MAPE for both one day period prediction and prediction for the peak hours (3:00pm-8:00pm), at every 5 minutes time interval (5 minutes, 10 minutes, 15 minutes, 20 minutes, 25 minutes and 30 minutes).

To summarize the results shown in Tables 17 through 28, Tables 29 and 30 list the total average MAPE and the variance of the MAPE for each model over the 30 days, both for one day and peak hour periods at each prediction time interval.

The comparison results of the prediction accuracy for the 30 days shown in the above tables indicate that the proposed extension of the KNN-Integrated model outperforms the KNN, KNN-T and ARIMA models in prediction accuracy. To better illustrate the comparison results between the KNN-Integrated model and the ARIMA model, the average MAPE for the two models are compared for both one day and peak hour periods for each of the 30 selected days, for 5 minutes prediction and 30 minutes prediction. The results are shown in Figures 24 to 27.

Table 17 Model performance results of MAPE for one day prediction - 5 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	2.983	3.285	3.382	3.248
1/25/2010	1.871	2.171	2.078	1.866
2/9/2010	2.490	2.697	2.853	2.626
2/10/2010	7.900	8.061	8.699	7.774
2/19/2010	7.210	7.489	8.461	8.706
3/12/2010	4.428	4.267	4.234	4.937
4/8/2010	2.024	2.223	2.190	2.156
5/31/2010	2.054	2.060	2.035	2.144
6/29/2010	2.139	2.412	2.467	2.281
7/3/2010	1.756	1.829	1.823	1.874
8/16/2010	1.872	1.975	1.997	2.066
8/18/2010	3.669	3.899	3.864	4.161
9/6/2010	2.009	2.054	2.041	2.186
9/27/2010	4.349	4.421	4.634	4.649
9/30/2010	4.122	3.826	3.903	3.695
10/1/2010	2.015	2.017	2.036	2.111
10/6/2010	1.750	1.882	1.886	1.923
10/11/2010	2.522	2.423	2.573	2.630
10/14/2010	1.938	1.948	1.949	2.111
11/4/2010	1.769	1.712	1.721	1.731
11/20/2010	1.657	1.734	1.719	1.814
12/16/2010	12.353	12.826	16.340	33.658
12/19/2010	1.349	1.400	1.396	1.524
12/31/2010	1.620	1.742	1.740	1.724
1/22/2011	1.654	1.767	1.718	1.813
1/26/2011	8.562	9.827	9.069	9.679
2/22/2011	6.614	6.801	7.042	8.013
5/24/2011	1.519	1.463	1.457	1.605
6/15/2011	1.586	1.631	1.621	1.687
7/4/2011	1.602	1.538	1.533	1.698
Average MAPE	3.313	3.446	3.615	4.270
Variance of MAPE	7.020	7.845	10.793	36.127

Table 18 Model performance results of MAPE for peak hour prediction - 5 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	6.816	7.417	7.412	8.798
1/25/2010	2.006	2.208	2.214	2.018
2/9/2010	2.416	2.748	3.142	3.166
2/10/2010	9.401	11.157	11.030	9.381
2/19/2010	19.922	23.452	25.862	23.880
3/12/2010	12.299	11.848	15.264	12.892
4/8/2010	3.039	3.458	3.373	3.337
5/31/2010	2.642	3.120	3.577	2.796
6/29/2010	3.751	4.689	5.154	4.270
7/3/2010	1.924	2.348	2.391	2.091
8/16/2010	2.559	3.374	3.185	3.282
8/18/2010	9.566	10.851	10.587	11.331
9/6/2010	1.810	1.958	2.070	1.864
9/27/2010	2.633	2.920	2.956	2.842
9/30/2010	8.564	9.241	11.819	9.979
10/1/2010	3.325	4.297	4.298	3.337
10/6/2010	2.096	2.270	2.336	2.368
10/11/2010	2.541	2.576	2.264	2.535
10/14/2010	2.940	3.226	3.308	3.692
11/4/2010	2.323	2.354	2.500	2.679
11/20/2010	1.515	1.684	1.780	1.724
12/16/2010	25.605	31.841	31.900	24.820
12/19/2010	1.159	1.231	1.209	1.132
12/31/2010	2.100	2.453	2.454	2.189
1/22/2011	1.720	1.823	1.902	1.733
1/26/2011	8.562	9.827	9.069	9.679
2/22/2011	1.269	1.405	1.472	1.318
5/24/2011	2.705	3.156	3.262	3.244
6/15/2011	2.371	2.506	2.603	2.729
7/4/2011	1.884	2.018	1.946	2.153
Average MAPE	5.049	5.782	6.078	5.575
Variance of MAPE	32.336	46.486	51.727	36.737

Table 19 Model performance results of MAPE for one day prediction - 10 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	3.822	4.602	4.560	4.265
1/25/2010	2.493	2.982	2.919	2.627
2/9/2010	3.954	4.351	4.458	4.168
2/10/2010	11.682	12.691	13.334	11.254
2/19/2010	12.053	12.288	13.927	14.116
3/12/2010	7.185	7.448	7.372	8.604
4/8/2010	3.047	3.381	3.367	3.337
5/31/2010	2.873	3.192	3.173	3.387
6/29/2010	3.186	3.742	3.666	3.521
7/3/2010	2.550	2.663	2.670	2.818
8/16/2010	2.888	3.044	3.072	3.311
8/18/2010	6.134	6.474	6.325	6.907
9/6/2010	3.027	3.197	3.117	3.411
9/27/2010	6.777	7.463	7.303	7.331
9/30/2010	7.490	6.265	6.190	6.216
10/1/2010	3.138	2.907	3.012	3.315
10/6/2010	2.524	2.754	2.782	2.914
10/11/2010	3.939	3.777	4.088	4.064
10/14/2010	3.109	3.123	3.061	3.491
11/4/2010	2.742	2.758	2.749	2.906
11/20/2010	2.482	2.615	2.585	2.783
12/16/2010	20.690	21.452	22.459	62.777
12/19/2010	2.040	2.104	2.056	2.313
12/31/2010	2.547	2.702	2.757	2.710
1/22/2011	2.412	2.637	2.609	2.773
1/26/2011	12.103	13.954	13.253	14.364
2/22/2011	10.487	11.292	11.410	13.645
5/24/2011	2.493	2.364	2.378	2.601
6/15/2011	2.269	2.328	2.366	2.461
7/4/2011	2.460	2.519	2.508	2.777
Average MAPE	5.153	5.436	5.517	7.039
Variance of MAPE	18.553	20.763	22.700	124.151

Table 20 Model performance results of MAPE for peak hour prediction - 10 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	8.410	11.056	11.099	10.536
1/25/2010	3.096	3.483	3.415	3.072
2/9/2010	3.930	5.177	4.902	5.235
2/10/2010	12.572	17.779	17.098	14.026
2/19/2010	34.326	35.921	42.253	36.543
3/12/2010	20.537	21.583	24.078	23.099
4/8/2010	5.527	5.608	5.857	6.065
5/31/2010	3.447	4.314	4.435	3.949
6/29/2010	6.620	7.994	7.235	7.073
7/3/2010	2.825	3.104	3.195	3.201
8/16/2010	4.541	5.439	5.417	5.443
8/18/2010	17.298	22.762	16.696	19.717
9/6/2010	2.737	2.945	2.882	2.718
9/27/2010	3.916	4.591	4.617	4.255
9/30/2010	15.355	14.063	16.700	18.723
10/1/2010	5.510	6.001	6.246	5.572
10/6/2010	3.478	3.811	3.666	3.814
10/11/2010	4.233	4.474	3.812	4.356
10/14/2010	4.822	4.913	5.286	6.105
11/4/2010	4.266	4.234	4.090	4.757
11/20/2010	2.503	2.676	2.747	2.728
12/16/2010	40.394	52.636	49.295	41.641
12/19/2010	1.634	1.599	1.592	1.621
12/31/2010	3.100	4.015	3.799	3.424
1/22/2011	2.521	2.754	2.813	2.736
1/26/2011	12.103	13.954	13.253	14.364
2/22/2011	1.991	2.211	2.166	1.959
5/24/2011	4.962	5.686	5.322	5.623
6/15/2011	3.995	4.368	4.046	4.443
7/4/2011	3.298	3.501	3.290	3.598
Average MAPE	8.132	9.422	9.377	9.013
Variance of MAPE	86.315	125.828	128.332	98.128

Table 21 Model performance results of MAPE for one day prediction - 15 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	4.598	5.484	5.010	5.038
1/25/2010	2.807	3.274	3.128	2.944
2/9/2010	4.531	5.554	5.580	5.181
2/10/2010	13.782	16.737	17.177	13.510
2/19/2010	16.087	16.027	18.144	17.863
3/12/2010	8.697	9.603	9.555	11.496
4/8/2010	3.550	4.204	4.144	4.163
5/31/2010	3.203	3.676	3.655	3.959
6/29/2010	4.002	4.682	4.514	4.390
7/3/2010	2.726	2.975	2.995	3.193
8/16/2010	3.696	3.962	3.892	4.222
8/18/2010	7.959	8.593	7.986	8.829
9/6/2010	3.481	3.647	3.574	3.973
9/27/2010	7.850	8.196	8.427	9.215
9/30/2010	10.628	8.041	8.195	8.236
10/1/2010	3.738	3.559	3.616	4.156
10/6/2010	3.089	3.129	3.094	3.415
10/11/2010	4.893	4.816	5.206	5.220
10/14/2010	4.022	4.086	4.022	4.629
11/4/2010	3.352	3.478	3.381	3.772
11/20/2010	2.782	2.990	2.944	3.320
12/16/2010	25.779	27.987	27.191	87.690
12/19/2010	2.299	2.366	2.299	2.693
12/31/2010	2.877	3.226	3.140	3.117
1/22/2011	2.636	2.945	2.922	3.150
1/26/2011	13.948	17.474	16.824	17.568
2/22/2011	12.926	14.514	14.369	17.960
5/24/2011	3.222	2.877	2.901	3.362
6/15/2011	2.787	2.844	2.826	2.915
7/4/2011	2.768	2.875	2.837	3.166
Average MAPE	6.290	6.794	6.785	9.078
Variance of MAPE	29.377	36.023	36.321	242.392

Table 22 Model performance results of MAPE for peak hour prediction - 15 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	11.021	13.312	13.283	12.849
1/25/2010	3.622	4.495	4.355	3.607
2/9/2010	5.013	7.043	6.378	6.524
2/10/2010	13.291	21.533	21.754	16.734
2/19/2010	47.583	46.133	54.183	44.411
3/12/2010	25.211	27.481	32.629	30.642
4/8/2010	7.545	8.212	8.381	8.272
5/31/2010	4.530	4.665	4.848	4.784
6/29/2010	8.576	9.410	10.501	9.270
7/3/2010	2.994	3.734	3.882	3.776
8/16/2010	6.823	6.650	6.508	7.428
8/18/2010	22.961	33.466	23.235	25.242
9/6/2010	3.271	3.581	3.385	3.189
9/27/2010	4.421	6.004	5.997	5.334
9/30/2010	23.689	18.476	19.504	27.127
10/1/2010	8.144	7.287	8.200	7.548
10/6/2010	4.687	4.996	5.026	5.257
10/11/2010	5.436	5.776	5.137	5.782
10/14/2010	6.848	7.095	7.455	8.420
11/4/2010	6.353	6.578	6.145	6.731
11/20/2010	2.919	3.165	3.134	3.291
12/16/2010	48.734	63.876	54.590	55.440
12/19/2010	1.865	1.690	1.666	1.731
12/31/2010	3.414	4.846	4.391	3.985
1/22/2011	2.605	3.256	3.266	3.269
1/26/2011	13.948	17.474	16.824	17.568
2/22/2011	2.210	2.361	2.456	2.185
5/24/2011	7.160	7.973	7.415	7.739
6/15/2011	5.753	5.562	5.655	5.722
7/4/2011	3.830	4.165	3.914	4.126
Average MAPE	10.482	12.010	11.803	11.599
Variance of MAPE	143.935	198.900	185.712	165.694

Table 23 Model performance results of MAPE for one day prediction - 20 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	5.572	6.303	6.250	5.869
1/25/2010	3.033	3.654	3.565	3.166
2/9/2010	5.289	6.392	6.574	5.901
2/10/2010	14.528	19.007	19.734	15.090
2/19/2010	19.141	19.169	21.969	20.965
3/12/2010	9.863	11.272	11.053	13.653
4/8/2010	4.136	4.908	4.788	4.804
5/31/2010	3.264	3.914	3.940	4.238
6/29/2010	4.711	5.584	5.282	5.029
7/3/2010	2.839	3.171	3.204	3.375
8/16/2010	4.257	4.760	4.811	4.883
8/18/2010	9.078	9.641	9.069	10.159
9/6/2010	3.517	3.818	3.760	4.332
9/27/2010	8.722	9.840	9.994	11.029
9/30/2010	13.248	9.082	9.265	9.812
10/1/2010	4.394	4.006	4.043	4.747
10/6/2010	3.481	3.367	3.384	3.773
10/11/2010	5.733	5.675	6.219	6.272
10/14/2010	4.790	4.919	4.833	5.296
11/4/2010	3.932	4.069	3.959	4.414
11/20/2010	2.876	3.126	3.112	3.529
12/16/2010	27.715	32.886	31.547	105.097
12/19/2010	2.435	2.417	2.336	2.882
12/31/2010	2.965	3.392	3.365	3.409
1/22/2011	2.675	3.008	3.013	3.324
1/26/2011	14.794	19.056	18.613	19.708
2/22/2011	14.064	16.816	16.750	21.106
5/24/2011	3.987	3.434	3.413	3.883
6/15/2011	3.085	3.172	3.132	3.195
7/4/2011	2.857	2.976	2.940	3.224
Average MAPE	7.033	7.761	7.797	10.539
Variance of MAPE	35.616	49.220	49.674	349.117

Table 24 Model performance results of MAPE for peak hour prediction - 20 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	17.262	24.079	23.952	15.881
1/25/2010	3.507	5.102	4.815	3.747
2/9/2010	5.986	8.346	7.512	7.204
2/10/2010	14.669	23.919	23.626	17.899
2/19/2010	57.526	54.258	64.704	52.170
3/12/2010	28.995	33.365	38.854	36.506
4/8/2010	9.634	10.583	10.423	10.046
5/31/2010	5.562	5.704	6.099	4.872
6/29/2010	11.291	12.995	13.810	10.991
7/3/2010	3.172	4.025	4.217	4.029
8/16/2010	9.143	7.869	7.925	8.893
8/18/2010	26.716	37.102	28.461	29.482
9/6/2010	3.471	3.681	3.681	3.376
9/27/2010	5.556	6.967	7.004	5.596
9/30/2010	27.915	23.673	24.976	32.857
10/1/2010	9.902	9.420	9.646	9.458
10/6/2010	6.003	6.043	5.922	6.338
10/11/2010	6.387	6.618	5.998	6.944
10/14/2010	9.274	10.198	9.927	10.274
11/4/2010	8.399	7.105	7.389	7.986
11/20/2010	3.325	3.349	3.180	3.487
12/16/2010	53.425	68.921	62.116	61.112
12/19/2010	2.010	1.788	1.753	1.876
12/31/2010	3.758	5.410	4.455	4.303
1/22/2011	2.694	3.424	3.443	3.301
1/26/2011	14.794	19.056	18.613	19.708
2/22/2011	2.242	2.418	2.431	2.286
5/24/2011	9.062	10.320	9.328	9.254
6/15/2011	6.275	6.831	6.969	6.181
7/4/2011	3.946	4.269	4.136	4.305
Average MAPE	12.397	14.228	14.179	13.345
Variance of MAPE	191.888	251.439	259.159	218.798

Table 25 Model performance results of MAPE for one day prediction - 25 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	5.864	6.829	6.618	6.051
1/25/2010	3.332	3.954	3.838	3.278
2/9/2010	6.061	7.207	7.343	6.368
2/10/2010	14.746	20.099	21.131	15.287
2/19/2010	21.371	19.553	22.036	23.503
3/12/2010	10.595	12.688	12.617	15.606
4/8/2010	4.831	5.649	5.596	5.384
5/31/2010	3.393	4.058	4.092	4.315
6/29/2010	5.550	6.373	6.415	5.696
7/3/2010	2.905	3.304	3.343	3.528
8/16/2010	4.799	5.400	5.395	5.358
8/18/2010	9.872	10.674	10.103	11.327
9/6/2010	3.615	3.925	3.848	4.335
9/27/2010	9.239	11.011	11.231	12.248
9/30/2010	14.464	9.709	10.728	10.957
10/1/2010	4.986	4.561	4.537	5.312
10/6/2010	3.921	3.608	3.651	3.919
10/11/2010	6.151	6.443	6.963	6.983
10/14/2010	5.471	5.658	5.509	5.913
11/4/2010	4.409	4.513	4.446	5.041
11/20/2010	2.886	3.206	3.232	3.487
12/16/2010	28.763	36.424	35.133	118.929
12/19/2010	2.371	2.488	2.432	2.827
12/31/2010	3.059	3.539	3.434	3.449
1/22/2011	2.715	3.076	3.120	3.383
1/26/2011	15.736	20.299	19.269	21.833
2/22/2011	15.055	19.131	19.179	23.837
5/24/2011	4.649	4.077	3.951	4.332
6/15/2011	3.404	3.503	3.499	3.422
7/4/2011	2.869	3.009	2.965	3.173
Average MAPE	7.569	8.466	8.522	11.636
Variance of MAPE	39.762	58.423	58.387	448.959

Table 26 Model performance results of MAPE for peak hour prediction - 25 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	16.135	23.574	23.464	16.277
1/25/2010	3.371	5.369	5.056	3.481
2/9/2010	7.023	9.098	8.370	7.637
2/10/2010	14.919	21.550	23.140	16.817
2/19/2010	55.332	49.430	67.256	53.122
3/12/2010	32.828	41.366	45.621	41.781
4/8/2010	11.669	12.974	13.018	12.074
5/31/2010	7.107	5.996	6.302	5.053
6/29/2010	14.753	16.760	17.851	13.290
7/3/2010	3.011	4.401	4.776	3.998
8/16/2010	10.761	9.404	8.898	10.160
8/18/2010	30.312	42.722	30.501	32.797
9/6/2010	3.584	3.695	3.654	3.334
9/27/2010	6.597	7.757	7.671	6.291
9/30/2010	30.251	24.919	29.233	37.118
10/1/2010	12.256	12.045	12.089	11.098
10/6/2010	7.472	7.224	7.020	7.417
10/11/2010	7.146	8.450	6.884	7.988
10/14/2010	11.821	13.470	12.779	12.022
11/4/2010	10.265	8.159	7.732	9.106
11/20/2010	3.373	3.603	3.262	3.490
12/16/2010	51.132	70.078	66.768	66.260
12/19/2010	2.157	1.925	1.847	2.142
12/31/2010	3.471	5.500	4.142	4.018
1/22/2011	2.787	3.321	3.473	3.297
1/26/2011	15.736	20.299	19.269	21.833
2/22/2011	2.539	2.630	2.429	2.242
5/24/2011	11.077	13.074	12.046	10.972
6/15/2011	6.931	8.053	7.887	6.880
7/4/2011	3.903	4.443	4.186	4.250
Average MAPE	13.324	15.376	15.554	14.541
Variance of MAPE	185.141	257.456	296.015	254.638

Table 27 Model performance results of MAPE for one day prediction - 30 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	6.584	7.396	7.360	6.653
1/25/2010	3.599	4.007	3.994	3.203
2/9/2010	6.730	7.856	7.948	6.759
2/10/2010	15.515	20.032	21.136	15.270
2/19/2010	21.872	20.749	22.314	25.101
3/12/2010	10.977	13.894	14.086	17.456
4/8/2010	5.650	6.229	6.268	5.787
5/31/2010	3.437	4.118	4.155	4.334
6/29/2010	6.035	7.041	7.201	6.274
7/3/2010	2.839	3.271	3.272	3.520
8/16/2010	5.149	5.874	5.855	5.678
8/18/2010	10.699	11.656	11.198	12.161
9/6/2010	3.568	4.023	3.977	4.485
9/27/2010	9.733	12.120	12.280	13.413
9/30/2010	15.984	9.752	11.183	12.033
10/1/2010	5.451	5.202	5.012	5.672
10/6/2010	4.093	3.887	3.803	3.988
10/11/2010	6.466	7.059	7.543	7.532
10/14/2010	6.259	6.466	6.114	6.478
11/4/2010	4.875	4.977	4.957	5.561
11/20/2010	2.941	3.269	3.328	3.495
12/16/2010	30.169	39.082	36.935	129.241
12/19/2010	2.412	2.585	2.594	2.846
12/31/2010	3.171	3.705	3.643	3.423
1/22/2011	2.819	3.112	3.226	3.324
1/26/2011	16.896	21.045	19.954	22.638
2/22/2011	15.901	21.408	21.469	27.286
5/24/2011	5.332	4.490	4.390	4.760
6/15/2011	3.712	3.745	3.794	3.659
7/4/2011	2.933	3.075	3.008	3.230
Average MAPE	8.060	9.037	9.067	12.509
Variance of MAPE	43.704	66.318	63.679	532.089

Table 28 Model performance results of MAPE for peak hour prediction - 30 min

Date	KNN-Integrated	KNN-T	KNN	ARIMA
1/21/2010	16.789	25.779	25.942	17.907
1/25/2010	3.781	5.973	5.186	3.457
2/9/2010	8.195	9.820	8.678	7.675
2/10/2010	14.597	21.030	18.859	15.621
2/19/2010	44.140	50.082	71.291	51.622
3/12/2010	36.097	47.904	51.941	45.935
4/8/2010	14.795	14.652	15.512	13.552
5/31/2010	9.841	5.487	5.880	5.090
6/29/2010	17.523	19.475	19.550	15.498
7/3/2010	2.903	3.799	4.224	3.862
8/16/2010	10.927	10.868	10.528	11.068
8/18/2010	33.929	41.562	34.625	35.171
9/6/2010	3.497	3.604	3.439	3.450
9/27/2010	7.185	8.365	8.293	6.584
9/30/2010	32.796	29.197	31.938	40.649
10/1/2010	14.360	14.046	13.443	12.241
10/6/2010	8.699	8.315	8.139	8.167
10/11/2010	7.972	9.807	7.296	8.711
10/14/2010	14.075	16.155	14.539	13.317
11/4/2010	11.344	9.707	9.001	10.053
11/20/2010	3.686	3.664	3.554	3.459
12/16/2010	50.629	68.429	71.965	70.890
12/19/2010	2.248	2.000	1.918	2.247
12/31/2010	3.527	4.962	3.864	3.576
1/22/2011	2.782	3.795	3.538	3.341
1/26/2011	16.896	21.045	19.954	22.638
2/22/2011	2.563	2.949	2.750	2.290
5/24/2011	13.071	14.642	14.213	12.113
6/15/2011	7.360	9.429	8.805	7.616
7/4/2011	3.948	4.545	4.189	4.220
Average MAPE	14.005	16.370	16.768	15.401
Variance of MAPE	165.228	263.958	347.566	282.901

Table 29 Model performance results of MAPE for one day prediction

One Day period	5 min	10 min	15 min	20 min	25 min	30 min
Average MAPE						
KNN-Integrated	3.313	5.153	6.290	7.033	7.569	8.060
KNN-T	3.446	5.436	6.794	7.761	8.466	9.037
KNN	3.615	5.517	6.785	7.797	8.522	9.067
ARIMA	4.270	7.039	9.078	10.539	11.636	12.509
Var. of MAPE						
KNN-Integrated	7.020	18.553	29.377	35.616	39.762	43.704
KNN-T	7.845	20.763	36.023	49.220	58.423	66.319
KNN	10.793	22.700	36.321	49.674	58.387	63.679
ARIMA	36.127	124.151	242.392	349.117	448.959	532.089

Table 30 Model performance results of MAPE for peak hour prediction

Peak hour period	5 min	10 min	15 min	20 min	25 min	30 min
Average MAPE						
KNN-Integrated	5.049	8.132	10.482	12.397	13.324	14.005
KNN-T	5.782	9.422	12.010	14.228	15.376	16.370
KNN	6.078	9.377	11.803	14.179	15.554	16.769
ARIMA	5.575	9.013	11.599	13.345	14.541	15.401
Var. of MAPE						
KNN-Integrated	32.336	86.315	143.935	191.888	185.141	165.228
KNN-T	46.486	125.828	198.900	251.439	257.456	263.958
KNN	51.727	128.332	185.712	259.159	296.015	347.566
ARIMA	36.737	98.128	165.694	218.798	254.638	282.901
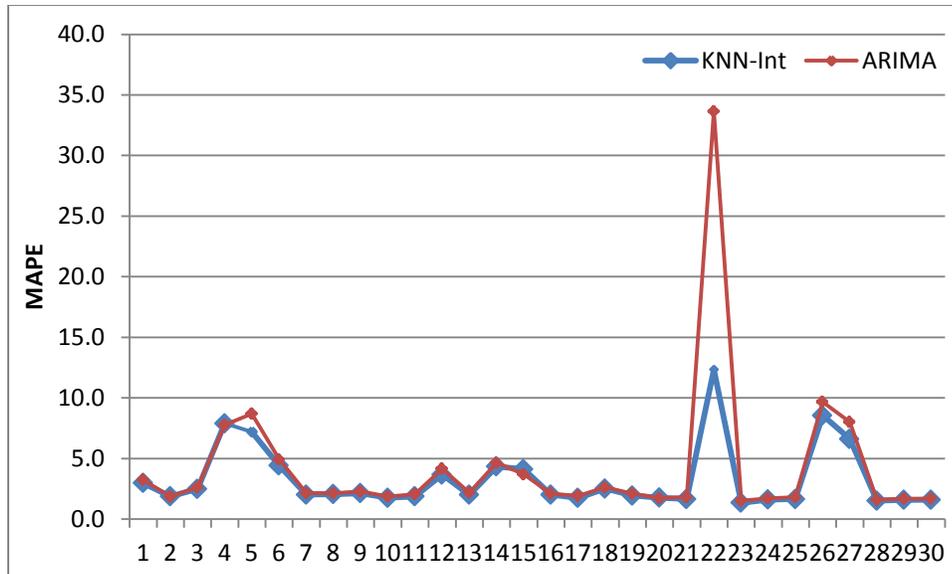
Figure 24 Comparison of Average MAPE of KNN-Integrated and ARIMA-one day-5 min
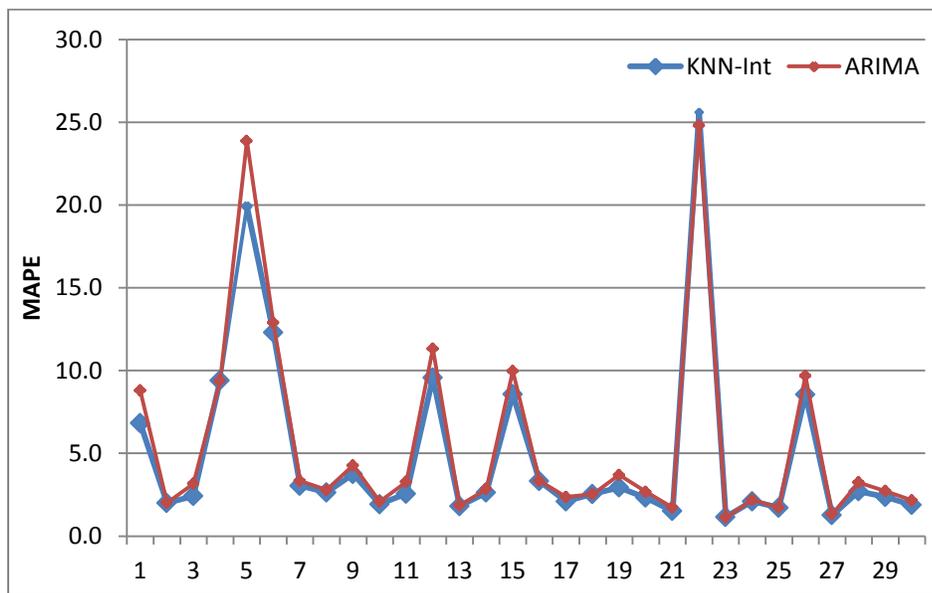
Figure 25 Comparison of Average MAPE of KNN-Integrated and ARIMA-peak hour-5 min
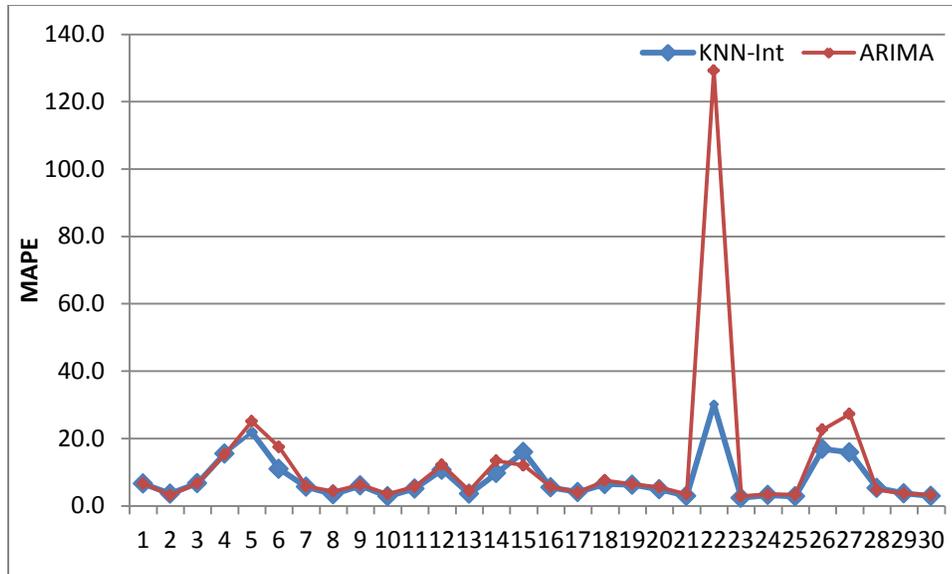
Figure 26 Comparison of Average MAPE of KNN-Integrated and ARIMA-one day-30 min
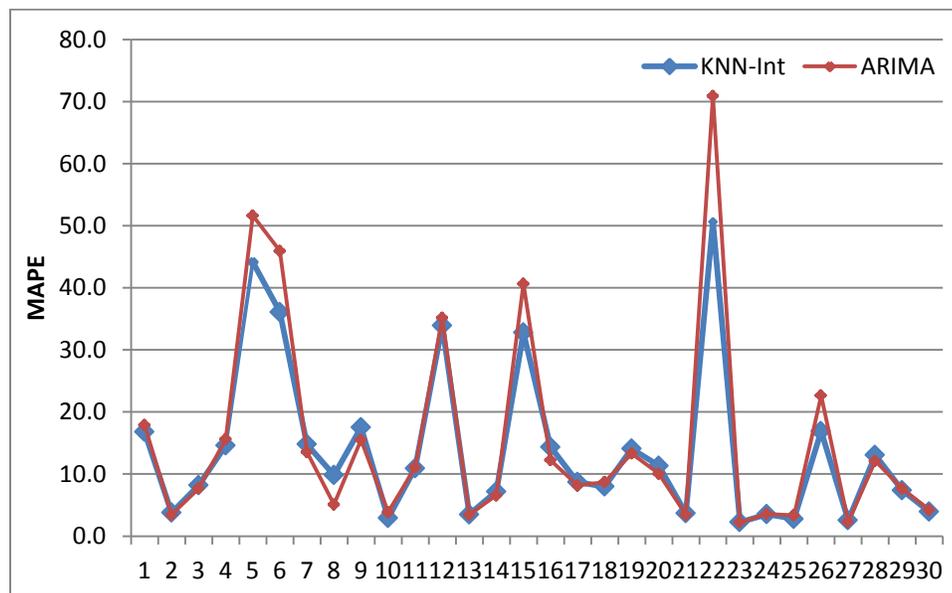
Figure 27 Comparison of Average MAPE of KNN-Integrated and ARIMA-peak hour-30 min

From the comparison results in the above Figures, the extended KNN-Integrate model shows better accuracy for most cases than the ARIMA model, especially when the prediction time interval is longer.

To investigate the improvement of the proposed KNN-Integrated model in terms of the prediction accuracy, Table 31 presents the percentage of reduced MAPE between ARIMA and KNN-Integrated models both for one day and peak hour prediction periods. As seen from the table, for 5 minutes prediction time interval, the KNN-Integrated model decreased the MAPE of ARIMA model by approximately 22.4% for all day period and 9.4% for peak hour period. For 30 minutes prediction time interval, the KNN-Integrated model decreased the MAPE of ARIMA model by approximately 35.6% for all day period and 9.1% for peak hour period.

Table 31 Model improvements in prediction accuracy

	5 min	10 min	15 min	20 min	25 min	30 min
One Day period Average MAPE						
KNN-Integrated	3.313	5.153	6.290	7.033	7.569	8.060
ARIMA	4.270	7.039	9.078	10.539	11.636	12.509
Improvement %	22.4%	26.8%	30.7%	33.3%	35.0%	35.6%
Peak hour period Average MAPE						
KNN-Integrated	5.049	8.132	10.482	12.397	13.324	14.005
ARIMA	5.575	9.013	11.599	13.345	14.541	15.401
Improvement %	9.4%	9.8%	9.6%	7.1%	8.4%	9.1%

To further investigate the model prediction performances, paired *t*-tests were

conducted for the model performance comparisons. 95% confidence intervals were

constructed comparing the KNN-Integrated model against the ARIMA, KNN and

KNN-T models. The *t*-test results in Tables 32 and 33 indicated that KNN-Integrated

model outperformed KNN and KNN-T model, passing every paired *t*-test. For its

comparison with ARIMA model for one day prediction, KNN-Integrated did not

show significant better performance at the 0.05 significance level. Paired *t*-tests at the

0.1 significance level were passed. For the peak hour prediction, paired *t*-tests at the

0.05 significance level were passed except for the 30 minutes prediction. It can be

seen that the extended KNN-Integrated model exhibits better prediction accuracy than

the other models, especially during the peak hour period.

Table 32 T- test results for model comparison -one day prediction

One day period	5 min	10 min	15 min	20 min	25 min	30 min
P-Value (α=0.05)						
KNN-Integrated vs KNN-T	0.0050	0.0023	0.0073	0.0126	0.0170	0.0202
KNN-Integrated vs KNN	0.0182	0.0017	0.0066	0.0081	0.0065	0.0081
KNN-Integrated vs ARIMA	0.0928	0.0933	0.0924	0.0913	0.0925	0.0940
CI						
KNN-Integrated vs KNN-T	-0.0510	-0.1264	-0.1738	-0.2043	-0.2122	-0.2034
KNN-Integrated vs KNN	-0.0681	-0.1708	-0.1763	-0.2554	-0.3403	-0.3366
KNN-Integrated vs ARIMA	0.2420	0.4828	0.6984	0.8563	1.0231	1.1582

Table 33 T-test results for model comparison –peak hour prediction

Peak hour period	5 min	10 min	15 min	20 min	25 min	30 min
P-Value (α=0.05)						
KNN-Integrated vs KNN-T	0.0018	0.0042	0.0163	0.0093	0.0139	0.0037
KNN-Integrated vs KNN	0.0007	0.0024	0.0047	0.0025	0.0050	0.0177
KNN-Integrated vs ARIMA	0.0011	0.0000	0.0014	0.0262	0.0372	0.0576
CI						
KNN-Integrated vs KNN-T	-0.3404	-0.5163	-0.3710	-0.5826	-0.5476	-0.9678
KNN-Integrated vs KNN	-0.5369	-0.5514	-0.5155	-0.7849	-0.8535	-0.6358
KNN-Integrated vs ARIMA	-0.2597	-0.5928	-0.5348	-0.1515	-0.1002	0.0644

In conclusion, based on the test results comparisons in the average MAPE and the variance of MAPE, along with the test results from t-tests, the extended KNN-Integrated model outperforms the other models, ARIMA, KNN and KNN-T, both in prediction accuracy and reliability. For the 5 minutes prediction time interval, the proposed KNN-Integrated model decreased the MAPE of ARIMA model by approximately 22.4% for all day period and 9.4% for peak hour period. For 30 minutes prediction time interval, the KNN-Integrated model decreased the MAPE of ARIMA model by approximately 35.6% for all day period and 9.1% for peak hour period.

(3) Examples of model prediction results from typical weather conditions

In this section, three days are selected and the predicted travel times for 5 minutes time interval are presented to compare with the real travel times. The three day selected are August 16, February 9 and October 14, 2010. August 16th, 2010 was a

typical weekday day without rain or snow or incidents. February 9th, 2010 was a typical snow day where snow started in the afternoon and continued until midnight. October 14th, 2010 was a rainy weekday where rain started in the late morning. The 5 minutes ahead predicted travel times from the extended KNN-Integrated model are presented and compared with the real travel time for each time interval for both one day and peak hour period on each day in Figures 28-33.

These figures indicate that the predicted travel time is very close to the real travel time and the KNN-Integrated model performs well for both normal and rainy/snowy weather scenarios. Another small example is presented here to illustrate the comparisons between the extended KNN-Integrated model and the ARIMA model under snow weather condition. Figures 34 and 35 present the predicted travel times from both KNN-Integrated and ARIMA model along with the real travel time for a snow day February 9th, 2010.
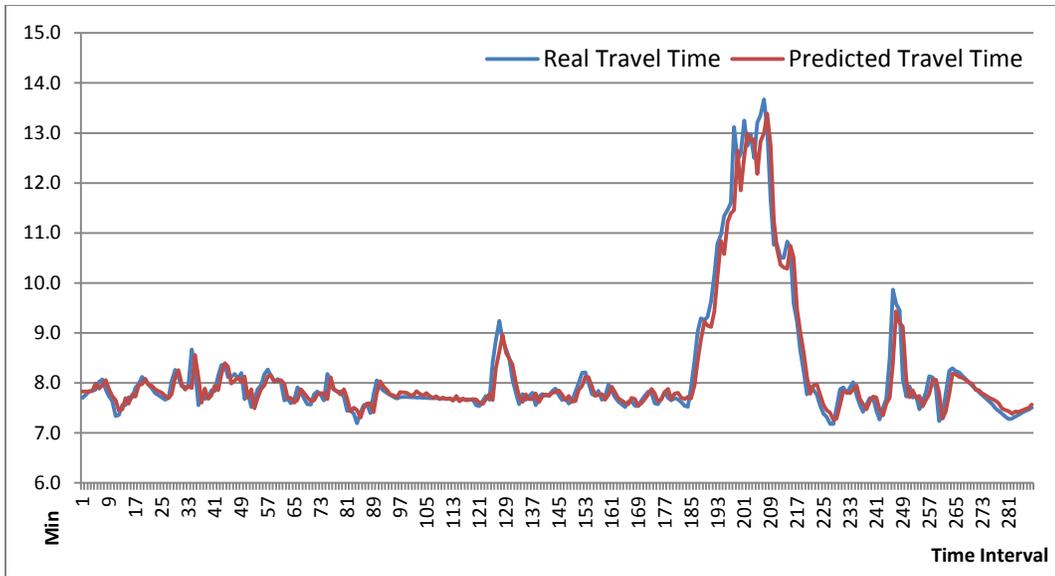
Figure 28 Comparison of Real and Predicted Travel Times-one day-08/16/2010
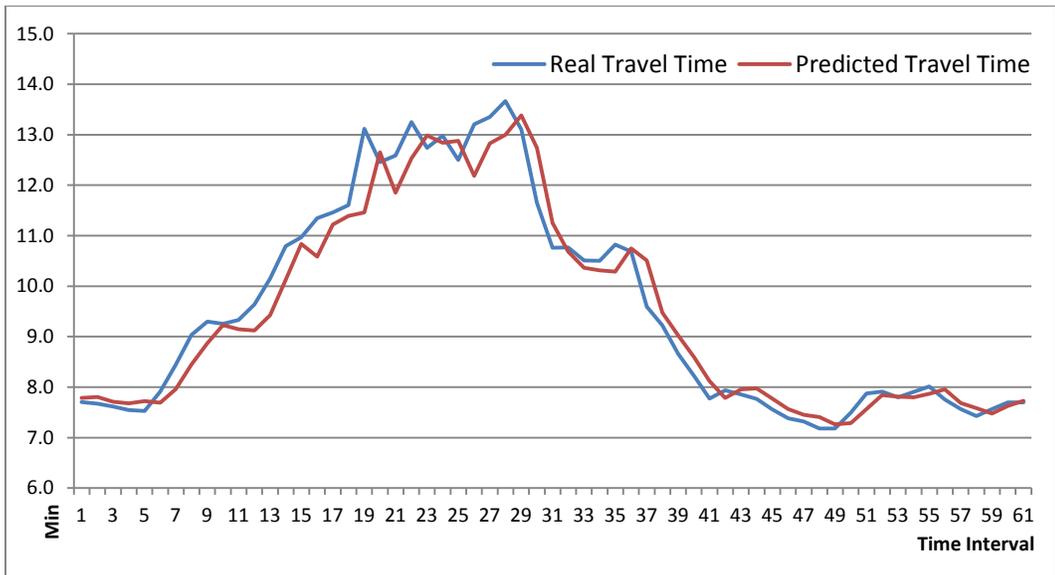
Figure 29 Comparison of Real and Predicted Travel Times-peak hour-08/16/2010

110
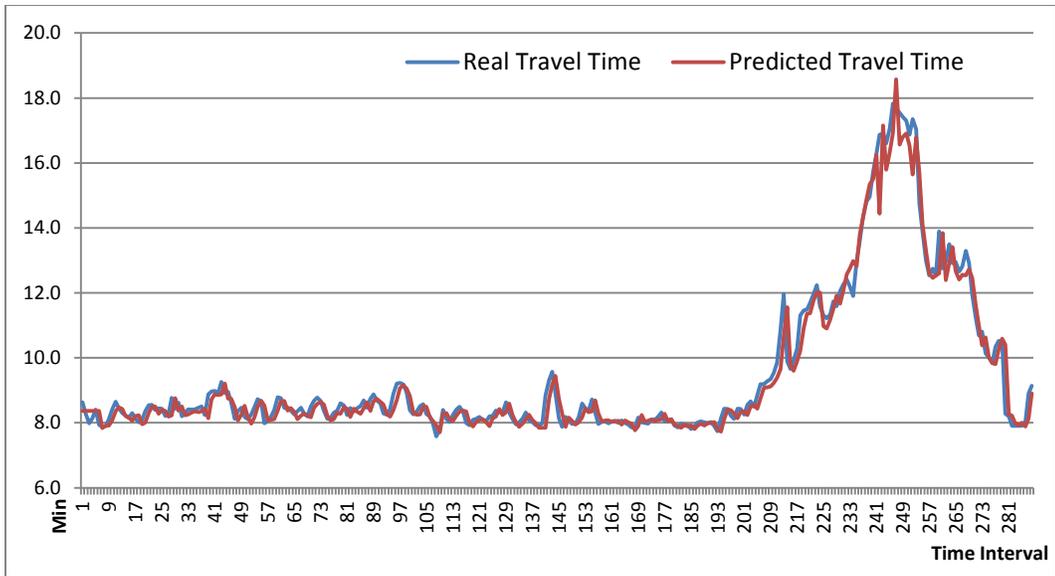
Figure 30 Comparison of Real and Predicted Travel Times-one day-02/09/2010
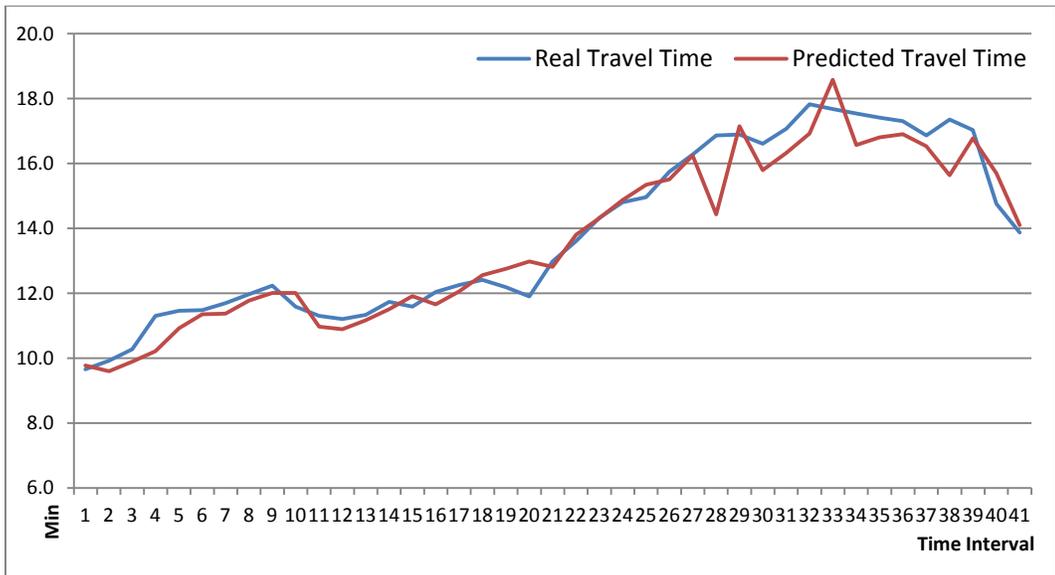
Figure 31 Comparison of Real and Predicted Travel Times- peak hour -02/09/2010

111
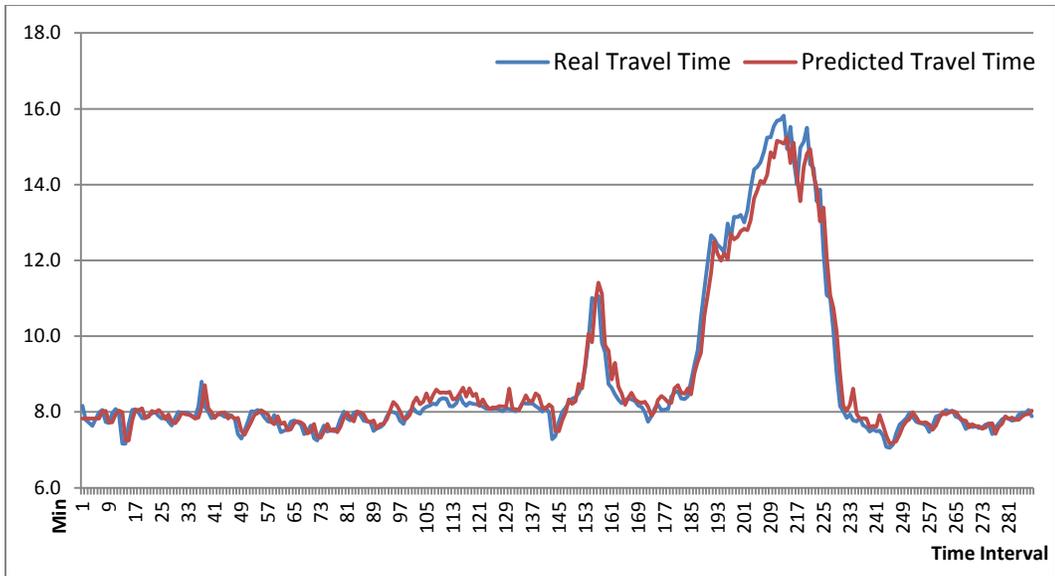
**Figure 32 Comparison of Real and Predicted Travel Times-one day-10/14/2010**
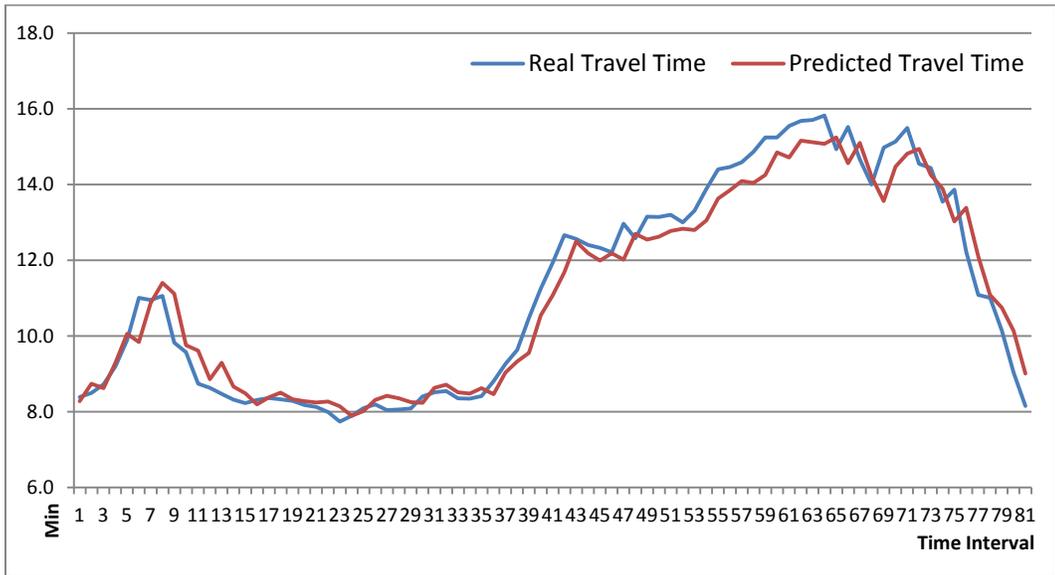
Figure 33 Comparison of Real and Predicted Travel Times- peak hour -10/14/2010

112
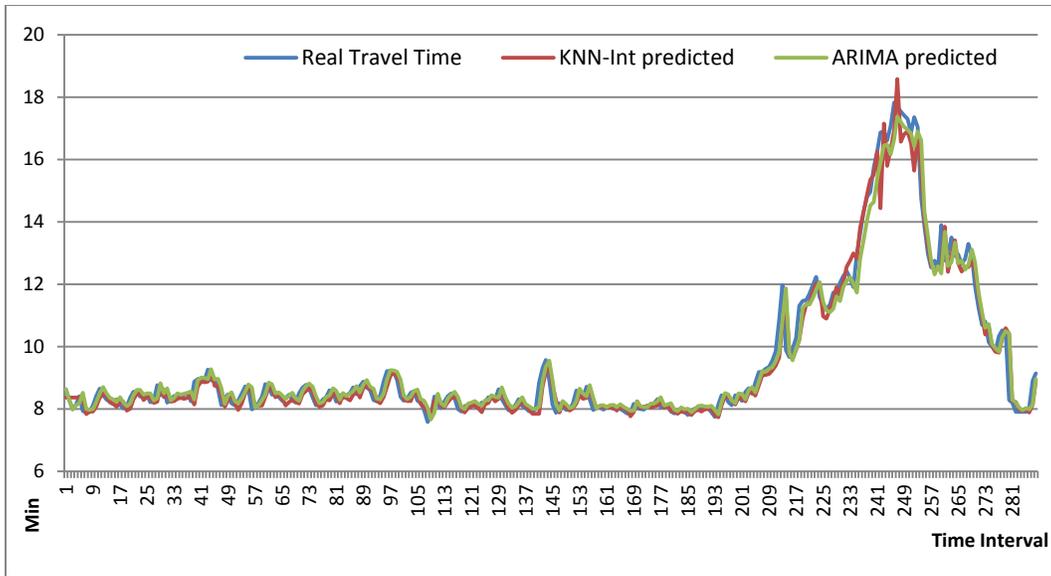
Figure 34 Comparison of KNN-Int and ARIMA Predicted Travel Times- one day -02/09/2010
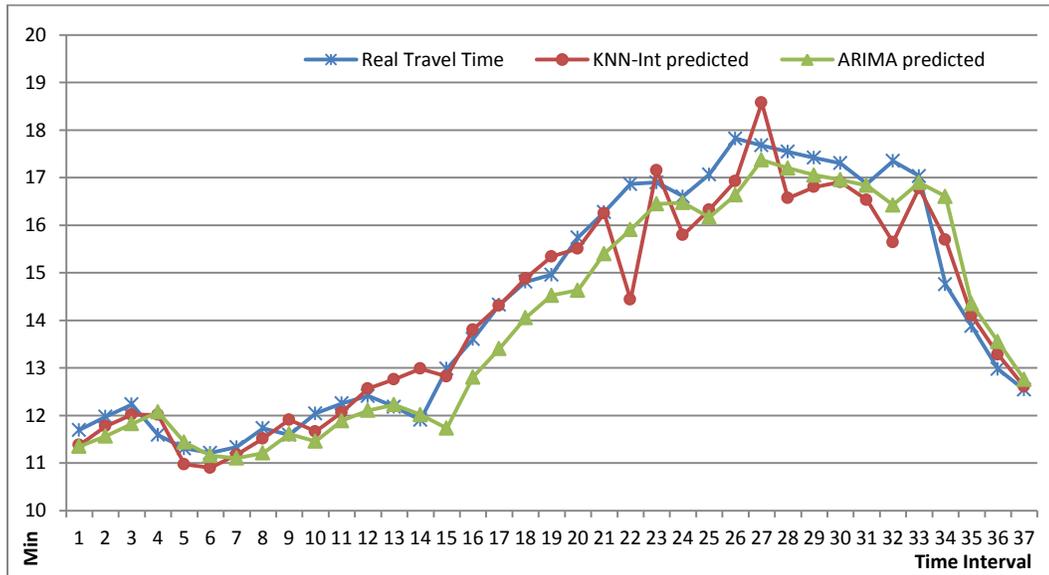
Figure 35 Comparison of KNN-Integrated and ARIMA Predicted Travel Times- Peak hour -

02/09/2010

In Figure 34, over the one day period, both models perform well. However, during the peak hour period, when traffic starts to build up under the snow condition, KNN-Integrated model exhibits better performance in prediction accuracy compared to the ARIMA model, especially during the transition period, as shown in Figure 35. Figure 35 presents the prediction results for the peak hour period. As seen from the figure, when travel time starts to increase from time interval 15 and continues increasing for about 35 minutes, KNN-Integrated provides high prediction accuracy compared to the real time data, while ARIMA exhibits an obvious time lagging effect. Similarly, when traffic congestion starts to relieve and travel time starts to reduce at time interval 33, KNN-Integrated model fits the real time data well and outperforms the ARIMA model during this 25 minutes transition period.

Furthermore, in order to investigate the performance of the proposed model under congested conditions, the MAPEs of the KNN-Integrated and ARIMA models are computed over the selected 30 day sample period for the individual time intervals with longer travel times. The threshold between normal traffic condition and congested condition is set at 16 minutes, which is twice the travel time under free flow condition on this freeway path (approximately 8 minutes). The results indicate that the MAPEs of KNN-Integrated and ARIMA models under congested conditions are 14.66% and 14.73% respectively, indicating that the KNN-Integrated model has higher prediction accuracy during congested period compared with the ARIMA model.

To summarize, in this section, first, the prediction performances between KNN-Integrated one shot and KNN-Integrated recursive models were compared. Then, the test results from 30 selected days were presented to compare the model performances. Last, three typical days including rain and snow days were selected to present the differences between the predicted travel times and the real travel time data. Based on the comparisons and results from the t-tests, the comparisons of model performance results indicate that the KNN-Integrated one shot model outperforms the KNN-Integrated recursive model both in terms of accuracy and computation time. From the test results comparisons in the average MAPE and the variance of MAPE, along with the comparison results from t-test, the proposed extended KNN-Integrated model outperforms the other models, ARIMA, KNN and KNN-T, both in prediction accuracy and reliability. For the 5 minutes prediction time interval, the extended KNN-Integrated model decreased the MAPE of ARIMA model by approximately 22.4% for all day period and 9.4% for peak hour period. For 30 minutes prediction time interval, the KNN-Integrated model decreased the MAPE of ARIMA model by approximately 35.6% for all day period and 9.1% for peak hour period.

The predicted travel times from KNN-Integrated are very close to the real travel times and the proposed model performs well under both normal and rain/snow weather scenarios. During the peak hour period, when traffic starts building up under the snow weather, KNN-Integrated model exhibits much better performance in prediction accuracy compared to the ARIMA model, especially during the two transition periods when traffic starts to build up and decrease. In conclusion, the proposed extended

KNN-Integrated model provides satisfactory prediction results through various performance tests.

5.2.3 Model Parameter Setting Guideline

In this part, a general guideline for the KNN-Integrated model's parameter selection is presented. In the extended KNN-Integrated model, there are four parameters required: (K, T, α, β), where K is the number of nearest neighbors selected, T is the number of previous continuous time intervals selected and α, and β are the weighting parameters for travel time and traffic volume, respectively. The KNN-Integrated model calibrates the optimal set of these parameters for each day and uses this set for the travel time prediction. However, for the practitioners, a general guideline for the parameter set selection is needed and will benefit greatly in the practical application of the travel time prediction model. As a result, a few pre-determined parameter sets corresponding to several traffic and weather conditions are given and the prediction results from the pre-set parameter sets are presented here. The following procedure is applied to establish these pre-determined parameters.

Step 1: To provide the pre-determined parameter sets, first, all of the 30 days selected in the previous section are listed with their best 6 sets of parameters that provide the lowest MAPEs for each day. Then, for each day, one set of parameters is generated based on the top 6 parameter sets by selecting the most frequently appeared parameters.

Step 2: Based on the parameter sets generated from the first step several traffic and weather conditions share the common parameter set and these scenarios are clustered into three groups: general condition, holiday and weekends, and snow condition. These 30 days are then classified into these three groups by their day and weather features.

Step 3: The most frequently appeared set of parameters is selected for each group. The pre-determined parameter sets for these three groups are listed in Table 34.

A similar approach is used for the peak hour period prediction. And the same three groups are identified with their pre-determined parameter sets listed in Table 35.

Table 34 Pre-set Parameter Sets for KNN-Integrated model – one day prediction

Parameter Set - one day	T	K	α	β
General condition	2	30	0.1	0.1
Holiday & weekends	2	30	0.2	0.7
Snow condition	4	10	0.1	0.2

Table 35 Pre-set Parameter Sets for KNN-Integrated model – peak hour prediction

Parameter Set - peak hour	T	K	α	β
General condition	2	10	0.1	0.1
Holiday & weekends	2	30	0.3	0.7
Snow condition	4	10	0.1	0.2

The prediction performances are tested on these 30 days with the pre-determined parameter sets given in the above tables, for both one day and peak hour periods. Tables 36 through 39 list the extended KNN-Integrated model prediction results with the calibrated parameter sets and the pre-determined parameter sets, respectively, for both one day and peak hour periods. The average MAPE for the 30 days are calculated and compared to see the differences in the model prediction accuracy between using calibrated and pre-determined parameter sets.

To summarize the results from the Tables 36 through 39, Table 40 lists the average of the MAPEs calculated from the 30 days with the calibrated parameter sets and the pre-determined parameter sets. The results for one day and peak hour periods are presented for each prediction time intervals. The percent differences between the prediction accuracy from the two sets of parameters are approximately 0.5% for one day prediction period and 1.2% for peak hour period. Figures 36 and 37 present the differences of average MAPE for 5 minutes prediction between the two sets of parameters of each day for both one day and peak hour periods. These comparison results present no significant difference in terms of accuracy between the two parameter sets and the pre-determined parameter sets are recommended to be used as the general guideline in practice for travel time prediction.

Table 36 Calibrated Parameter Sets for KNN-Integrated model – one day prediction MAPE

Date	5 min	10 min	15 min	20 min	25 min	30 min
01/21/2010	2.983	3.822	4.598	5.572	5.864	6.584
01/25/2010	1.871	2.493	2.807	3.033	3.332	3.599
02/09/2010	2.490	3.954	4.531	5.289	6.061	6.730
02/10/2010	7.900	11.682	13.782	14.528	14.746	15.515
02/19/2010	7.210	12.053	16.087	19.141	21.371	21.872
03/12/2010	4.428	7.185	8.697	9.863	10.595	10.977
04/08/2010	2.024	3.047	3.550	4.136	4.831	5.650
05/31/2010	2.054	2.873	3.203	3.264	3.393	3.437
06/29/2010	2.139	3.186	4.002	4.711	5.550	6.035
07/03/2010	1.756	2.550	2.726	2.839	2.905	2.839
08/16/2010	1.872	2.888	3.696	4.257	4.799	5.149
08/18/2010	3.669	6.134	7.959	9.078	9.872	10.699
09/06/2010	2.009	3.027	3.481	3.517	3.615	3.568
09/27/2010	4.349	6.777	7.850	8.722	9.239	9.733
09/30/2010	4.122	7.490	10.628	13.248	14.464	15.984
10/01/2010	2.015	3.138	3.738	4.394	4.986	5.451
10/06/2010	1.750	2.524	3.089	3.481	3.921	4.093
10/11/2010	2.522	3.939	4.893	5.733	6.151	6.466
10/14/2010	1.938	3.109	4.022	4.790	5.471	6.259
11/04/2010	1.769	2.742	3.352	3.932	4.409	4.875
11/20/2010	1.657	2.482	2.782	2.876	2.886	2.941
12/16/2010	12.353	20.690	25.779	27.715	28.763	30.169
12/19/2010	1.349	2.040	2.299	2.435	2.371	2.412
12/31/2010	1.620	2.547	2.877	2.965	3.059	3.171
01/22/2011	1.654	2.412	2.636	2.675	2.715	2.819
01/26/2011	8.562	12.103	13.948	14.794	15.736	16.896
02/22/2011	6.614	10.487	12.926	14.064	15.055	15.901
05/24/2011	1.519	2.493	3.222	3.987	4.649	5.332
06/15/2011	1.586	2.269	2.787	3.085	3.404	3.712
07/04/2011	1.602	2.460	2.768	2.857	2.869	2.933
Average MAPE	**3.313**	**5.153**	**6.290**	**7.033**	**7.569**	**8.060**

Table 37 Pre-set Parameter Sets for KNN-Integrated model – one day prediction MAPE

Date	5 min	10 min	15 min	20 min	25 min	30 min
01/21/2010	3.289	4.263	4.915	5.836	6.072	6.632
01/25/2010	1.968	2.661	2.983	3.163	3.516	3.804
02/09/2010	2.761	4.022	4.828	5.519	5.912	6.425
02/10/2010	8.936	12.260	14.444	14.654	15.054	15.617
02/19/2010	7.917	12.048	14.291	15.916	17.451	17.452
03/12/2010	4.550	7.483	9.049	10.100	10.843	11.333
04/08/2010	2.155	3.383	4.229	4.920	5.721	6.380
05/31/2010	2.147	3.038	3.487	3.694	3.926	4.059
06/29/2010	2.247	3.448	4.281	4.776	5.317	5.704
07/03/2010	1.835	2.652	2.849	3.000	3.055	3.054
08/16/2010	1.990	3.113	3.955	4.639	5.174	5.613
08/18/2010	4.042	6.506	8.186	9.275	10.007	10.380
09/06/2010	2.147	3.163	3.650	3.746	3.684	3.594
09/27/2010	4.671	7.053	8.519	9.942	10.753	11.166
09/30/2010	5.688	10.666	15.422	18.519	19.574	20.950
10/01/2010	2.220	3.526	4.349	4.972	5.628	6.016
10/06/2010	2.046	3.242	3.965	4.532	5.095	5.525
10/11/2010	2.858	4.405	5.327	5.965	6.302	6.643
10/14/2010	2.343	3.755	4.893	5.830	6.680	7.623
11/04/2010	1.996	3.008	3.822	4.460	4.938	5.464
11/20/2010	1.793	2.579	2.895	3.017	3.050	3.039
12/16/2010	12.967	21.825	27.896	29.114	31.274	32.184
12/19/2010	1.417	2.150	2.457	2.579	2.525	2.434
12/31/2010	1.748	2.557	2.882	3.042	3.146	3.246
01/22/2011	1.772	2.541	2.781	2.788	2.756	2.849
01/26/2011	8.913	12.621	14.633	16.001	17.417	18.401
02/22/2011	6.910	10.994	13.679	15.454	16.059	16.818
05/24/2011	1.802	3.043	4.146	5.231	6.109	6.862
06/15/2011	1.700	2.523	3.048	3.393	3.834	4.204
07/04/2011	1.737	2.606	2.956	3.005	2.908	2.921
Average MAPE	**3.619**	**5.571**	**6.827**	**7.569**	**8.126**	**8.546**

Table 38 Calibrated Parameter Sets for KNN-Integrated model – peak hour prediction MAPE

Date	5 min	10 min	15 min	20 min	25 min	30 min
01/21/2010	6.816	8.410	11.021	17.262	16.135	16.789
01/25/2010	2.006	3.096	3.622	3.507	3.371	3.781
02/09/2010	2.416	3.930	5.013	5.986	7.023	8.195
02/10/2010	9.401	12.572	13.291	14.669	14.919	14.597
02/19/2010	19.922	34.326	47.583	57.526	55.332	44.140
03/12/2010	12.299	20.537	25.211	28.995	32.828	36.097
04/08/2010	3.039	5.527	7.545	9.634	11.669	14.795
05/31/2010	2.642	3.447	4.530	5.562	7.107	9.841
06/29/2010	3.751	6.620	8.576	11.291	14.753	17.523
07/03/2010	1.924	2.825	2.994	3.172	3.011	2.903
08/16/2010	2.559	4.541	6.823	9.143	10.761	10.927
08/18/2010	9.566	17.298	22.961	26.716	30.312	33.929
09/06/2010	1.810	2.737	3.271	3.471	3.584	3.497
09/27/2010	2.633	3.916	4.421	5.556	6.597	7.185
09/30/2010	8.564	15.355	23.689	27.915	30.251	32.796
10/01/2010	3.325	5.510	8.144	9.902	12.256	14.360
10/06/2010	2.096	3.478	4.687	6.003	7.472	8.699
10/11/2010	2.541	4.233	5.436	6.387	7.146	7.972
10/14/2010	2.940	4.822	6.848	9.274	11.821	14.075
11/04/2010	2.323	4.266	6.353	8.399	10.265	11.344
11/20/2010	1.515	2.503	2.919	3.325	3.373	3.686
12/16/2010	25.605	40.394	48.734	53.425	51.132	50.629
12/19/2010	1.159	1.634	1.865	2.010	2.157	2.248
12/31/2010	2.100	3.100	3.414	3.758	3.471	3.527
01/22/2011	1.720	2.521	2.605	2.694	2.787	2.782
01/26/2011	8.562	12.103	13.948	14.794	15.736	16.896
02/22/2011	1.269	1.991	2.210	2.242	2.539	2.563
05/24/2011	2.705	4.962	7.160	9.062	11.077	13.071
06/15/2011	2.371	3.995	5.753	6.275	6.931	7.360
07/04/2011	1.884	3.298	3.830	3.946	3.903	3.948
Average MAPE	**5.049**	**8.132**	**10.482**	**12.397**	**13.324**	**14.005**

Table 39 Pre-set Parameter Sets for KNN-Integrated model – peak hour prediction MAPE

Date	5 min	10 min	15 min	20 min	25 min	30 min
01/21/2010	9.034	10.787	13.666	18.303	18.360	21.328
01/25/2010	2.325	3.671	4.206	4.371	4.551	5.222
02/09/2010	3.085	4.679	5.239	6.859	8.353	9.645
02/10/2010	11.071	14.155	14.320	14.995	14.452	14.709
02/19/2010	23.782	34.600	40.131	44.405	47.130	45.112
03/12/2010	12.574	21.384	26.134	29.994	33.299	35.944
04/08/2010	3.523	5.883	8.686	10.904	12.798	14.886
05/31/2010	2.738	3.561	4.020	4.134	4.678	5.436
06/29/2010	3.971	6.956	9.406	11.729	14.058	16.422
07/03/2010	2.089	2.763	2.935	3.104	3.055	3.152
08/16/2010	3.006	5.101	6.723	8.841	10.237	12.224
08/18/2010	11.253	19.309	25.569	29.797	33.074	35.180
09/06/2010	1.970	2.716	3.060	3.385	3.394	3.335
09/27/2010	3.094	4.725	5.459	6.035	6.331	6.503
09/30/2010	10.695	19.048	26.943	31.848	33.334	35.659
10/01/2010	4.265	7.203	9.167	11.532	13.595	15.145
10/06/2010	2.578	4.241	5.889	7.326	8.835	9.934
10/11/2010	4.748	7.930	9.888	11.189	12.153	12.847
10/14/2010	3.924	5.822	8.663	10.625	13.606	15.556
11/04/2010	2.702	4.934	7.389	9.889	12.168	13.824
11/20/2010	1.753	2.560	2.802	2.964	2.990	3.119
12/16/2010	31.215	44.519	51.876	52.536	55.335	56.731
12/19/2010	1.192	1.799	2.262	2.369	2.419	2.377
12/31/2010	2.279	3.209	3.659	3.771	3.756	3.793
01/22/2011	1.902	2.600	2.766	2.642	2.530	2.598
01/26/2011	16.680	24.613	31.170	30.438	28.605	31.006
02/22/2011	1.789	2.910	3.229	3.722	4.308	3.769
05/24/2011	3.445	5.606	8.129	10.891	13.200	14.593
06/15/2011	2.664	4.459	5.493	6.223	7.456	8.116
07/04/2011	2.120	3.206	3.473	3.794	3.895	3.954
Average MAPE	6.249	9.498	11.745	13.287	14.398	15.404

Table 40 Comparisons of average MAPE of Calibrated and Pre-set Parameter Sets for KNN-Integrated

|	5 min	10 min	15 min	20 min	25 min	30 min
One day average MAPE %						
Calibrated parameter set	3.313	5.153	6.290	7.033	7.569	8.060
Pre-set parameter set	3.619	5.571	6.827	7.569	8.126	8.546
Peak hour average MAPE %						
Calibrated parameter set	5.049	8.132	10.482	12.397	13.324	14.005
Pre-set parameter set	6.249	9.498	11.745	13.287	14.398	15.404
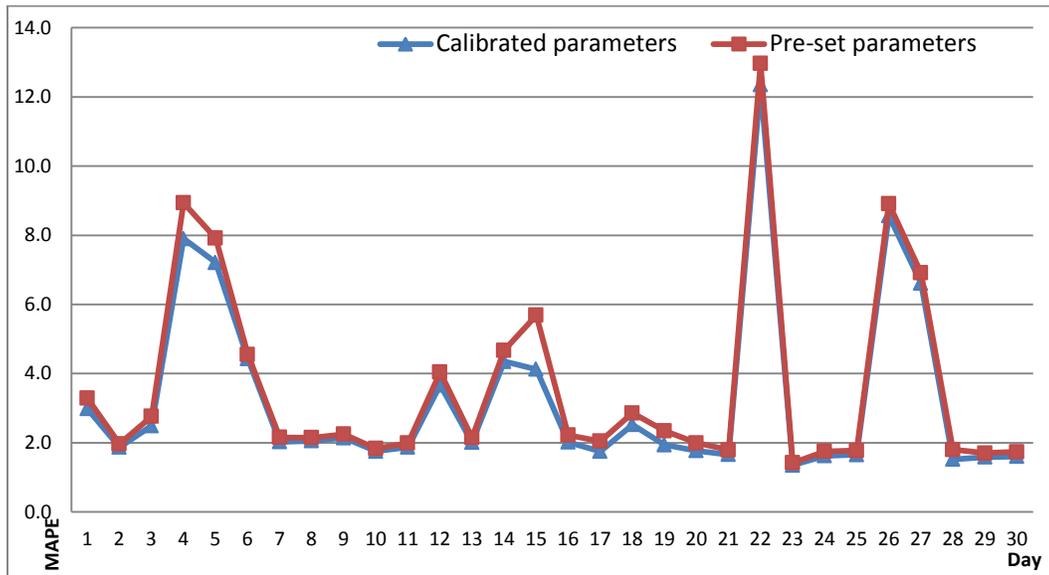
**Figure 36 Average MAPE of Calibrated and Pre-set Parameter Sets – one day**
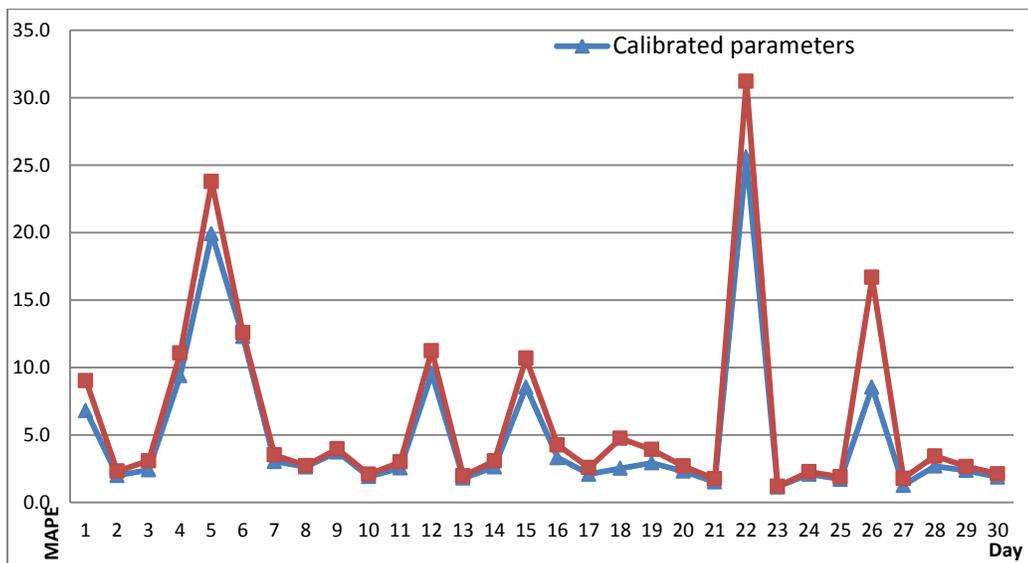
Figure 37 Average MAPE of Calibrated and Pre-set Parameter Sets – peak hour

In this section, a general guideline for the KNN-Integrated model's parameter set selection was presented. Pre-determined parameter sets for three groups of different traffic and weather conditions were provided for both one day and peak hour periods and the prediction results of the calibrated and pre-determined parameter sets by KNN-Integrated model were presented and compared. There is no significant difference between the prediction results from the two set of parameters, which indicates that the pre-determined parameter sets can be used as the general guideline and benefit in practical application of the travel time prediction model.

5.2.4 Model Sensitivity analysis on the efficient size of historical data

In this section, the efficient size of the historical database is studied and discussed through several prediction performance tests. Sensitivity analysis is conducted to evaluate the changes in prediction accuracy while varying the historical data sizes.

In the previous case studies, the historical dataset covers 19 months of traffic and weather data. However, a smaller size of the historical dataset may be sufficient to provide satisfactory prediction results, or even better prediction results. In addition, the computation time may reduce significantly in the prediction as well. To evaluate the efficient size of the historical database, the following durations of the historical data are tested by the extended KNN-Integrated model: 3 months, 6 months, 9 months, 12 months and 15 months. The same 30 days from the previous section are used for prediction and the average MAPE for each day and computation times for model calibrations are listed and compared for both one day and peak hour periods. The results are shown in Tables 41-44.

124

To better illustrate the effects of the historical data sizes on prediction accuracy and calibration computing time, Figures 38-41 present the total average MAPE of the 30 days for different historical data sizes and their calibration computing time for both one day and peak hour periods.

Based on the above test results, there is no significant difference in prediction accuracy while varying the historical database sizes. The 12 months of historical data size has the highest prediction accuracy for one day prediction period and for the peak hour prediction, the 9 months of historical data size has the highest accuracy.

Table 41 Comparisons of average MAPE for different historical data size –one day

Date	3 months	6 months	9 months	12 months	15 months
01/21/2010	3.124	3.080	3.102	3.079	3.070
01/25/2010	2.013	1.972	1.992	1.950	1.935
02/09/2010	2.571	2.584	2.536	2.530	2.562
02/10/2010	7.712	7.851	7.739	7.732	8.055
02/19/2010	7.167	7.569	7.358	7.437	7.369
03/12/2010	4.502	4.550	4.619	4.572	4.496
04/08/2010	2.705	2.159	2.166	2.137	2.144
05/31/2010	2.092	2.092	2.070	2.093	2.086
06/29/2010	2.236	2.223	2.194	2.190	2.164
07/03/2010	1.778	1.781	1.791	1.769	1.785
08/16/2010	2.015	2.040	2.024	2.015	1.993
08/18/2010	4.493	4.305	4.068	3.885	3.881
09/06/2010	2.114	2.084	2.050	2.056	2.055
09/27/2010	5.003	5.032	4.720	4.545	4.510
09/30/2010	4.284	4.611	4.394	4.378	4.316
10/01/2010	2.303	2.271	2.260	2.182	2.194
10/06/2010	2.207	2.109	2.014	1.977	1.967
10/11/2010	3.210	2.589	2.584	2.565	2.570
10/14/2010	2.610	2.609	2.231	2.231	2.218
11/04/2010	1.748	1.924	1.845	1.777	1.778
11/20/2010	1.731	1.746	1.750	1.734	1.738
12/16/2010	13.586	13.453	12.466	11.392	12.152
12/19/2010	1.457	1.410	1.387	1.358	1.366
12/31/2010	1.667	1.697	1.696	1.695	1.698
01/22/2011	1.733	1.715	1.721	1.705	1.700
01/26/2011	10.267	9.812	10.441	8.864	8.852
02/22/2011	6.904	6.855	6.853	6.853	6.643
05/24/2011	1.754	1.764	1.760	1.813	1.822
06/15/2011	1.721	1.722	1.708	1.708	1.686
07/04/2011	1.642	1.645	1.655	1.655	1.655
Average MAPE	**3.612**	**3.575**	**3.506**	**3.396**	**3.415**

Table 42 Comparisons of average MAPE for different historical data size –peak hour

Date	3 months	6 months	9 months	12 months	15 months
01/21/2010	7.940	6.843	7.190	7.430	7.335
01/25/2010	2.285	2.176	2.173	2.144	2.108
02/09/2010	2.608	2.648	2.674	2.578	2.565
02/10/2010	9.613	10.397	9.802	9.655	9.576
02/19/2010	21.285	21.958	20.844	20.974	21.113
03/12/2010	12.576	12.650	12.623	12.535	12.578
04/08/2010	3.565	3.426	3.449	3.437	3.465
05/31/2010	2.664	2.585	2.649	2.608	2.573
06/29/2010	4.108	4.161	3.826	3.845	3.794
07/03/2010	1.972	2.018	2.016	1.984	2.016
08/16/2010	2.928	2.915	3.002	2.970	2.967
08/18/2010	11.033	10.788	10.866	10.914	10.773
09/06/2010	1.721	1.670	1.684	1.767	1.812
09/27/2010	2.883	3.130	2.955	2.822	3.033
09/30/2010	9.693	9.729	9.713	9.720	9.759
10/01/2010	3.820	3.894	3.850	3.932	3.903
10/06/2010	2.497	2.341	2.467	2.443	2.430
10/11/2010	2.786	2.759	2.792	2.854	2.568
10/14/2010	3.546	3.558	3.568	3.381	3.399
11/04/2010	2.232	2.381	2.443	2.461	2.591
11/20/2010	1.586	1.587	1.573	1.595	1.602
12/16/2010	27.607	26.865	24.545	25.457	24.239
12/19/2010	1.094	1.100	1.078	1.046	1.106
12/31/2010	2.143	2.088	2.073	2.078	2.150
01/22/2011	1.798	1.739	1.791	1.762	1.769
01/26/2011	15.860	15.666	15.669	15.196	18.335
02/22/2011	1.852	1.390	1.319	1.236	1.266
05/24/2011	3.251	3.348	3.339	3.128	3.189
06/15/2011	2.420	2.473	2.445	2.620	2.550
07/04/2011	2.012	1.966	1.949	1.945	1.945
Average MAPE	**5.713**	**5.675**	**5.545**	**5.550**	**5.617**

Table 43 Comparisons of calibration computing time for different historical data size –one day

Date	3 months	6 months	9 months	12 months	15 months
01/21/2010	24.7	52.1	79.4	103.3	124.6
01/25/2010	17.8	36.2	55.1	72.1	88.9
02/09/2010	26.9	57.9	89.4	116.6	139.6
02/10/2010	5.3	9.2	13.2	17.4	21.0
02/19/2010	30.3	64.6	99.4	129.4	155.6
03/12/2010	5.6	8.5	11.3	15.8	20.0
04/08/2010	29.1	61.3	94.1	122.5	148.0
05/31/2010	3.4	5.1	8.8	14.4	18.0
06/29/2010	59.8	106.9	167.4	218.1	262.5
07/03/2010	28.1	49.6	72.7	98.7	121.2
08/16/2010	64.8	117.5	167.0	217.5	262.3
08/18/2010	39.4	71.9	101.4	132.3	159.4
09/06/2010	4.2	6.7	9.1	14.9	18.6
09/27/2010	3.8	7.8	10.9	16.6	19.2
09/30/2010	2.5	3.9	5.2	7.0	8.3
10/01/2010	7.0	10.5	15.6	21.1	24.4
10/06/2010	8.7	12.3	17.3	21.4	24.7
10/11/2010	4.3	7.7	10.0	14.5	18.0
10/14/2010	36.1	77.2	108.4	145.8	175.3
11/04/2010	15.8	32.9	47.4	63.3	75.5
11/20/2010	26.2	50.9	76.7	99.0	122.4
12/16/2010	37.5	85.2	127.8	164.6	198.4
12/19/2010	25.4	51.2	77.1	98.1	120.9
12/31/2010	7.5	10.8	13.1	15.8	18.3
01/22/2011	27.6	49.5	75.9	100.0	121.3
01/26/2011	8.4	14.4	21.6	29.3	35.2
02/22/2011	8.7	19.9	26.8	36.8	41.9
05/24/2011	22.8	41.0	58.2	82.7	100.1
06/15/2011	59.8	108.6	155.0	222.6	272.0
07/04/2011	4.1	7.1	13.9	16.8	19.6
Average computing time (min)	**21.5**	**41.3**	**61.0**	**80.9**	**97.8**

Table 44 Comparisons of calibration computing time for different historical data size –peak hour

| Date	3 months	6 months	9 months	12 months	15 months
01/21/2010	7.2	15.5	24.2	31.3	37.7
01/25/2010	6.8	14.6	22.7	29.5	37.0
02/09/2010	2.6	5.1	7.8	10.2	12.2
02/10/2010	1.9	3.5	5.3	6.9	8.3
02/19/2010	5.6	11.7	18.0	23.4	28.4
03/12/2010	0.8	1.4	1.9	2.5	3.1
04/08/2010	5.1	10.2	15.7	20.5	25.1
05/31/2010	0.8	1.3	2.0	3.2	3.9
06/29/2010	11.9	21.1	33.1	43.0	51.8
07/03/2010	5.6	10.1	14.8	20.0	24.3
08/16/2010	13.1	23.6	33.6	43.7	52.6
08/18/2010	8.2	15.1	21.3	27.7	33.2
09/06/2010	1.1	1.6	2.1	3.2	4.0
09/27/2010	1.1	2.2	3.0	4.3	5.1
09/30/2010	0.8	1.1	1.3	1.7	1.9
10/01/2010	1.8	2.7	4.1	5.4	6.2
10/06/2010	2.0	2.7	3.9	4.7	5.5
10/11/2010	1.1	1.8	2.2	3.1	3.9
10/14/2010	8.8	19.1	26.7	36.0	43.3
11/04/2010	5.7	12.5	18.1	23.9	28.6
11/20/2010	5.3	10.3	15.4	20.0	24.4
12/16/2010	6.8	15.6	23.3	30.0	36.2
12/19/2010	5.3	10.5	15.6	19.9	24.4
12/31/2010	1.6	2.0	2.7	3.2	3.7
01/22/2011	5.5	10.1	15.4	20.2	24.4
01/26/2011	0.4	0.7	0.9	1.4	1.6
02/22/2011	1.2	3.3	3.8	5.2	5.7
05/24/2011	8.1	15.1	21.6	30.7	37.4
06/15/2011	12.1	22.0	31.3	45.3	54.8
07/04/2011	1.0	1.6	2.9	3.4	4.0
Average computing time (min)	**4.6**	**8.9**	**13.1**	**17.4**	**21.1**
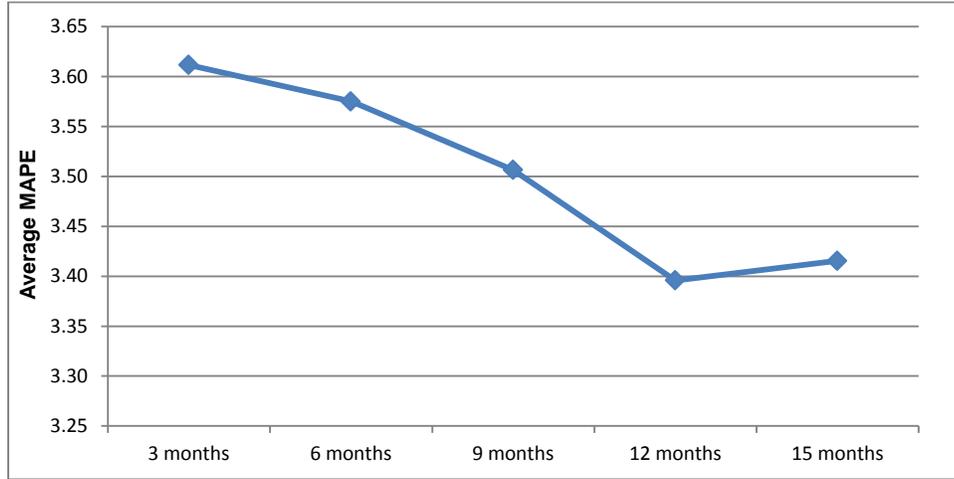
**Figure 38 Average MAPE for different historical data sizes– one day**
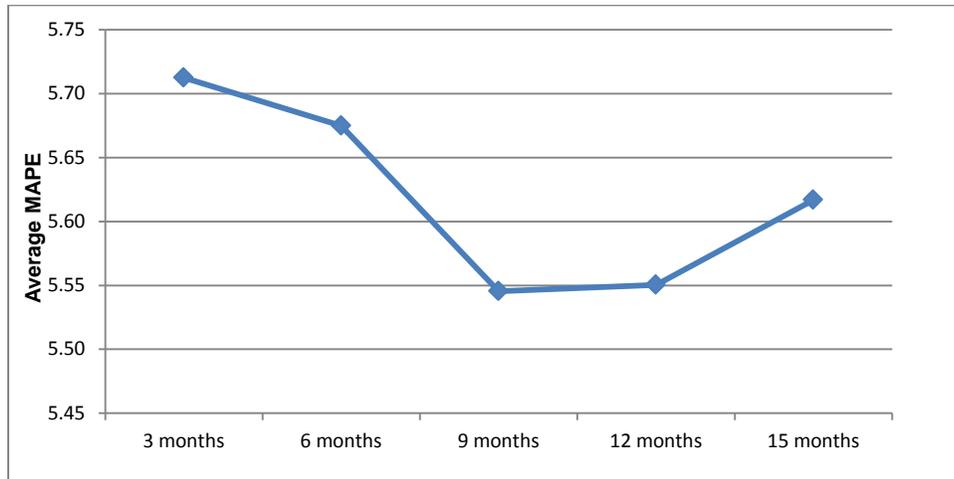
Figure 39 Average MAPE for different historical data sizes– peak hour
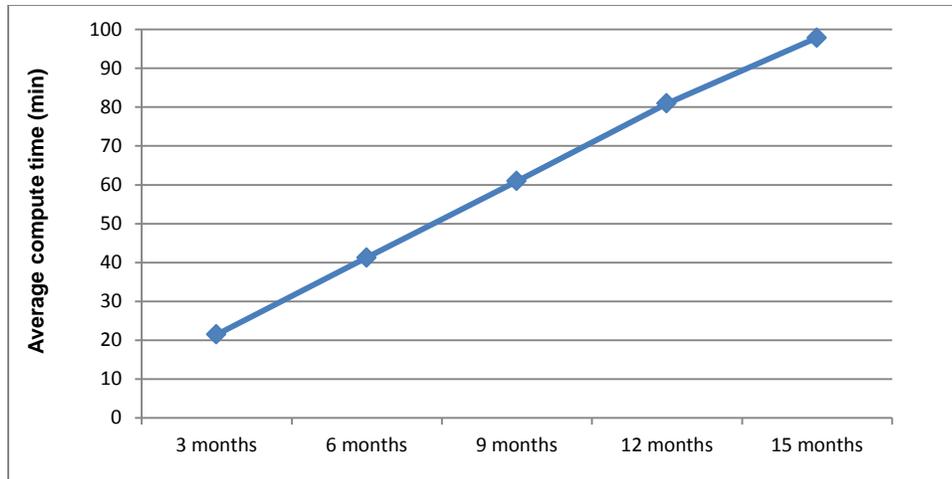
Figure 40 Average calibration computing time for different historical data sizes– one day
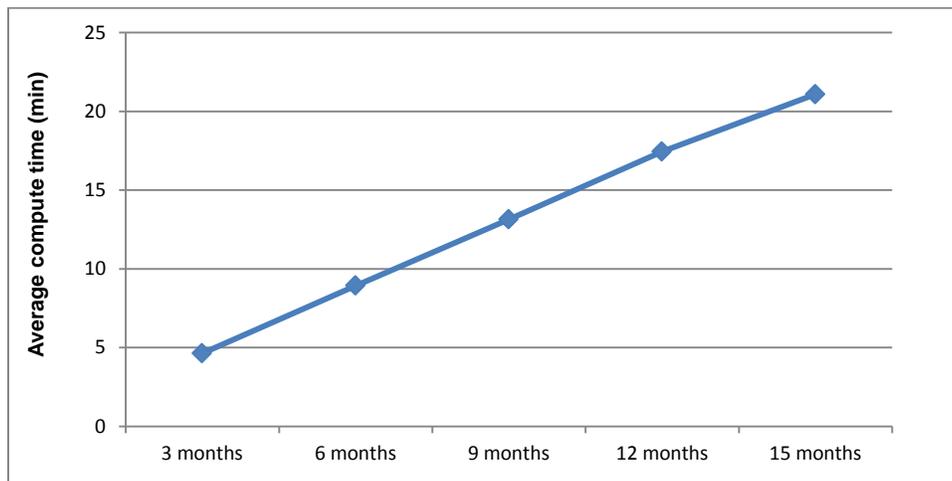

Figure 41 Average calibration computing time for different historical data sizes– peak hour

The variation ranges for the average MAPE over different historical database sizes are from about 3.3% to 3.5% for one day period prediction and 6.3% to 6.5% for peak hour period prediction. However, the computing times for model calibration have large variations over the different historical data sizes. For the one-day period prediction, the average calibration computing time grows with the increase of the historical data size linearly from 20 minutes to 100 minutes. Similarly, for the peak

hour period prediction, the average calibration computing time varies from 5 minutes to 20 minutes. Selecting the optimal historical data size provides slightly higher prediction accuracy at the cost of much higher computation time. As a result, 6 months to 9 months of historical dataset sizes are recommended for KNN-Integrated model calibration and prediction based on practical needs.

In conclusion, varying the historical data sizes from 3 months to 15 months does not indicate significant differences in prediction accuracy. However, large differences are presented in the calibration computation time while varying the data sizes. In general, 6 months to 9 months of historical dataset sizes are recommended for KNN-Integrated model calibration and prediction.

5.2.5 Conclusion

In this chapter, the travel time prediction model KNN-Integrated proposed in the previous chapter was extended with the ability to perform path travel time prediction and multi-step ahead travel time prediction. Prediction results from 5 minutes to 30 minutes ahead of time were investigated and discussed. A set of numerical experiments were conducted to evaluate the performance of the extended KNN-Integrated model. Comparison results of different models and their prediction performance were presented as well as the recommended pre-determined parameter sets. The results of a sensitivity analysis performed to examine the effect of the size of historical dataset were also presented.

First, three tasks on model performance test were completed and the test results were discussed.

(1) Model performance comparison between KNN-Integrated one shot and recursive models. The prediction comparisons results indicate that the KNN-Integrated one shot model outperforms the KNN-Integrated recursive model both in terms of accuracy and computation time.

(2) Model performance comparisons between KNN-Integrated and other models. The test results in the average MAPE and the variance of MAPE, along with the *t*-test results draw the conclusion that the proposed extended KNN-Integrated model outperforms the other models, ARIMA, KNN and KNN-T, both in prediction accuracy and reliability. For 5 minutes prediction time interval, the extended KNN-Integrated model decreased the MAPE of ARIMA model by approximately 22.4% for all day period and 9.4% for peak hour period.

(3) Model prediction results from typical weather conditions indicate that the predicted travel times from KNN-Integrated are very close to the real travel times and the proposed model performed well for both normal and rain/snow weather scenarios. KNN-Integrated model exhibits much better performance in prediction accuracy compared to the ARIMA model, especially during the transition periods when traffic starts to build up and decrease.

Second, a general guideline for the KNN-Integrated model's parameter set selection was recommended for practitioners. Pre-set parameter sets for three groups of different traffic and weather conditions were given for both one day and peak hour

periods respectively and the comparison of prediction results of the calibrated and pre-determined parameter sets indicate no significant difference in accuracy.

Last, the efficient size of the historical database is studied and the results show that varying the historical data sizes from 3 months to 15 months does not result in significant differences in prediction accuracy. However, large differences are presented in the calibration computation time while varying the data sizes. In general, 6 months to 9 months of historical dataset sizes are recommended for KNN-Integrated model calibration and prediction.

In conclusion, the extended KNN-Integrated model provides satisfactory prediction results in various performance tests and a general guideline for selecting parameter sets and the efficient size of historical data were provided in this chapter.

Chapter 6: Conclusions and Directions for Future Studies

6.1 Conclusions

This research utilized multiple data sources to predict travel times under various traffic and weather conditions, especially severe weather conditions. Since accurate travel time prediction under various weather conditions would help users to make better trip decisions, an integrated non-parametric short term travel time prediction model was proposed that incorporated the weather impacts on travel times based on a large historical traffic and weather information dataset, along with the available real time traffic information.

Literature indicated that non-parametric models work well under stochastic traffic conditions when a sufficient historical database is available, exhibiting advantages over parametric models especially under non-recurrent traffic conditions. Meanwhile, although there are studies on the quantified weather impact on traffic, incorporating weather impact in short term travel time prediction was much less explored.

In this study, three short term travel time prediction models for freeways were proposed: KNN-T, KNN-Integrated and the extended KNN-Integrated model. The model KNN-T enhanced traditional KNN model with trend adjustment where travel time trends were considered both qualitatively and quantitatively. KNN-Integrated further improved the KNN-T model by considering both stationary and non-stationary

traffic conditions. The extended KNN-Integrated model was enhanced by adding the features to perform path travel time prediction as well as multi-step ahead predictions. Compared with existing non-parametric models, the KNN-T model considered the change of trend of the historical travel times in the neighbors searching process. This modeling effort allowed the proposed model to capture not only the value but also the time-varying trend which would lead to a more precise match with the current traffic condition. The improvements in prediction accuracy of KNN-T were well supported by the tests results where the average MAPE of KNN-T decreased over 10% for all day and 20% for peak hours compared with ARIMA and Kalman filter models. As a result, both travel time values and their trends should be considered in the prediction process.

The KNN-Integrated model incorporated the weather impact as well as other important factors such as traffic volume, day features and incident occurrence into the model. Tests results indicated that this proposed model outperformed other models under inclement weather conditions since the changes of prevailing traffic conditions exhibited different patterns under various weather conditions. On average, KNN-Integrated model decreased the MAPE of ARIMA by approximately 8.0%. As a result, considering weather impacts is very important in the prediction process.

The extended KNN-Integrated model was proposed to predict path travel time and multi-step ahead travel times and test results indicated its better performances both in accuracy and reliability under various weather conditions, especially during transition

periods. The proposed model decreased the MAPE of ARIMA model by 22.4% for all day period and 9.4% for peak hour period.

In conclusion, KNN model is a generic non-parametric model that is widely used for prediction purposes in various areas. The accuracy of the KNN model relies heavily on the selection of the neighbors. This study improved the performance of KNN model for short term freeway travel time prediction by introducing the trend effects and weather impacts as discussed earlier. The improvements in prediction accuracy and reliability were well supported by the performance test results and the proposed models were not sensitive to the parameters, which made off-line calibration sufficient for real world applications.

6.2 Directions for Future Studies

This section presents directions for future research related to travel time prediction. In this proposed KNN–Integrated model, the travel time, volume, weather condition and incidents' occurrence on the target freeway path are used for prediction. However, including the volume and incident information from upstream or downstream segments may help improve the prediction accuracy, especially for longer prediction intervals such as 30 minutes ahead prediction.
Also, study on the impacts of traffic mix, such as the percentage of trucks in the mixed traffic, on travel time prediction may help improve the prediction.

The reliability of predicted travel time is another crucial issue along with the prediction of its mean value. The range of future travel times can be obtained by building appropriate confidence intervals around the predicted mean travel time. For example, when performing predictions, the standard deviation of the observed travel times from the previous five minutes can be used as the standard deviation for the future travel time value. Accordingly, a 95% confidence interval can be built based on the observed historical standard deviation.

The transferability of the model can be tested on other freeway locations where traffic data is available. Freeway segments with different geographical locations containing different characteristics can be selected to test the model's transferability. Studies may be conducted on how to use the historical data on a freeway segment and make the model more location independent with the purpose of reducing the model calibration efforts on each new location.

There are several systems and applications that may be studied for future research: (1) Develop an efficient and reliable traffic data filtering system. In this research, multiple data sources were used to form the historical dataset. However, not all of these traffic data are accurate due to the various reasons, for example, detector failure or error in transmission. Filtering the data outliers and dealing with the missing data are important in building a reliable historical database. As a result, a reliable traffic data processing system may be developed to provide accurate and reliable traffic datasets.

(2) Develop a travel time prediction model for applications on arterials with signalized intersections along the path. The proposed model in this dissertation may be modified to adjust to the prediction requirements on arterial segments.

(3) Develop an incident detection and duration estimation model. In this study, the main emphasis was on weather impact on travel time prediction. Incidents are another important impacting factor on the traffic condition on freeways. The prompt detection of an incident's occurrence and accurate estimation of the incident duration will help improve the reliability of the travel time prediction models greatly.

(4) Develop a location based travel time prediction system. In this study, the prediction models were proposed for applications on freeway segments. For future research, origin-destination based travel time prediction may be studied for individual road users' benefits.

(5) Develop a traveler's routing guidance system. One of the main purposes for travel time prediction is to provide information for trip users to make routing decision when there are alternatives. The estimated travel times are displayed on many dynamic messages signs along the freeway corridors, giving users updated travel time information for their en-route decision making. The study on routing guidance system will benefit from the predicted travel times for both pre-trip and en-route decision making, giving users the option to take an alternative route to reach destination faster.

(6) Develop a traffic management system using the predicted travel time for operation strategy making. The predicted travel times help traffic controllers to understand the near future traffic condition better and these information will facilitate the decision makers to prepare better control strategies for the entire road network.

Reference

2008 Status of the Nation's Highways, Bridges, and Transit: Conditions and Performance, Report to Congress, 2008, U.S. Department of Transportation, Federal Highway Administration and Federal Transit Administration.

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

Altman, N. S., (1992), "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression", The American Statistician, Vol. 46, No. 3, pp. 175-185

Bajwa, S. I., E. Chung, and M. Kuwahara. (2004), "An Adaptive Travel Time Prediction Model Based on Patter Matching." Presented at 11th Intelligent Transportation Systems World Congress, Nagoya, Japan.

Bajwa, S., Chung, E. and Kuwahara, M. (2003). "A travel time prediction method based on pattern matching technique." Proc. 21st ARRB Conference and 11th REAAA Conference, (CD), Cairns, Australia. (ARRB Transport Research: Vermont South, Victoria, Australia).

Ben-Akiva, M., Cantarellla, G., Cascetta, E., Ruiter, J., Whittaker, J., and Kroes, E., (1992) "Real-Time Prediction of Traffic Congestion", The Third international Conference on Vehicle Navigation and Information Systems. pp 557-562.

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

Butler S., Ringwood J. and Fay D. (2007) "Use of weather inputs in traffic volume forecasting", Submitted to ISSC 2007 Conference, Sept 13 – 14.

Byun, J., Daniel, J., Chien, S., (2010), "Speed-Flow Relationships Under Adverse Weather Conditions", 89th Transportation Research Board Annual Meeting, Washington, D.C., 2010

Cambridge Systematics and Texas Transportation Institute. Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation. Federal Highway Administration, 2005. ops.fhwa.dot.gov/congestion_report/congestion_report_05.pdf.

Cetin, M. and G. Comert, (2006) "Short-Term Traffic Flow Prediction with Regime-Switching Models", Transportation Research Record: Journal of the Transportation Research Board, No. 1965, pp 23-31.

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.

Chung E., Ohtani O., Warita H., Kuwahara M. and Morita H. (2006) "Does weather affect highway capacity?" Presented at the 5th International Symposium on Highway Capacity and Quality of Service, Yakoma, Japan.

Chung, E., O. Ohtani, and M. Kuwahara. (2005), Effect of rainfall on travel time and travel demand. 5th ITS European Congress, Hannover, Germany.

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.

Cools, M., E. Moons, and G. Wets. Assessing the Impact of Weather on Traffic Intensity. Transportation Research Board Annual Meeting. CD-ROM. Transportation Research Board of the National Academies, Washington, D.C., 2008, Paper No. 08-1903.

Daganzo C., (1997) "Fundamentals of Transportation and Traffic Operations," Pergamon-Elsevier, Oxford, U.K.

Datla, Sandeep, and Satish Sharma. "Impact of Cold and Snow on Temporal and Spatial Variations of Highway Traffic Volumes." Journal of Transport Geography 16.5 (2008): 358-72.

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

Disbro, J.E. and Frame, M. (1989). "Traffic flow theory and chaotic behavior." Transportation Research Record No. 1225, Transportation Research Board of the National Academies, Washington, D.C., pp. 109-115.

Dong, J., Mahmassani, H.S., Alfelor, R. (2010a). Incorporating Adverse Weather Impacts in Dynamic Traffic Simulation-assignment Models: Methodology and Application. Submitted to Transportation Research Board Annual Meeting 2010.

Dong, J., Mahmassani, H.S., Alfelor, R. (2010b), Weather Responsive Traffic Management: Deployment of Real-Time Traffic Estimation and Prediction Systems. Submitted to Transportation Research Board Annual Meeting 2010.

Edwards, J.B. The Relationship between Road Accident Severity and Recorded Weather. Journal of Safety Research, Vol. 29, No. 4, 1998, pp.249-262.

Farokhi, K. S., H. Masoud, and A. Haghani, (2010) "Evaluating Moving Average 26 Techniques in Short-Term Travel Time Prediction using an AVI Data Set", 27 Transportation Research Board Annual Meeting 2010, Washington 28 D.C.

Forbes, G. J., and Hall, F. L., (1990) "The applicability of catastrophe theory in modelling freeway traffic operations" Transportation Research Part A: General, Volume 24, Issue 5, Pages 335-344

Haghani, A., Hamedi, M., Sadabadi, K., Young, S. and Tarnoff, P. (2010) "Freeway Travel Time Ground Truth Data Collection Using Bluetooth Sensors", Presented at the 89th Transportation Research Board annual meeting, Washington, DC.

Highway Capacity Manual. TRB, National Research Council, Washington, D.C., 2000.

Hranac, R., Sterzin, E., Krechmer, D., Rakha, H., Farzaneh, M. (2006) Empirical Studies on Traffic Flow in Inclement Weather, FHWA-HOP-07-073

Huang S. and Ran B. (2003) "An application of Neural Network on traffic speed prediction under adverse weather condition", 82th Annual Meeting of the Transportation Research Board. Washington, D. C.

Ibrahim,A.T., and Hall,F.L., (1994). Effect of Adverse Weather Conditions on Speed-Flow-Occupancy Relationships. Transportation Research Record 1457, 184-191.

Jeffrey, D. J., Russam, K. and Robertson, D. I., (1987) "Electronic Route Guidance by AUTOGUIDE: the Research Background." Traffic Engineering and Control. pp. 525-529.

Jiang, X., Adeli, H., and Hon, M., (2005), "Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting", Journal of Transportation Engineering.

Juri, N. R., Unnikrishnan, A., and Waller S. T., (2007), "Integrated Traffic Simulation-Statistical Analysis Framework for Online Prediction of Freeway Travel Time", Transportation Research Record, Vol. 2039, Iss. 1, pp 24-31.

Kalman, R. E. (1960) "A New Approach to Linear Filtering and Prediction Problems," Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45.

Karlsson, M., and Yakowitz, S., (1987), Rainfall-runoff forecasting methods, old and new. Stochastic hydrology and hydraulics. Vol. 1, pp.303-318

Kaysi, I., Ben-Akiva, M. and Koutsopoulos, H. (1993). "An Integrated Approach to Vehicle Routing and Congestion Prediction for Real-Time Driver Guidance." 72nd Annual Meeting of the Transportation Research Board. Washington, D. C.

Keay, K. and I. Simmonds. (2006), Road Accidents and rainfall in a large Australian city. Accident Analysis and Prevention, Vol. 38, No. 3, pp. 445-454.

Kim, T., Kim, H., and Lovell, D. J.. (2005). "Traffic Flow Forecasting: Overcoming Memoryless Property in Nearest Neighbor Non-Parametric Regression." Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems Vienna, Austria, September 13-16, 2005.

Kwon J. and Petty K., (2005) "Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes", Transportation Research Record: Journal of the Transportation Research Board, No. 1935, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 147–153.

Kwon, J., B. Coifman, and P. Bickel. (2000), "Day-to-Day Travel-Time Trends and Travel-Time Prediction from Loop Detector Data." In Transportation Research Record: Journal of the Transportation Research Board, No. 1717, TRB, National Research Council, Washington, D.C., 2000,pp. 120–129.

Kwon, J., Coifman, B. and Bickel, P., "Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data," Transportation Research Record No. 1717, pp. 120-129, 2002.

Kyte, M., Khatib, Z., Shannon, P., and Kitchener, F. (2000) Effect of Environmental Factors on Free-Flow Speed. Transportation Research Circular, Proceedings of the Fourth International Symposium on Highway Capacity, Transportation Research Board, held in Maui, Hawaii, 108 – 119.

Lam, S. and Toan, T. D., (2008). "Short-Term Travel Time Prediction Using Support Vector Regression" 87th Annual Meeting of the Transportation Research Board. Washington, D. C.

Lamm, R., E. M. Choueiri, and T. Mailaender (1990). Comparison of Operating Speeds on Dry and Wet Pavements of Two-Lane Rural Highways, Transportation Research Record 1280, National Research Council, Washington D.C.

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

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

Liu, Y., Lin, P. W., Lai, X. R., Chang, G. L., Marquess, A., (2006), "Developments and Applications of Simulation-Based Online Travel Time Prediction System Traveling to Ocean City, Maryland", Transportation Research Record, No. 1959, Washington, D.C., 2006, pp. 92–104.

Maze,T.H., Agarwal,M., and Burchett,G.D., 2006. Whether Weather Matters to Traffic Demand, Traffic Safety, and Traffic Operations and Flow, Transportation Research Record 1948, 170-176.

Mulhern, F. J. and Caprara, R. J. (1994). "A nearest neighbor model for forecasting market response." International Journal of Forecasting, 10, pp.191-207.

Navin, F., (1986) "Traffic Congestion Catastrophes", Transportation Planning and Technology, Volume 11, Issue 1, pages 19 – 25.

OFCM. 2002b. (Office of the Federal Coordinator for Meteorological Services and Supporting Research), Weather Information for Surface Transportation National Needs Assessment Report. Silver Spring, MD: U.S. Department of Commerce/ National Oceanic and Atmospheric Administration.

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.

Park, D., and L. R. Rilett. (1998), "Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks." In Transportation Research Record: Journal of the Transportation Research Board, No. 1617, Transportation Research Board of the National Academies, Washington, D.C., pp. 163-170.

Pisano, A. Paul, Patrick J. Kennedy, B. C. Boyce, and A. D. Stern.(2008), Regional Demonstrations of the FHWA Clarus System. FHWA.

Qiu, L. and W.A. Nixon. Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis. In Transportation Research Record: Journal of the Transportation Research Board , No. 2055, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp 139-146.

Rakha, H., M. Farzaneh, M. Arafeh, and E. Sterzin. (2008), Inclement Weather Impacts on Freeway Traffic Stream Behavior. In Transportation Research Record: Journal of the Transportation Research Board , No. 2071, Transportation Research Board of the National Academies, Washington, D.C., pp 8-18.

Real-Time Traveler Information Systems, (2009), Transportation Research Board National Cooperative Highway Research Program (NCHRP) Synthesis 399.

Rice, J. and Zwet, E. (2004), "A Simple and Effective Method for Predicting Travel Times on Freeways", IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 3, September 2004.

Samba, D. and Park, B. (2010), Probabilistic Modeling of Inclement Weather Impacts on Traffic Volume , 89th Transportation Research Board Annual Meeting , Washington, D.C., 2010

Seasons of Achievement – Accomplishments of the Road Weather Management Program, 2010, Federal Highway Administration, Research & Innovative Technology Administration, FHWA-JPO-10-004

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

Stephanedes, Y. J., Michalopoulos, P. G. and Plum, R. A. (1981). "Improved Estimation of Traffic Flow for Real-Time Control." Transportation Research Record 795. pp.28-39.

Stern, A., Shah, V., Goodwin, L., and Pisano P. (2003). Analysis of Weather Impacts Flow in Metropolitan Washington D.C., Mitretek Systems, Washington, D.C.

Tanner, J. C. (1952). Effect of Weather on Traffic Flow, Nature, No. 4290.

Transportation Invest in Our Future – Future needs of the U.S. Surface Transportation System (2007), American Association of State Highway and Transportation Officials.

Tsirigotis L, Vlahogianni, E I, and Karlaftis, M. G. (2011). "Does Information on Weather Affect the Performance of Short-term Traffic Forecasting Models"? 90th Annual Meeting of the Transportation Research Board, January 22-28, Washington DC, US.

Tu, H., I.J.W.C. van Lint, H.J. van Zuylen. (2007), Impact of Adverse Weather on Travel Time Variability of Freeway Corridors. Transportation Research Board Annual Meeting. Transportation Research Board of the National Academies, Washington, D.C., Paper No. 07-1642.

Urban Mobility Report, 2010, Texas Transportation Institute.

van Hinsbergen, C. P. I., and J. W. C. van Lint. (2008), "Bayesian Combination of Travel Time Prediction Models." In Transportation Research Record: Journal of the Transportation Research Board, No. 2064, Transportation Research Board of the National Academies, Washington, D.C., pp. 73–80.

Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C., (2005), "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach", Transportation Research Part C 13 (2005) 211–234.

Wang, J., (2005) "Studies on Short-term Traffic Flow Forecasting Models and Methods". Doctor dissertation, Department of Civil Engineering, Tsinghua University.

Watkins, K. E. and Hallenbeck, M., (2010), Impact of Weather on Freeway Travel Times in the Rain City, TRB 2010 Annual Meeting, Washington D.C..

Williams, B. M. (1999) "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process." Ph.D. dissertation, Dept. of Civil Eng., Univ. of Virginia, Charlottesville, Virginia.

Williams, B.M., Durvasula, P.K. and Brown, D.E. (1998) "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," Transportation Research Record: Journal of the Transportation Research Board, No. 1644, TRB, National Research Council, Washington, D.C., 132-141.

Wu, C. H., Wei, C. C., Su, D. C., Chang, M. H. and Ho, J. M., (2003) "Travel Time Prediction with Support Vector Regression," in the Proceedings of IEEE Intelligent Transportation Systems Conference, October 2003.

Xiao, H., Sun, H. Y., and Ran, B. (2003) "Fuzzy-Neural Network Traffic Prediction Framework with Wavelet Decomposition" 82nd Annual Meeting of the Transportation Research Board. Washington, D. C.

Xie, Y., Y. Zhang, and Z. Ye (2006) "Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition", Transportation Research Board 85th Annual Meeting, Washington D.C..

Yang, F., Z. Yin, H. X. Liu, and B. Ran, (2004) "Online Recursive Algorithm for Short Term Traffic Prediction", Transportation Research Record No. 1879, pp. 1-8, Transportation Research Board.

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.

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, G., B. E. Patuwo, and M. Y. Hu. (1998) "Forecasting with Artificial Neural Networks: The State of the Art." International Journal of Forecasting, Vol. 14, No. 1, pp. 35-62.,

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

Zou, N., J. Wang, G. L. Chang, and J. Paracha. (2009) "Application of Advanced Traffic Information Systems: Field Test of a Travel Time Prediction System with Widely Spaced Detectors." In Transportation Research Record: Journal of the Transportation Research Board, No. 2129, Transportation Research Board of the National Academies, Washington, D.C., pp. 62-72.