Short-term Prediction of Freeway Travel Times Using Data from Bluetooth Detectors

by

Yaxin Hu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2013

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

There is increasing recognition among travelers, transportation professionals, and decision makers of the importance of the reliability of transportation facilities. An important step towards improving system reliability is developing methods that can be used in practice to predict freeway travel times for the near future (e.g. 5 – 15 minutes). Reliable and accurate predictions of future travel times can be used by travelers to make better decisions and by system operators to engage in pre-active rather than reactive system management.

Recent advances in wireless communications and the proliferation of personal devices that communicate wirelessly using the Bluetooth protocol have resulted in the development of a Bluetooth traffic monitoring system. This system is becoming increasingly popular for collecting vehicle travel time data in real-time, mainly because it has the following advantages over other technologies: (1) measuring travel time directly; (2) anonymous detection; (3) weatherproof; and (4) cost-effectiveness.

The data collected from Bluetooth detectors are similar to data collected from Automatic Vehicle Identification (AVI) systems using dedicated transponders (e.g. such as electronic toll tags), and therefore using these data for travel time prediction faces some of the same challenges as using AVI measurements, namely: (1) determining the optimal spacing between detectors; (2) dynamic outlier detection and travel time estimation must be able to respond quickly to rapid travel time changes; and (3) a time lag exists between the time when vehicles enter the segment and the time that their travel time can be measured (i.e. when the vehicle exits the monitored segment).

In this thesis, a generalized model was proposed to determine the optimal average spacing of Bluetooth detector deployments on urban freeways as a function of the length of the route for which travel times are to be estimated; a traffic flow filtering model was proposed to be applied as an enhancement to existing data-driven outlier detection algorithms as a mechanism to improve outlier detection performance; a short-term prediction model combining outlier filtering algorithm with Kalman filter was proposed for predicting near future freeway travel times using Bluetooth data with special attention to the time lag problem.

The results of this thesis indicate that the optimal detector spacing ranges from 2km for routes of 4km in length to 5km for routes of 20km in length; the proposed filtering model is able to solve the problem of tracking sudden changes in travel times and enhance the performance of the data-driven outlier detection algorithms; the proposed short-term prediction model significantly improves the

accuracy of travel time prediction for 5, 10 and 15 minutes prediction horizon under both free flow and non-free flow traffic states. The mean absolute relative errors (MARE) are improved by 8.8% to 30.6% under free flow traffic conditions, and 7.5% to 49.9% under non-free flow traffic conditions. The 90th percentile errors and standard deviation of the prediction errors are also improved.

Acknowledgements

I am lucky and honored to study and work under Professor Bruce Hellinga's supervision. Without his guidance and help this dissertation would not have been possible. I would like to express my sincere gratitude to him.

I appreciate the examining committee, Dr. Jeff Casello, Dr. Tarek Hegazi, Dr. John Zelek and Dr. Lina Kattan for their time and guidance towards improving this research.

I would like to acknowledge Mr. Robert Bruce from TPA North America for providing the sample of travel time data used in Chapter 3 of this dissertation

I am truly grateful to my friends and colleagues Tao Wei, Wenfu Wang, You-Jin Jung, Zeeshan Raza Abdy, David Duong, Mike Mandelzys, Mimi Ng, Soroush Salek Moghaddam, Reza Noroozisanani, Amir Zarinbal, Akram Nour, Pedram Izadpanah and Amir Hosein Ghods for their friendship and help during the past 5 years.

Finally, I would like to express my sincerest thanks to my husband Fei Yang, my parents Jirong Hu and Ruiping Tian for their love, encouragement and understanding. I wouldn't have been able to complete my PhD study without their support.

Dedication

This work is dedicated to my family, my father Jirong Hu, my mother Ruiping Tian, my husband Fei Yang and my daughter Lauren Yang for their love and sacrifice.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Ever increasing traffic demands makes the issue of traffic congestion a common concern for most urban areas throughout the world. A study entitled "The Cost of Urban Congestion in Canada" released by Transport Canada in 2006, indicates that the cost of recurrent urban congestion in Canada is between $2.3 billion and $3.7 billion per year (in 2002 dollar values). The report states that this is only a conservative estimate of the total cost of congestion as there isn't sufficient data to quantify the costs associated with non-recurrent congestion, air pollution, noise and stress. A more recent study reported by the Organization for Economic Co-operation and Development (OECD, 2010) estimates that traffic congestion in the Toronto area costs $3.3 billion in lost productivity annually.

Solutions to alleviating traffic congestion mainly focus on two aspects: (1) reducing traffic demand through demand management strategies, such as parking restrictions, road pricing, policy approaches to encourage more usage of public transport and so on; and (2) improving the transportation systems to increase capacity. Increasing road capacity by construction of new highways and expansion of existing highways is extremely costly and is often not feasible in urban centers with limited land resources. Consequently, advanced technologies and measures used to maximize the efficiency of the existing transportation systems become more and more attractive to traffic managers.

Those transportation management and control measures which use advanced technologies to improve transportation safety, mobility and enhance productivity are known collectively as Intelligent Transportation Systems (ITS). These technologies involve various sensing technologies, computer technologies, wire line and wireless communication technologies, electronic technologies, information processing technologies and so on, which enable the systems to sense, memorize, communicate, think (process information) and adapt (feedback information), and that is why we call it "intelligent".

Traffic monitoring technology is an integral part of ITS, because it provides important information about the operation of transportation systems and this information is essential for performance analysis, problem detection, management strategy implementation and traffic planning. Many monitoring technologies (or traffic sensors) have been developed for acquiring (directly or indirectly) traffic conditions, e.g. loop detectors, CCTV cameras, electronic toll tags, license plate recognition,

dedicated probes vehicles, cell phone probes, Bluetooth detectors, connected vehicles (a system still under development) (RITA ITS website,2013) and so on.

Data from different traffic monitoring systems have different forms and different characteristics, and these data have to be interpreted and transformed into valuable information that can be utilized by travelers and traffic managers. Travel time as a fundamental measure in transportation provides travelers and traffic managers with direct impression of traffic conditions, and it is used by travelers to plan trips (i.e. departure time, mode, and route), and used by traffic managers to manage transportation facilities and plan improvements. Automatic vehicle identification (AVI) systems (e.g. electronic toll tags, license plate recognition and Bluetooth detectors) can measure travel times directly and therefore the travel time data collected from AVI systems is expected more accurate than the travel times deduced from fixed-location detectors (e.g. loop detectors) (Haas et al., 2001).

Bluetooth traffic monitoring, as one of the AVI systems, is a relatively new technology, and it has gained popularity in the field of dynamic traffic data collection, mainly because it has the following advantages over other technologies: (a) measures travel time directly; (b) anonymous detection; (c) insensitive to weather conditions; (d) not installed in the road surface; (e) deployed and maintained easily, quickly and cost-effectively.

The Bluetooth technology has been implemented for monitoring traffic conditions in some places, e.g. the TranStar traffic monitoring center (Houston, USA) implemented this new low-cost traffic monitoring system in 2011 to obtain real-time traffic information, and this information is used for providing travelers with traffic information in various formats, including color-coded speed map on the Houston TranStar Website, travel time messages on roadside message signs, and traffic conditions reported through radio and television media (Houston TranStar Website, 2012). In 2012, a Bluetooth traffic monitoring system was deployed in the City of Calgary, Canada for monitoring real time traffic conditions and warning passengers and commuters about delays. The travel time data are also archived and used for planning purposes (City of Calgary Website, 2012).

However, most of these implementations provide only travel time estimation rather than travel time prediction. Travel time prediction has far greater value but also has many challenges that need to be addressed. The traffic information that is being disseminated in these implementations consists of some aggregation of individual vehicle travel times recently measured by the traffic monitoring system. That is to say, the information provided to motorists represents the traffic conditions from the

recent past, not the conditions that motorists will experience if they enter the road segment in the near future.

Therefore, the main goal in this research is to develop and evaluate methods to predict near-future travel times for freeways using data obtained from Bluetooth detectors.

1.1 Travel Time Estimation and Prediction

1.1.1 Definition of Travel Time

Travel time is an important system performance measure in the field of transportation. The actual travel time that an individual traveler takes to traverse a road segment is influenced by many factors such as traffic volume, weather conditions, behavior of drivers and vehicle characteristics. It is impossible to estimate or predict this stochastic process for all travelers exactly, consequently, the expected travel time is defined as the mean travel time during a specific time period.

Two types of travel time are defined in practice: arrival travel time (ATT) and departure travel time (DTT). ATT refers to the travel time measured after the vehicles have travelled through the entire road segment, while DTT is the estimated/predicted travel times that vehicles will experience if they enter the road segment now or at some specified future time. A time-space diagram (Figure 1.1) illustrates the definitions of ATT and DTT.

Figure 1.1 illustrates trajectories of several vehicles $(j, j+1, j+2, \dots)$ when they traverse a hypothetical road segment between two Bluetooth detectors i and $i-1$. The time taken for vehicle j to traverse the road segment i is denoted as $tt_{i,j}$ and is equal to $t_{i,j} - t_{i-1,j}$, where $t_{i,j}$ and $t_{i-1,j}$ are the times at which vehicle j was detected at Bluetooth detectors i and $i-1$ respectively.

The average arrival travel time ($\overline{\text{ATT}}$) for n_A vehicles traversing road segment i during time period k can be defined as follows:

$$\overline{ATT}_{i,k} = \frac{1}{n_A} \sum_{j=1}^{n_A} tt_{i,j}\,, \qquad t_{i,j} \in k \tag{1.1}$$

Where, n_A is the number of vehicles passing the downstream boundary of the road segment i during time period k. For example, in Figure 1.1, n_A is equal to 2, which refers to vehicles j and $j+1$.

The average departure travel time ($\overline{\text{DTT}}$) for n_D vehicles traversing road segment i during time period k can be defined as follows:

$$\overline{DTT}_{i,k} = \frac{1}{n_D}\sum_{j=1}^{n_D} tt_{i,j}\,, \qquad t_{i-1,j} \in k \tag{1.2}$$

Where, n_D is the number of vehicles passing the upstream boundary of the road segment i during time period k. For example, in Figure 1.1, n_D is equal to 3, which refers to vehicles $j+1$, $j+2$ and $j+3$.

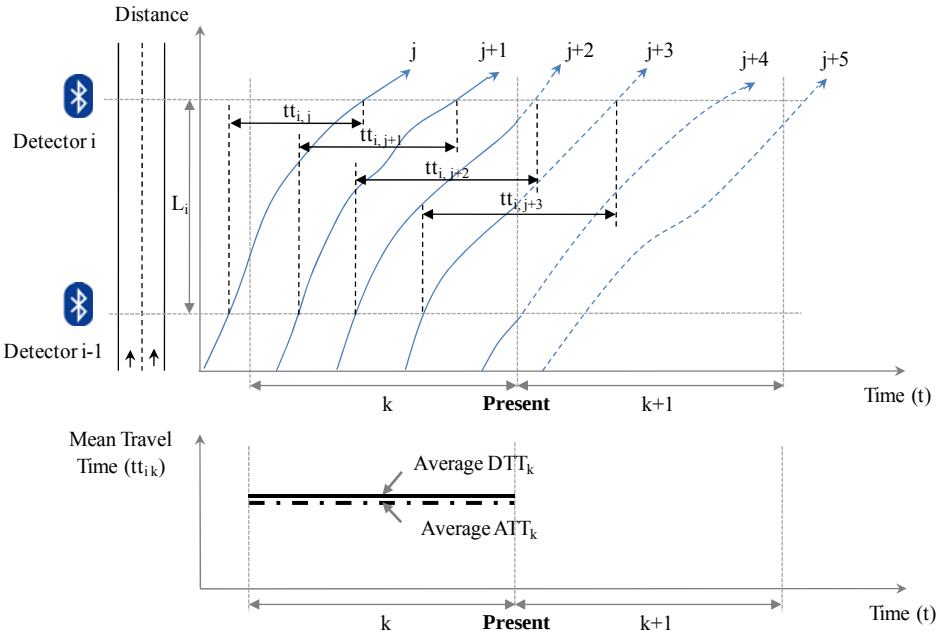


Figure 1.1: Illustration of travel time concepts based on a time-space diagram

The definitions (Equation 1.1 and Equation 1.2) imply that both $\overline{ATT}_{i,k}$ and $\overline{DTT}_{i,k}$ cannot be computed until all n_A or n_D vehicles have traversed the road segment. $\overline{ATT}_{i,k}$ can be computed in real time as the required data are available (i.e. vehicles j and $j+1$ have completely traversed the entire road segment before the present time). However, $\overline{DTT}_{i,k}$ cannot be computed in real time because the travel times experienced by some of the vehicles which enter the road segment during time period k may not be available yet as these vehicles have not yet travelled the entire road segment.

ATT can be of value for off-line analysis, but for real-time applications, such as posting travel time on variable message sign (VMS), the travel time of interest is the travel time that vehicles entering the segment during the given time interval will experience. Thus for these applications, DTT should be considered as the true travel time.

1.1.2 Concept of Travel Time Estimation and Prediction

Travel time estimation helps travelers and traffic managers to understand the current traffic conditions, and travel time prediction provides travelers and traffic managers with travel time information for vehicles that will traverse a road segment in the future. Figure 1.2 illustrates the concepts of travel time estimation and travel time prediction.

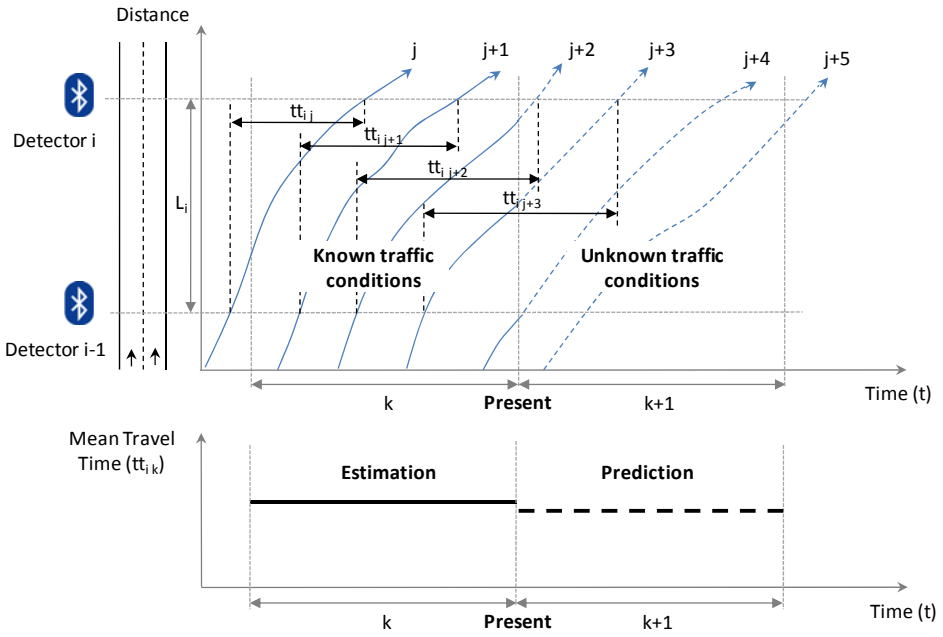


Figure 1.2: Concept of travel time estimation and prediction

Travel time estimation is a process of calculating (mean) travel time based on the "known" traffic conditions. Generally, the "known" conditions are various measures obtained from traffic monitoring systems. Data from different traffic monitoring systems have different forms and different characteristics, thus the method used to estimate travel time varies from system to system. This process of calculating the mean travel time based on the observed traffic data is a process of measuring travel times. Measuring travel times experienced by a sample of vehicles traversing a segment of roadway provides an estimate of the mean travel time experienced by the population of vehicles (i.e. entire traffic stream). However, by necessity this travel time is for the (recent) past and may not be a good estimate of future conditions. It is of far greater value to predict travel times in the future so that travelers and transport system managers can make informed decisions.

Travel time prediction is a process of estimating the travel time for the future and therefore a time period when traffic conditions are "unknown". Generally, the models of travel time prediction aim to

use previous "known" traffic condition information to forecast the traffic conditions (might not be the direct travel time) in the future. It is necessary to define how far into the future we are attempting to make the prediction, because different knowledge is required for short term prediction than long term prediction. The prediction horizon is defined as the length of time from present to a time in the future for which travel time is predicted.

Travel time prediction models are categorized into real-time travel time prediction, short term travel time prediction and long term travel time prediction based on the length of the prediction horizon (Van Lint, 2004). In real-time travel time prediction models, the prediction horizon is 0, but it is different from travel time estimation, because traffic conditions are unknown at the time when travel time is estimated (illustrated in Figure 1.3). In short-term travel time prediction models, the prediction horizon is greater than 0 but less than or equal to 60 minutes. The prediction horizon of long term travel time prediction models is typically longer than 60 minutes. In this research, we mainly focus on short-term travel time prediction for which the prediction horizon is in a range of 5 minutes to 15 minutes.

1.1.3 Control Strategy for Dynamic Travel Time Estimation and Prediction

Control strategy as one part of the dynamic travel time estimation and prediction models is very important, as it ensures the theoretical models can be applied in practice. Typically, the real time traffic data is transmitted via a high speed link to a database server at the transportation centre, and the database can be accessed by different users for different applications. For dynamic travel time estimation and prediction, especially a prediction model based on time-series traffic data, it is usually controlled by a rolling scheme on time scales. Figure 1.4 illustrates a typical control strategy for dynamic travel time estimation and prediction.

As illustrated in Figure 1.4, the rolling horizon is defined as the length of time in the past during which the traffic conditions observed affect the next prediction value. The rolling step is a time interval generated by dividing the rolling horizon into several parts, so that the average value for each part can be used for the time series model. The prediction horizon, as mentioned previously, is defined as the length of time from present to a time in the future for which travel time is predicted. The prediction step is defined as a time interval at which the prediction is updated.

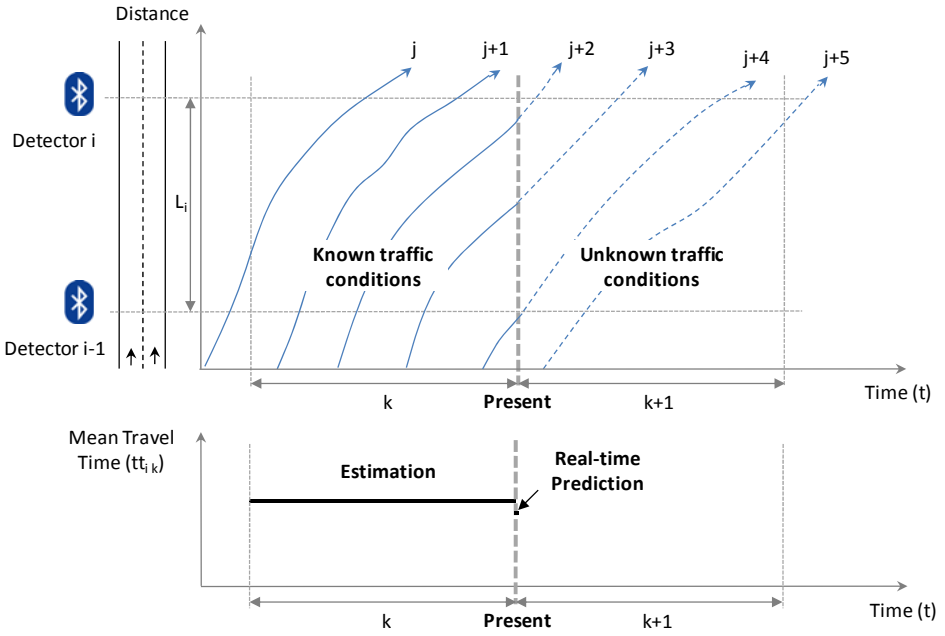


Figure 1.3: Difference between real-time travel time prediction and travel time estimation

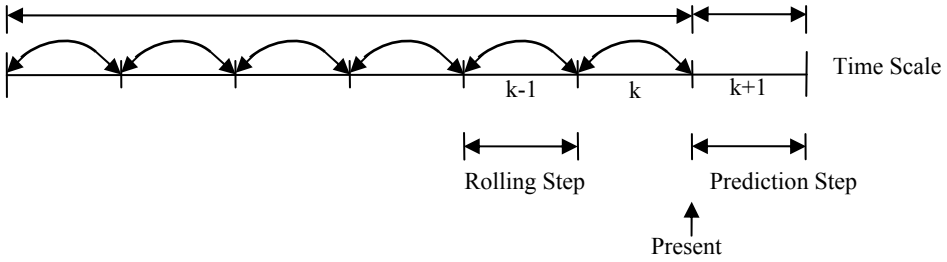


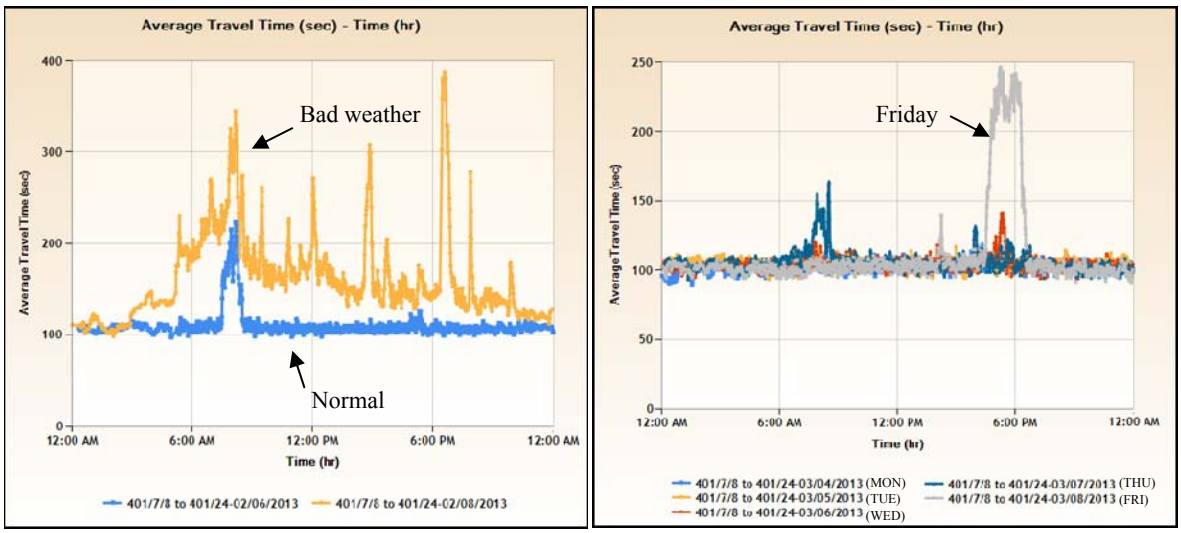
Figure 1.4: A typical control strategy for dynamic travel time prediction (Ishak et al., 2002)

In Figure 1.4, the rolling step, prediction step and prediction horizon are set to be equal, but in practice it is not necessary. The rolling step and prediction step can be set equal to the prediction horizon or set to be a value smaller than the prediction horizon (Ishak et al., 2002), which is determined according to the application requirements.

1.1.4 Travel Time Variability and Prediction Model Reliability

The time taken by vehicles to traverse a road segment is not constant, and it varies from hour to hour and from day to day. Well known factors that influence this variability in travel time include:

inadequate roadway capacity and traffic demand fluctuations, traffic incidents, work zones, weather, special events and traffic control devices (U.S. DOT, 2005).



(a) Different weather conditions

(b) Different traffic demands

Figure 1.5: Travel time variability (Waterloo traffic website 2013)

Data shown in Figure 1.5 are collected from Bluetooth detectors deployed along a 3.1 km section of suburban freeway (Highway 401 in Region of Waterloo, Canada). The variability of travel time caused by factors of weather conditions and demand fluctuations are shown in Figure 1.5 (a) and (b) respectively. From Figure 1.5 (a) we can see that in condition of bad weather, the travel time taken by vehicles to traverse a road segment is different from the travel time taken by vehicles to traverse the same road segment in normal weather condition, and the travel times that vehicles experienced during poor weather varies significantly with time of day. For the case shown in Figure 1.5 (b), we can see that travelers experienced more serious traffic congestion at Friday (pm peak) compared to other days in this week, and the main reason is that Friday pm peak has high traffic demand on this road segment than the same time period of other days.

It is the variability of travel time that makes travel time prediction challenging. Consequently, a reliable travel time prediction model is required. In the contest of this research, reliable is defined as accurate, robust and transferable. An accurate travel time prediction model is one that provides errors between the predicted travel times and "true" travel times that are less than some threshold. A robust travel time prediction model is one that performs acceptably over a range of traffic states (e.g.

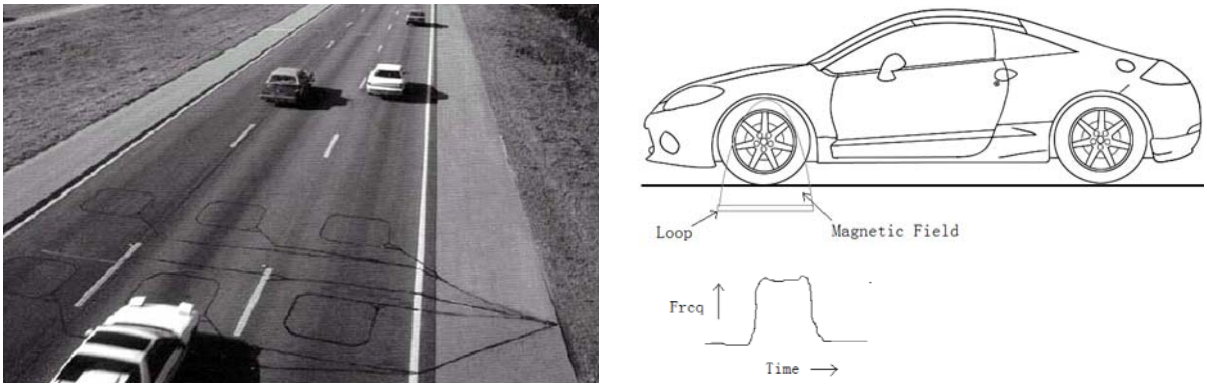
recurrent congestion and non-recurrent congestion). A transferable travel time prediction model is one that can be applied to any similar roadway and obtain accurate and robust results. (Izadpanah, 2007(a))

1.2 Methods of Collecting Real-time Travel Time Data

Three types of methods used in real-time travel time collection (direct/indirect) are introduced in this section: loop detectors, probe vehicles, and Bluetooth detectors. Loop detector is the traditional traffic monitoring technology and it is the most widely-used technology during the past several decades. Most of the previous travel time prediction models were developed based on data from loop detectors. Probe vehicle is another widely-used traffic monitoring technology, and it can be seen as the earliest wide-area wireless traffic monitoring technology. Bluetooth detector is one of the newest wireless traffic monitoring technologies, and data from Bluetooth detectors are the emphasis of this research.

1.2.1 Loop Detectors

Loop detectors are installed in the road surface (Figure 1.6(a)), and used for detecting the presence or passage of vehicles traveling along the roadway. The loop is a continuous run of wire with a magnetic field in the loop area. "When a vehicle passes over the wire loop or is stopped within the area enclosed by the loop, it reduces the loop inductance, which unbalances the tuned circuit of which the loop is a part. The increase in oscillator frequency is detected by the electronics unit and interpreted as a detected vehicle by the controller." (U.S. DOT, 2003) (Figure 1.6(b))



(a) (b)

Figure 1.6: Loop detectors (U.S. DOT, 2003)

data, namely spot speed and road section density. Spot speed is the traffic stream speed at a point or over a short roadway segment at a fixed location. The traditional method for estimating spot speeds

from single loop detectors is based on the assumption of a constant average effective vehicle length (Petty et al., 1998). And some studies made efforts to improve the speed estimation accuracy (Lin et al., 2004). Compared to the single loop detectors, double loop detectors can directly measure the spot speeds. For example, as the two loop detectors are very close to each other, the time mean speed between the two loop detectors can be seen as the spot speed at the location where the loop station (double loop detectors) is installed. In most cases, the data from loop detectors are aggregated over a defined polling period (typically 20 or 30 seconds), therefore the speed data from loop detectors are not individual vehicle speeds but the average speed of vehicles passing each loop station for each polling period.

Section density for a specific road segment during a defined time interval can be estimated based on the cumulative vehicle counts at the upstream and downstream loop detectors, and travel time can be calculated from the densities and interval volumes based on traffic flow theory.

1.2.2 Probe Vehicles

Probe vehicles are equipped with on-board electronics (such as a location and a communication device) acting as moving traffic detectors, participating in the traffic flow and are capable of determining experienced traffic conditions. The probe vehicle keeps track of its own geographic position by the equipped location device and transmits its traffic experiences (i.e. vehicle trajectory) via the communication device to traffic center. Continuous vehicle trajectories from sample vehicles are used to calculate the travel time taken by vehicles to traverse a specific road section.

Dedicated probe vehicles can be seen as the earliest wide-area wireless traffic monitoring technology, which might be the vehicles dispatched to the traffic stream for the purposes of data collection or those are already in the traffic network for other purpose (e.g. taxis, public transit buses or winter road maintenance vehicles) but can be used to collect traffic data.

The advantages of dedicated probe vehicles technology are: (1) the data collection process can be implemented in a large area; and (2) the real-time data can be automatically transmitted to traffic center for purpose of traffic control and management. The most evident disadvantage of dedicated probe vehicles technology is when using commercial fleets as probe vehicles the data may be biased towards specific driving styles, and when using test probe vehicles the data obtained may be limited as the number of probe vehicle is not sufficiently large.

In recent years, the anonymous cell phone tracking systems are used to extract traffic information. They are essentially probe vehicles in which drivers or passengers carry cell phones, and the cell phone can be anonymously tracked by wireless carriers with existing technologies. Taking advantage of the widespread using of cell phone communication system around the world, traffic data can be obtained at wide area scale without extensive instrumentations and at a low cost.

A new emerging technology called connected vehicles (i.e. vehicles equipped with wireless communication systems that will be able to communicate with roadside infrastructure and with each other) is expected to provide opportunities to collect traffic information in real-time. The connected vehicle technology (Figure 1.7) allows exchanging messages between vehicle's on-board equipment (OBE) and road side equipment (RSE), and therefore it has the potential to provide real-time traffic data if the connected vehicles are taken as probes in the traffic streams. However, the probe data collected by connected vehicles only provide partial vehicle trajectories rather than continuous vehicle trajectories covering the whole length of the road segment. (RITA ITS website 2013)

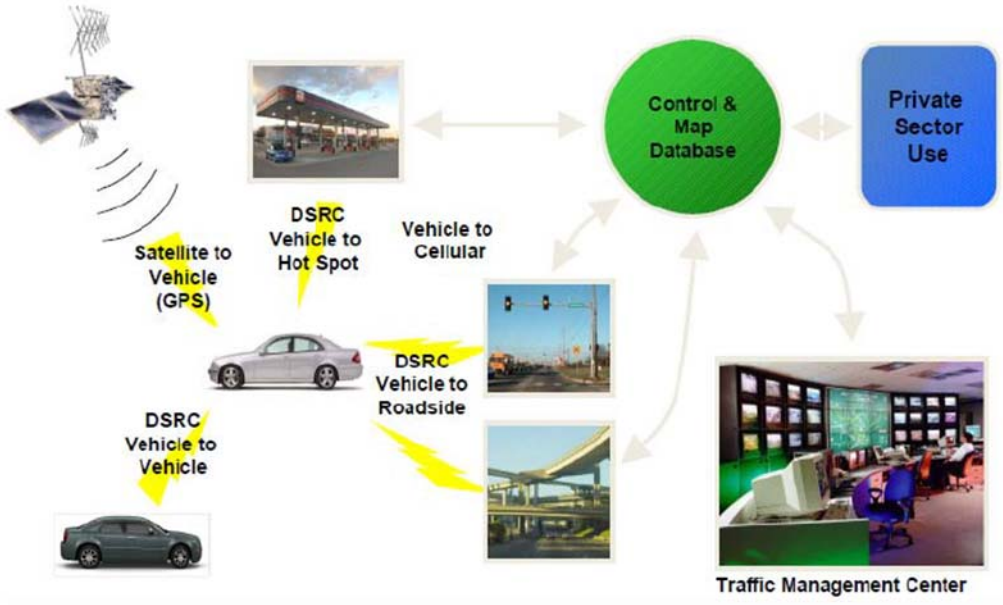


Figure 1.7: Connected vehicle environment (Pekilis, 2009)

1.2.3 Bluetooth Detectors

Traffic monitoring through Bluetooth detectors is a type of automatic vehicle identification (AVI) systems. The principles of Bluetooth traffic monitoring are similar as the technologies of radio frequency identification (RFID) and license plate recognition in traffic monitoring systems.

Bluetooth is a low powered, short range wireless communication technology with high transmission speed. Bluetooth was developed in the 1990s to replace wires for wired connection on electronic devices such as headsets for mobile phones. The prevalence of in-vehicle devices equipped with Bluetooth wireless communication technologies resulted in the development of this new technology for collecting traffic data, and this technology has become very popular in recent years.

The Bluetooth protocol uses a unique electronic identifier, called Machine Access Control address (i.e. MAC address) in each device so that electronic devices can be identified during data communications. It is these MAC addresses that are used as the basis for obtaining traffic information. MAC addresses in Bluetooth traffic monitoring systems are not directly associated with any specific user account or any specific vehicle. Consequently, using Bluetooth detectors alone, it is not possible to identify the attributes of a vehicle (e.g. year, make, color, license plate number) or the identity of a person. From this point of view, the Bluetooth traffic monitoring systems are anonymous. However, it is possible to re-identify a device at different points in the network and on different days.

The operational concept of collecting traffic data through Bluetooth detectors is illustrated in Figure 1.8.

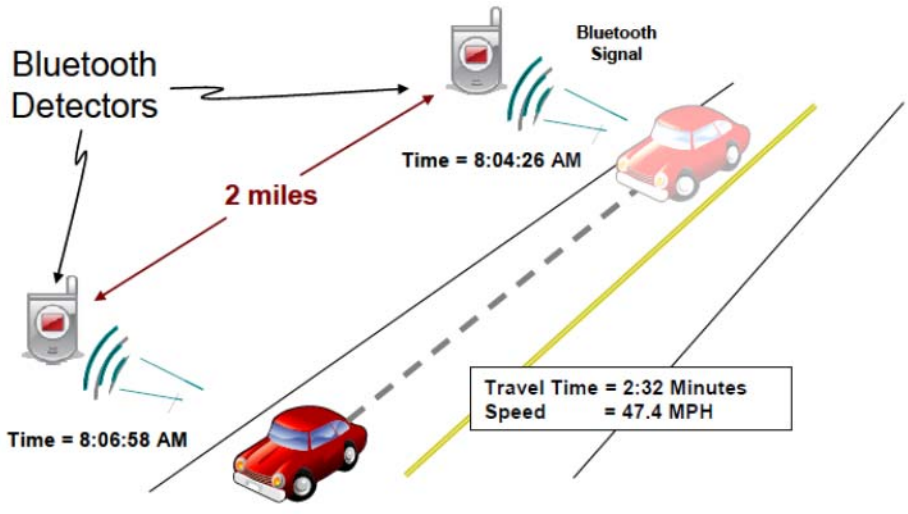


Figure 1.8: Operation Concept of Bluetooth Traffic Monitoring Systems (Haghani et al. 2010)

From Figure 1.8, we can see that when a vehicle containing a detectable Bluetooth device passes within the communication range (around 100m) of Bluetooth detectors (Haghani et al. 2010), it can be observed. If the vehicle is observed at two consecutive Bluetooth detectors, then travel time and the average speed for this vehicle over the road segment between these two detectors can be obtained. Obtaining similar data from more vehicles represents a statistical sample of the population of vehicles and provides an opportunity to estimate traffic conditions on this road segment. Experiments conducted by the University of Maryland (2008) have indicated that "approximately one in twenty vehicles contain a Bluetooth device that can be detected" indicating that a substantial fraction of the traffic stream can be monitored. In 2009, Ontario introduced a law prohibiting the use of hand-held devices while driving which resulted in an increase in the use of hands-free Bluetooth technology. (Ministry of Transportation Ontario website 2013)

The main advantages of the Bluetooth traffic monitoring system include:

- measures travel time over an entire road section;
- anonymous detection;
- monitoring traffic conditions for both travelling directions;
- insensitive to weather conditions;
- not installed in the road surface;
- deployed and maintained easily, quickly and cost-effectively.

Similar to data collected from other AVI systems (e.g. electronic toll tags, license plate recognition etc.), the travel time data collected by Bluetooth detectors typically contain outliers. These outliers represent measured travel times which are not representative of the traffic stream for which travel time measurements are desired. Outliers can arise from a number of sources including: (1) vehicles making an enroute stop or taking a detour between two consecutive Bluetooth detectors; (2) Bluetooth devices which are not within an automobile (e.g. the device may be in a public transit vehicle, on a pedestrian, cyclist, etc.); (3) vehicles in special purpose lanes; (4) vehicles on parallel roadways; (5) vehicles on off-ramps. The likelihood of the occurrence of outliers from a given source is a function of the roadway type (e.g. freeway vs. arterial), location of the Bluetooth detectors, traffic patterns, road network topology between the upstream and downstream Bluetooth detectors, etc. These outliers must be identified and removed before computing the mean travel time.

The data collected from Bluetooth detectors are transmitted wirelessly to a central server where the data processing (e.g. the matching, outlier detection, aggregation, and prediction) takes place, and the

information is used for traffic management applications (e.g. traveller information system or pre-active roadway management).

Based on communications with hardware system vendors, the approximate cost for deploying Bluetooth detectors is in the range of $5,000 to $10,000 per detector, depending on location, availability of power and communications, etc. In addition, for detectors equipped with wireless modems, there is a monthly fee for the wireless data transmission of approximately $25-$50.

1.3 Problem Statement

There is increasing recognition among travelers, transportation professionals, and decision makers of the importance of the reliability of freeway transportation facilities. An important step towards improving system reliability is developing methods that can be used in practice to predict freeway travel times for the near future. Reliable and accurate predictions of future travel times can be used by travelers to make better decisions and by system operators to engage in pre-active rather than reactive system management.

Over the past few decades, a considerable amount of work has been done on the subject of travel time prediction. However, most of this work has used data obtained from traditional traffic monitoring systems (e.g. loop detectors). Bluetooth traffic monitoring technologies provide the opportunity to collect wide area real-time travel time data with low cost, something that is not feasible with the traditional traffic monitoring technologies, but the data collected by Bluetooth detectors are different from the data collected by conventional sensors (e.g. loop detectors), and therefore new methods are required for predicting freeway travel times based on Bluetooth data. Although the core of the proposed prediction can be applied to any data source, the details of the prediction methods would likely change for different types of data.

As discussed previously, the travel time data collected by Bluetooth detectors typically contain outliers, and these outliers have to be identified before computing the mean travel time. Therefore, a dynamic filter is needed to remove those outliers. The challenge of detecting the outliers in real-time is distinguishing outliers from rapid changes in the underlying travel time. Existing real-time filtering algorithms (SwRI 1998; Mouskos et al. 1998; Dion and Rakha 2006) have limitations on responding to rapid fluctuations of traffic conditions, and consequently do not perform reliably when travel times are changing rapidly.

The travel time of an individual vehicle measured by Bluetooth detectors can only be obtained after the vehicle has passed through the entire road segment. Consequently there is a time lag that exists between the time when vehicles enter the segment and the time that their travel time can be measured. This is a commonly accepted limitation of AVI data, and is one of the challenges of using this type of traffic data for real-time travel time estimation/prediction (Waller et al. 2007, Chen and Chien 2001). This challenge becomes increasingly more difficult when the travel time between two successive AVI detectors becomes large (either because the AVI detectors are spaced far apart and/or because traffic is congested). A practical solution was suggested to divide the long freeway route into shorter segments to reduce the magnitude of the time lag and therefore increase the probability that the measured travel time from AVI systems can be used as a reliable estimate of true travel time. However, to-date, no study has quantitatively analyzed the optimal spacing between AVI detectors with respect to maximizing the accuracy of the real-time travel time estimations or predictions.

Historical data have long been considered as an important input to travel time prediction models because there is an expectation that time series of traffic state data collected from different days at the same site in the same situation have similar time-varying traffic patterns (Chen et al. 2012). The most common method of selecting historical data for travel time prediction is aggregating travel times from past consecutive days, while distinguishing between work days and weekends/holidays. This method is attractive because it is intuitive, simple to implement and easy to understand. However, this method can only provide a primary pattern, and the large variation in travel times within each day caused by variations in demand and on capacity.

In summary, the prediction of mean travel time on the basis of Bluetooth travel time data requires: (1) the Bluetooth detectors used to collect real-time travel time data are properly deployed (i.e. determining the optimal spacing between detectors); (2) a reliable real-time outlier detection algorithm; (3) an improved method of selecting historical data for travel time prediction; and (4) a method that is able to address the data gaps caused by the time lag inherent in the Bluetooth measurements. This research is focused on addressing the above four issues.

1.4 Research Goal and Objectives

The main goal of this research is to develop and evaluate a method which can be used for reliable prediction of near future travel times on freeways using data collected from Bluetooth detectors. An immediate application of the travel time predictions based on Bluetooth data is the posting of travel

times on overhead variable message signs or portable message signs for use in construction zones. Other potential applications include the use of these data for pro-active freeway traffic control, including variable speed limits, variable speed advisories, ramp metering, etc. It is anticipated that the required accuracy of the travel time predictions may vary depending on the application. However, there is currently no standard metric quantifying the accuracy of the predictions and there is no objective threshold to define acceptable versus unacceptable accuracy. Consequently, within this thesis, the objective is to develop a travel time prediction method that performs better than existing methods as measured by the mean absolute relative error and the 90th percentile error. The 90th percentile error is particularly important as we wish to avoid large travel time predictions errors, even if they are relatively infrequent (and may have very little impact on the mean error).

To achieve this goal the proposed research has the following objectives:

1. Determine the optimal spacing between Bluetooth detectors with respect to maximizing the accuracy of real-time travel time estimations and predictions.
2. Develop a dynamic filtering algorithm to address the problems of reliable estimation of travel times using Bluetooth data.
3. Improve the method of selecting historical data for travel time prediction.
4. Develop a model for predicting near future freeway travel times using Bluetooth data with special attention to the time lag exists between the time when vehicles enter the segment and the time that their travel time can be measured (i.e. when the vehicle exits the monitored segment).
5. Calibrate and validate the proposed prediction model.
6. Demonstrate the performance of the proposed model by comparing the application results of the proposed model to that obtained from existing methods.

The proposed prediction method is a dynamic adaptive traffic control system, and the framework of this system is shown in Figure 1.9. Based on this system framework, the dynamic outlier detection and travel time prediction are combined together, and these two models are operated recursively. A feedback control mechanism is used to update previous estimations before each prediction once new measurements are available.

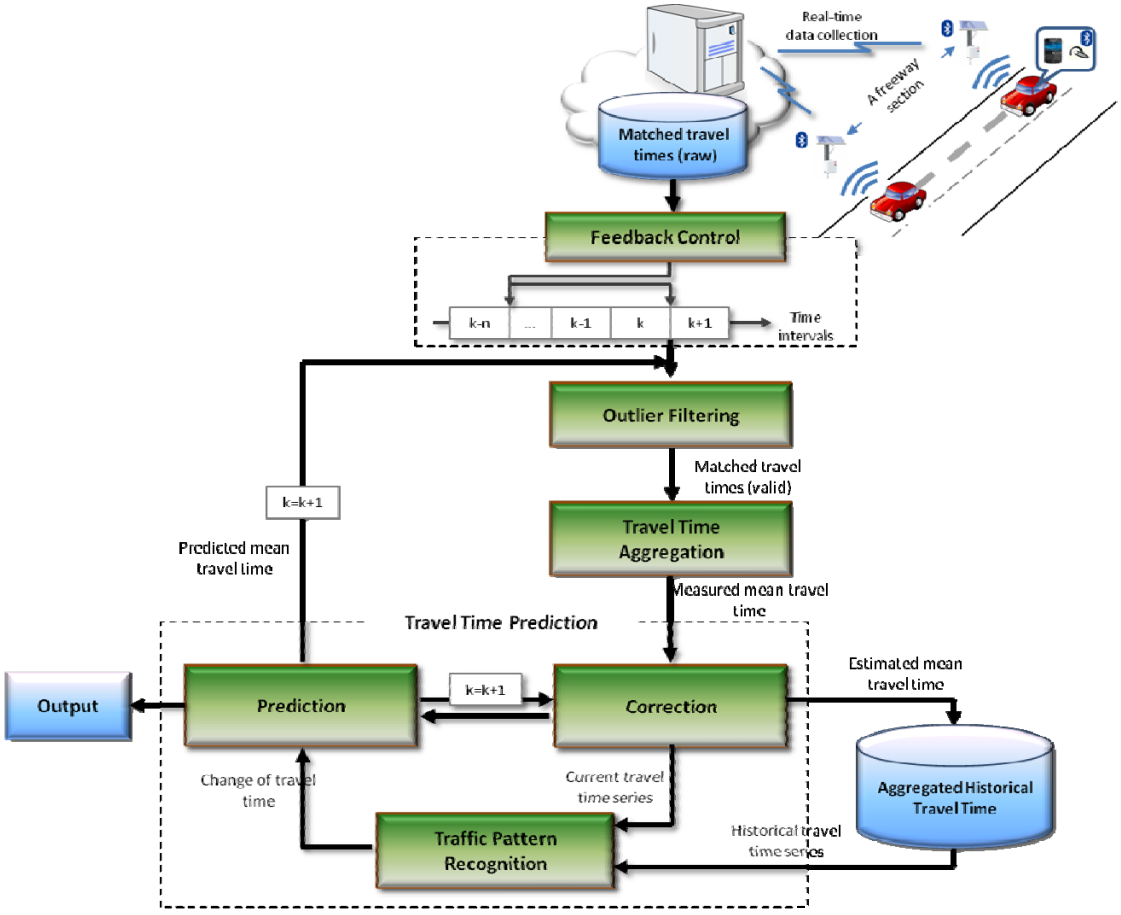


Figure 1.9: Proposed System Framework

1.5 Thesis Outline

The remainder of this dissertation is organized as follows:

Chapter 2 provides a review of the efforts that researchers have made to develop travel time estimation and travel time prediction models.

Chapter 3 examines, quantitatively, the difference between Bluetooth measured travel time and true travel time using field data collected from Bluetooth detectors, and analyzes the impacts that spacing between Bluetooth detectors has on the real-time estimation errors based on simulation data. A generalization model is developed which can be used to find the optimal average spacing between Bluetooth detectors as a function of the freeway route length.

Chapter 4 evaluates the existing dynamic travel time outlier filters and proposes an enhanced filter based on traffic flow theory. Also this chapter demonstrates the performance of the proposed filtering model by incorporating the model into two existing data driven outlier filtering algorithms and applying the enhanced algorithms to a dataset of freeway travel times collected from Bluetooth detectors.

Chapter 5 compares the performances of two methods of selecting historical data - simple aggregation (SA) and K nearest neighbor technique (KNN), and recommended to use the KNN as a method of selecting historical data in the proposed prediction model (described in Chapter 6). Then, off-line calibration of the parameters associated with KNN method is performed in this chapter using freeway travel time data collected from Bluetooth detectors.

Chapter 6 describes the proposed short-term travel time prediction model, then calibrates and validates the proposed model using freeway travel time data collected from Bluetooth detectors through statistical test and sensitivity analysis.

Chapter 7 shows application results of the proposed model to datasets collected from different freeway segments, and demonstrates the performance of the proposed model through performance comparisons between the proposed model and two benchmark models.

Chapter 8 summarizes the conclusions and contributions of this study, and provides recommendations for future research.

Chapter 2
Review of Travel Time Estimation and Prediction Methods

In this chapter, a review of methodologies for travel time estimation and prediction corresponding to the three traffic monitoring technologies mentioned in the previous chapter is provided. Although the objective of this research is travel time prediction, the accurate estimation of travel time is foundational, and the study of travel time estimation provides insight into the mechanism of travel time generation and variation over time which is crucial to developing models for travel time prediction.

2.1 Travel Time Estimation

Generally, the travel time estimation methods described in the literature differ as a function of the input data used by the method. Loop detectors detect the presence or passage of vehicles traveling along the roadway, and the measured spot speed is usually used to estimate travel time; probe vehicles provide continuous vehicle trajectories for sample vehicles, and the trajectory data are used to derive the travel time for a specific road segment; and Bluetooth detectors provide directly measured travel time of sample vehicles for the entire road segment, but the observed travel time data typically contain outliers which have to be removed before computing the average travel time for the population vehicles.

The following section describes travel time estimation methods associated with these three different input data: (1) spot speed algorithms; (2) vehicle trajectory algorithms; and (3) travel time outlier filtering algorithms.

2.1.1 Spot Speed Algorithms

Spot speed algorithms are a family of travel time estimation algorithms that rely on the speed measures obtained at a fixed location (e.g. detected by loop detectors). There are four basic algorithms used to estimate travel time using spot speed data (Zhang, 2006), namely: (a) average spot speed algorithm; (b) average link speed algorithm; (c) half distance algorithm; (d) and minimum speed algorithm. Those are described in the following with symbols referred to Figure 2.1.

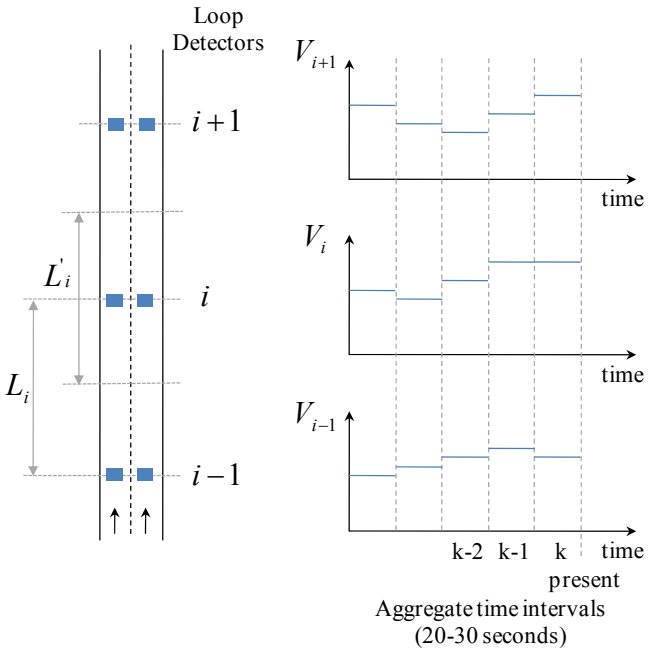


Figure 2.1: System illustration for spot speed algorithms

(a) Average spot speed algorithm

$$\hat{tt}_{i',k} = \frac{L'_i}{V_i} \tag{2.1}$$

L'_i: length of link i, which is usually defined as the distance from a middle point between the upstream detector $(i-1)$ and the current detector (i) to the middle point between the current detector (i) and the next downstream detector $(i+1)$;

$\hat{tt}_{i',k}$: estimated travel time over link i during time interval k;

$V_{i,k}$: speed at current loop detector i during time interval k.

(b) Average Link Speed Algorithm

$$\hat{tt}_{i,k} = \frac{L_i}{(V_{i-1,k} + V_{i,k})/2} \tag{2.2}$$

L_i: length of link i between loop detector $i-1$ to i;

$\hat{tt}_{i,k}$: estimated travel time over link i between loop detector $i-1$ to i;

$V_{i-1,k}$: speed at loop detector $i-1$;

V_i: speed at loop detector i.

(c) Half Distance Algorithm

$$\hat{tt}_{i,k} = \frac{1}{2}\left(\frac{L_i}{V_{i-1,k}} + \frac{L_i}{V_{i,k}}\right) \tag{2.3}$$

(d) Minimum Speed Algorithm

$$\hat{tt}_{i,k} = \frac{L_i}{V_{min,k}} \tag{2.4}$$

V_{min}: minimum speed among $V_{i-1,k}$ and $V_{i-1,k}$.

These algorithms are relative simple and easy to understand, as they make the assumption that the spot speed measured at a loop station represents the speed of vehicle travel over a fixed section of roadway. Attempts have been made to improve upon these basic algorithms.

The Vehicle Trajectories Algorithm (Coifman, 2002) assumes that loop detectors are able to measure and report the speeds and arrival times of individual vehicles and vehicle trajectories can be developed based on the assumptions about shockwave speeds. Travel time can be estimated by constructing an estimated vehicle trajectory (Figure 2.2). The performance of Coifman's travel time estimation method was evaluated using data from I-880 in California. Actual travel times were obtained from dedicated probe vehicles. It was concluded that the proposed method is relatively good as long as the road segment is not partially covered by a queue. However, Coifman's method requires that loop detectors are able to provide the speeds and arrival times of individual vehicles, something not possible with conventional loop detector systems.

The Iterative Travel Time Algorithm proposed by Cortes et. al. (2001) combines the simple Average Speed Algorithm and Vehicle Trajectory Algorithm. It is based on the assumption that the travel time experienced by vehicles on section L_i at time interval k is a linear combination of the speeds measured by loop stations i and $i-1$ (Equation 2.5).

$$\hat{tt}_{i,k} = \frac{L_i}{\alpha V_{i-1,k-\hat{tt}_{i\,k}} + (1-\alpha)V_{i,k}} \tag{2.5}$$

Where, α is a weighting factor that must be calibrated; $\hat{tt}_{i,k}$ is the estimated travel time from loop station $i-1$ to i during polling interval k; $V_{i-1,k-\hat{tt}_{i\,k}}$is the average speed reported by loop station

$i-1$ during interval $k-\widehat{tt}_{i,k}$; $V_{i,k}$ is the average speed reported by loop station i during polling interval k.

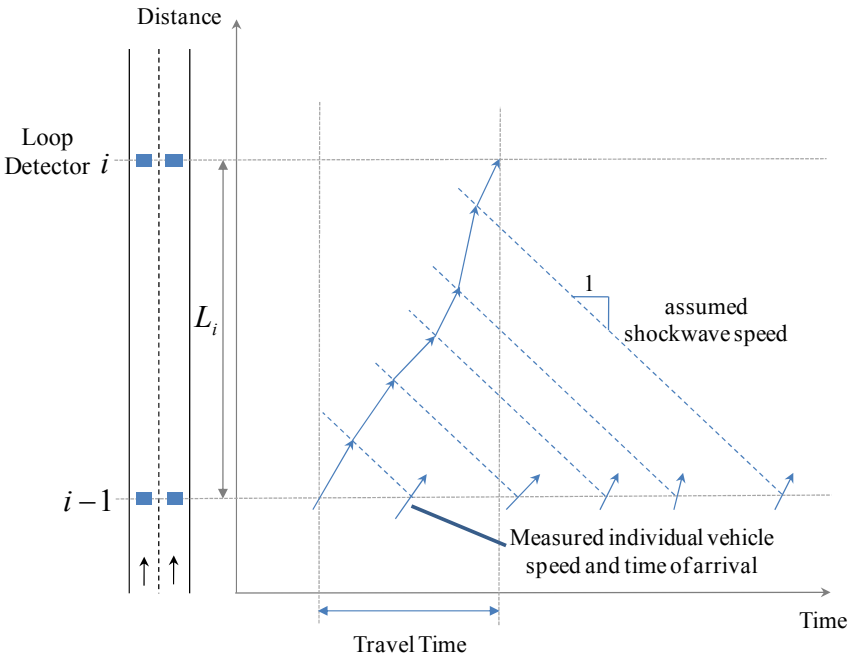


Figure 2.2: Coifman's travel time estimation method based on estimated trajectory

The unknown travel time appears on both sides of the equation meaning that this equation can only be solved iteratively. This algorithm was evaluated using only simulated data. The reported mean absolute percent error in the estimate of link travel times is in the range of 7% while they suggested that the corresponding mean absolute percentage error for the Average Speed Algorithm is in the range of 20-25%.

Besides the spot speeds algorithms, the Section Density Algorithm (Oh et. al., 2002) is another type of travel time estimation method using data from loop detectors. It estimates travel time based on the relationship among traffic volume, density and speed. The key of this algorithm is the calculation of section densities from cumulative vehicle counts and then the estimation of travel time from the section densities and interval volumes.

2.1.2 Vehicle Trajectory Algorithms

Traditional probe vehicles (e.g. vehicles dispatched to the traffic stream for the purposes of data collection or those already in the traffic network for other purposes, such as taxis, public transit buses

or fleet vehicle, but can be used as traffic data collection tools) can provide continuous vehicle trajectories which can be used to estimate travel times.

The basic definition of travel time (defined in Section 1.1.1) can be used to explain the common method using probe vehicle data to estimate travel times. The mean travel time of segment L_i (Figure 1.1) during a time period k can be calculated by averaging travel times of individual vehicles traversing the road segment (only including the vehicles that either arrive at the downstream boundary or departure from the upstream boundary within time period k). The travel time of individual vehicles traversing a road segment can be calculated by subtracting the time stamp associated with the vehicle passing the upstream boundary from the time stamp of the same vehicle passing the downstream boundary. The time stamps of individual vehicles passing the upstream boundary and downstream boundary may be provided by probe vehicle data directly or may be estimated indirectly based on the nearest data records.

Once the vehicle trajectory data are transformed to travel times of individual vehicles traversing a roadway segment, the population mean of travel time for the traffic stream during a specific time interval can be estimated using the sample mean of travel times collected from probe vehicles.

Travel times data transformed from data collected using cell phone probes require some kind of filter to remove the outliers, and the filtering algorithms are described in the following section. The data obtained from connected vehicles technology will only provide partial vehicle trajectories rather than continuous vehicle trajectories covering the whole length of the road segment because of privacy concerns, and therefore the common probe vehicle data analysis methods cannot be directly applied to connected vehicles data. The characteristic of data collected from connected vehicles and specific methods of travel time estimation using connected vehicles data can be found in the literature (Wunderlich, et al. 2007).

2.1.3 Travel Time Outlier Filtering Algorithms

Travel time of individual vehicles traversing a roadway segment can be obtained directly using AVI systems, and these travel times typically contain outliers. The outlier of sample travel times from individual vehicles can arise from a number of sources including: (1) vehicles making an enroute stop or taking a detour within a roadway segment; (2) probe devices (e.g. Bluetooth enabled devices, cell phone probes) which are not within an automobile (e.g. the device may be in a public transit vehicle,

on a pedestrian, cyclist, etc.); (3) vehicles in special purpose lanes; (4) vehicle on parallel roadways; (5) vehicles on off-ramps.

Existing filtering algorithms (Traffax 2009; Robinson and Polak 2006; Dion and Rakha 2006; Clark et al. 2002; SwRI 1998; Mouskos et al. 1998; Fowkes 1983) developed based on measurements from AVI systems using electronic toll tags or license plate recognition can be applied to data collected from Bluetooth detectors; however only a few of these algorithms are suitable for real-time detection which is the focus of this research.

TranStar system in Houston (TranStar, 2001), TransGuide system in San Antonio (SwRI, 1998), and the TRANSMIT system in the New York/New Jersey metropolitan area (Mouskos et al., 1998) collect data in real-time using in-vehicle toll tags and use three different data filtering algorithms to estimate travel time dynamically.

The TransGuide and TranStar Algorithms are generally similar, and use a rolling average algorithm that automatically filters out the travel times that exceed a user defined threshold. Equation 2.6 defines the set of valid travel times $S_{tt_{i,k}}$ that are observed between two AVI detectors i and $i-1$ during time interval k, and Equation 2.7 defines the method used to calculate the average travel time during time interval k ($\widehat{tt}_{i,k}$) based on the valid observations identified by Equation 2.6. In the TransGuide system, the length of the time interval k (i.e. rolling-average window) was set to 2 minutes and the threshold parameter (δ) was set to 0.2, so that any observed travel times that differed from the average travel time calculated in the previous interval ($k-1$) by more than 20 percent would be considered as invalid.

$$S_{tt_{i,k}} = \left\{ t_{i,j} - t_{i-1,j} \middle| t_k - t_{k-1} < t_{i,j} \le t_k \; and \; tt_{i,lower,k} \le t_{i,j} - t_{i-1,j} \le tt_{i,upper,k} \right\} \tag{2.6}$$

$$\widehat{tt}_{i,k} = \frac{\sum_{j=1}^{n_{v,k}} \left(t_{i,j} - t_{i-1,j} \right)}{n_{v,k}} \tag{2.7}$$

In the above equations, $t_{i,j}$ and $t_{i-1,j}$ are the time at which vehicle j was detected at AVI detectors i and $i-1$ respectively; t_k is the end time of time interval k; $tt_{i,lower,k} = tt_{i,k-1}(1-\delta)$ and is the lower bound of the validity window at time interval k; $tt_{i,upper,k} = tt_{i,k-1}(1+\delta)$ and is the upper bound of the validity window at time interval k; $n_{v,k}$ is the number of valid observations identified at interval k.

The TRANSMIT system uses a fixed 15-minute observation interval which contains up to a maximum of 200 observations as a sample of individual link travel time. This sample is used to estimate average travel time for this link during this time interval. There isn't any filtering before this process, but historical data from the same 15-minute interval in the previous week or weekend day would be used to smooth the estimated travel times. The smoothing process consists of an exponential smoothing algorithm (Equation 2.8), in which the factor α was set at 0.1.

$$tts_{i,k} = \alpha \cdot tts'_{i,k} + (1-\alpha) \cdot tts_{i,k-1} \tag{2.8}$$

Where, $tts'_{i,k}$ represents the historical smoothed travel time over section i of roadway during time interval k, while $tts_{i,k}$ and $tts_{i,k-1}$ represent smoothed travel times over section i of roadway during current interval k and previous interval $k-1$ respectively. The historical average travel time $tts'_{i,k}$ would be updated by current estimated travel time $tts_{i,k}$ continually.

Dion and Rakha (2006) investigated the above three existing real-time filtering algorithms. They identified the limitations of these algorithms and proposed a low-pass adaptive filtering algorithm to address the problems of reliable estimation of travel times in real-time using AVI data.

The algorithm developed by Dion and Rakha (referred to as D&R) is a low-pass adaptive filtering algorithm which identifies valid observations within a dynamically varying validity window. The set of the valid data is defined in the same way as the TransGuide algorithm (Equation 2.6), and boundaries of the validity window are determined using a user-defined number of standard deviations (n_σ) above and below the expected smoothed average travel time (Equations 2.9 and 2.10), where the expected smoothed average travel time $tts_{i,k}$ and the smoothed travel time variance $\sigma^2{}_{tts_{i,k}}$ between detectors i and $i-1$ during time interval k can be computed using Equations 2.11 and 2.12. In these calculations, the travel times are assumed to follow a lognormal distribution and therefore the validity window boundaries are not symmetrically distributed about the mean.

$$tt_{i,lower,k} = e^{\left[ln(tts_{i,k}) - n_{\sigma,k} \cdot \left(\sigma_{tts_{i,k}}\right)\right]} \tag{2.9}$$

$$tt_{i,upper,k} = e^{\left[ln(tts_{i,k}) + n_{\sigma,k} \cdot \left(\sigma_{tts_{i,k}}\right)\right]} \tag{2.10}$$

$$tts_{i,k} = \begin{cases} e^{\left[(\alpha_k) \cdot ln(tt_{i,k-1}) + (1-\alpha_k) \cdot ln(tts_{i,k-1})\right]} & if\ n_{v,k-1} > 0 \\ tts_{i,k-1} & if\ n_{v,k-1} = 0 \end{cases} \tag{2.11}$$

$$\sigma^2{}_{tts_{i,k}} = \begin{cases} \alpha_k \cdot \left(\sigma^2{}_{tt_{i,k-1}}\right) + (1-\alpha_k) \cdot \left(\sigma^2{}_{tts_{i,k-1}}\right) & if\ n_{v,k-1} > 1 \\ \sigma^2{}_{tts_{i,k-1}} & if\ n_{v,k-1} = \{0,1\} \end{cases} \quad (2.12)$$

The α_k in Equations 2.11 and 2.12 is a smoothing factor varying between 0 and 1 to determine the level of confidence that should be placed on the data $tt_{i,k-1}$ (or $\sigma^2{}_{tt_{i,k-1}}$) observed in the previous interval and the smoothed estimation $tts_{i,k-1}$ (or $\sigma^2{}_{tts_{i,k-1}}$) of the previous interval when the expected smoothed average travel time $tts_{i,k}$ (or smoothed travel variance $\sigma^2{}_{tts_{i,k}}$) of the current interval is estimated. The value of α_k depends on the number of valid observations $\left(n_{v,k-1}\right)$ identified in the previous interval and a calibrated sensitivity parameter β (Equation 2.13).

$$\alpha_k = 1 - (1-\beta)^{n_{v,k-1}} \quad (2.13)$$

The D&R algorithm described up to this point is referred to as D&R algorithm 1 (D&R1) in the remainder of this thesis. On the basis of the D&R1, two modifications were made by Dion and Rakha to increase the algorithm's responsiveness to abrupt changes in travel times and deal with the problem caused by low level of sampling rates. These modifications are: (1) "allow the algorithm to consider as valid the third of three consecutive points outside the validity window, provided that all three observations are either above or below the validity window" (Dion and Rakha 2006); and (2) "dynamically adjusts the size of the validity window based on the number of preceding sampling intervals without AVI observations" (Dion and Rakha 2006).

To implement the first modification, Equation 2.13 is substituted by Equation 2.14, and the travel time standard deviation is calculated by Equation 2.15.

$$\alpha_k = \begin{cases} 1-(1-\beta)^{n_{v,k-1}} & for\ n_a < 3\ and\ n_b < 3 \\ max(0.5, 1-(1-\beta)^{n_{v,k-1}}) & for\ n_a \geq 3\ or\ n_b \geq 3 \end{cases} \quad (2.14)$$

$$\sigma^2{}_{tt_{i,k}} = \begin{cases} 0 & for\ n_{v,k-1} = 0\ and\ n_a < 3\ and\ n_b < 3 \\ \dfrac{\left[ln\left(t_{i,j} - t_{i-1,j}\right)_k - ln\left(tts_{i,k}\right)\right]^2}{n_{v,k-1}} & for\ n_{v,k-1} = 1\ and\ n_a < 3\ and\ n_b < 3 \\ \dfrac{\sum_{j=1}^{n_{v-1,k}} \left[ln\left(t_{i,j} - t_{i-1,j}\right)_k - ln\left(tts_{i,k}\right)\right]^2}{n_{v,k-1} - 1} & for\ n_{v,k-1} \geq 2\ and\ n_a < 3\ and\ n_b < 3 \\ 0.01 \cdot \left(tt_{i,k}\right) & for\ n_a \geq 3\ or\ n_b \geq 3 \end{cases} \quad (2.15)$$

Where, n_a and n_b are counters for the number of consecutive observations above or below the validity window.

To implement the second modification, Equation 2.16 is defined to provide a model that dynamically adjusts the size of the validity window based on the number of consecutive intervals without AVI observations.

$$n_{\sigma,k} = \lambda + \lambda[1 - (1 - \beta_\sigma)^{n_{0,k}}] \quad (2.16)$$

Where, $n_{0,k}$ represents the number of consecutive intervals without AVI observations; λ represents a minimum number of standard deviations to be considered in the process of calculating lower bound and upper bound of validity window; and β_σ is a sensitivity parameter. The D&R algorithm including the modifications described by Equations 2.14 through 2.16 is referred to in the remainder of this thesis as D&R2.

Dion and Rakha applied the proposed algorithm to datasets collected by the San Antonio AVI system, and concluded that the algorithm can respond to abrupt changes in traffic conditions and function at a relatively low level of market penetration (less than 1 percent of the traffic volume). They also note that the parameters, such as β and β_σ need to be calibrated under local conditions.

2.2 Travel Time Prediction

The travel time prediction methods typically are not governed by the type of data collection technology used, and their purpose is to find the relationship between given inputs (i.e. speed, flow, travel time of previous time intervals, etc.) and output (i.e. travel time of the predicted time intervals), and then predict travel time according to the relationship. Several commonly used travel time prediction models are introduced as follows.

2.2.1 Naïve Models

The instantaneous method is one of the naïve models, which assumes that the traffic condition remains consistent during a short period and the predicted traffic condition for the next time interval is equal to the traffic condition of the previous time interval.

The historical average method is also a type of naïve model, which assumes that the traffic conditions which will be experienced in the next time interval can be estimated as the average of the historical conditions observed for the same time of day over previous days.

The naive models are all relatively simple and easy to understand, but they typically don't have a high degree of prediction accuracy (Van Hinsbergen et al., 2007).

2.2.2 Linear Regression Models

Liner regression based prediction models assume that the future traffic condition can be predicted based on a linear combination of historical and current traffic conditions. These models have the advantage that they are intuitive, have known statistical properties, and are computationally easy to apply in real-time.

Nevertheless, these models suffer from two main limitations (Zhang et al., 2003; Rice et al., 2004; Nikovski et al., 2005):

- The assumption of a linear relationship between the independent and dependent variables is frequently violated.
- The parameters that indicate how much each covariate contributes to the outcome must be calibrated off-line. The values of these parameters may be functions of a variety of factors including weather, traffic conditions elsewhere on the network, the duration of the prediction horizon, etc.

2.2.3 Nearest Neighbors Models

The Nearest neighbors (NN) model is a nonparametric regression model, which aims to find some day or days in the past that is (are) most similar to the present day in some appropriate sense. When historical day(s) have been identified that are sufficiently similar to the conditions observed so far on the current day, then the conditions observed after the current time on these historical days are used to predict the future conditions for today (Smith et al., 2002; Wild, 1997; Myung et al. 2011). The key of the NN model is finding a suitable measurement to represent the similarity (normally termed the "distance") between the traffic conditions of the present day and conditions in the past days. The effectiveness of this method is heavily influenced by the quality of the historical database. It is not possible to predict future travel times accurately if similar traffic conditions are not present in the historical database (Smith et al., 2002).

2.2.4 Time Series Models

Time series models are based on the concept that the data are not generated independently, rather their dispersion varies in time, and they are often governed by a trend and sometimes have cyclic components. As observed traffic data are usually arranged by time, many studies predicted travel times using time series models. However, the time series models require the process to be stationary

(i.e. the mean does not change over time), and usually the seasonality needs to be modeled (i.e. SARIMA). In practice, the traffic processes are not stationary, therefore, the traditional time series models are unable to capture rapid fluctuations in the traffic stream, and typically are confounded by non-recurrent congestion. Modified time series models may overcome this limitation and provide a high accuracy prediction, but the low computation speed makes them unsuitable for dynamic traffic prediction (Lee et al., 1999; Chrobok et al., 2004; Yang et al., 2005; Guin, 2006; Suarez et al., 2009; Wang et al., 2010).

2.2.5 Kalman Filter-based Models

The Kalman filter-based model is widely used in engineering applications. The Kalman filter is a recursive estimator that estimates the state of a linear dynamic system from a series of noisy measurements. As it has the ability to estimate the current or predict the future state of the system, the Kalman filter has been used in traffic estimation and prediction in many studies. The new (corrected) estimations are added into the dataset dynamically to be used for prediction, which makes the predictor reflect the traffic fluctuation quickly and that gives it significant advantage over many other methods in dynamic traffic system prediction.

The Kalman filter is a feedback control process. A priori state estimate x_k for step k is made given the knowledge of the process. Then, a posterior state estimate based on new measurements is incorporated into the priori estimate to improve the estimated value. The priori estimate can be considered as prediction, while the posterior estimate can be considered as correction. Equation 2.17 and Equation 2.18 demonstrate the main concept of the Kalman filter being used to address the problem of travel time prediction (Welch and Bishop, 2006).

A priori state prediction process:

$$x_k = \varphi_k x_{k-1} + \eta_k u_k + \omega_{k-1} \tag{2.17}$$

Where,

x_k : The state vector (i.e. travel time, section density, average speed etc.) at time interval k that is to be predicted

φ_k : Transition parameter (matrix) at time interval k which is externally determined

ω_{k-1} : Noise term that has a normal distribution with zero mean and a variance of Q_{k-1}

u_k : Optional control input from the state at the previous time interval $k-1$ to the state at the current time interval k

η_k : Transition parameter (matrix) relates the optional control input u_k to the state x_k which is externally determined

A posterior correction process:

$$z_k = \theta_k x_k + \delta_k \tag{2.18}$$

Where,

z_k : The measurement of target vector (i.e. travel time) on time interval k

θ_k : Parameter (matrix) relates the state x_k to measurement z_k

δ_k : Measurement error that has a normal distribution with zero mean and a variance of R_k

In the application, the parameter φ_k describes the relationship between the state variable x_k in two time periods k and $k+1$. Different methods are used to describe this relationship.

A well-known limitation of using Kalman filter is that the noise terms in both the state prediction process and the measurement correction process are assumed to be known. Typically, the noise terms are estimated through analysis of empirical data or simulation data and assumed constant. However, in the real-world problem, the stochastic noises always change with time. Myers and Tapley (1976) proposed a covariance matching method to adaptively estimate the unknown noises for both state and measurement process. The main concept of this method is described in the following with symbols defined in Equations 2.17 and 2.18:

Estimation of measurement noise:

The measurement noise δ_k cannot be determined because the "true" state x_k is unknown. But an intuitive approximation of the measurement noise is given by Equation 2.19:

$$\delta_j = z_j - \theta_j x_j \tag{2.19}$$

Where, δ_j is defined as the measurement noise sample based on the last l_δ observations ($j = k - l_\delta + 1, \dots, k$). If the noise samples δ_j are assumed to be representative of δ_k, and they are considered independent and identically distributed, an unbiased estimate for δ_k is taken as the sample mean, where, N is the number of noise samples, Equation 2.20:

$$\delta_k = \frac{1}{N}\sum_{j=1}^{N} \delta_j \tag{2.20}$$

The estimate for covariance of δ_k ($\hat{C}_{\delta_k}$) is given by Equation 2.21:

$$\hat{C}_{\delta_k} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\delta_j - \delta_k\right)\cdot\left(\delta_j - \delta_k\right)^T \tag{2.21}$$

Based on Equation 2.20 and 2.21, the expected value of $\hat{C}_{\delta_k}$ is:

$$E\left(\hat{C}_{\delta_k}\right) = \frac{1}{N}\sum_{j=1}^{N} \theta_j \, P_j^- {\theta_j}^T + R_k \tag{2.22}$$

Where, $P_j^- = E\left(e_j^- e_j^{-T}\right)$ is defined as the a priori estimate error covariance at time interval j; $e_j^- = x_j - \hat{x}_j^-$ is defined as the a priori estimate error; $\hat{x}_j^-$ is defined as the a priori state estimate at time interval j given knowledge of the state prior to time interval j.

Then after substitution of Equation 2.21, the unbiased estimate of R_k is given by Equation 2.23:

$$R_k = \frac{1}{N-1}\sum_{j=1}^{N}\left\{\left(\delta_j - \delta_k\right)\cdot\left(\delta_j - \delta_k\right)^T - \left(\frac{N-1}{N}\right)\theta_j \, P_j^- {\theta_j}^T\right\} \tag{2.23}$$

Estimation of state noise:

Similar to estimation of measurement noise, an intuitive approximation of the state noise is given by Equation 2.24:

$$\omega_j = x_k - \varphi_j x_{j-1} - \eta_j u_j \tag{2.24}$$

Where, ω_j is defined as the state noise sample based on the last l_ω observations ($j = k - l_\omega + 1, \ldots, k$). If the noise samples ω_j are assumed to be representative of ω_k, and they are considered independent and identically distributed, an unbiased estimate for ω_k is taken as the sample mean, where, N is the number of noise samples, Equation 2.25:

$$\omega_k = \frac{1}{N}\sum_{j=1}^{N} \omega_j \tag{2.25}$$

The estimate for covariance of ω_k ($\hat{C}_{\omega_k}$) is given by Equation 2.26:

$$\hat{C}_{\omega_k} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\omega_j - \omega_k\right)\cdot\left(\omega_j - \omega_k\right)^T \tag{2.26}$$

Based on Equation 2.24 and 2.17, the expected value of $\hat{C}_{\omega_k}$ is:

$$E\left(\hat{C}_{\omega_k}\right) = \frac{1}{N}\sum_{j=1}^{N}\left(\varphi_j P_{j-1}\varphi_j^T + P_j\right) + Q_k \tag{2.27}$$

Where, $P_j = E\left(e_j e_j^T\right)$ is defined as the posterior estimate error covariance at time interval j. $e_j = x_j - \hat{x}_j$ is defined as the posterior estimate error; $\hat{x}_j$ is defined as the posterior state estimate at time interval j given measurement z_j.

Then after substitution of Equation 2.26, the unbiased estimate of Q_k is given by Equation 2.28:

$$Q_k = \frac{1}{N-1}\sum_{j=1}^{N}\left\{\left(\omega_j - \omega_k\right)\cdot\left(\omega_j - \omega_k\right)^T - \left(\frac{N-1}{N}\right)\left(\varphi_j P_{j-1}\varphi_j^T + P_j\right)\right\} \tag{2.28}$$

This method was applied to simulation data of a typical near-earth satellite orbit determination problem by Myers and Tapley (1976). The overall results indicated that the adaptive Kalman filter performs better than the Kalman filter with constant noise terms.

The key to solving the problem addressed by the Kalman filter is to obtain the Kalman gain (K_k) for each interval k (Equation 2.29) so as to minimizing the posterior estimate error.

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k\hat{x}_k^-) \tag{2.29}$$

The basic operation of the Kalman filter is a cycle consisting of time update ("Predict") and measurement update ("Correction") processes. The operation process and equations of Kalman filter are shown in Figure 2.3. A detailed derivation of the associated equations can be found in the literature (Welch and Bishop, 2006).

Specific to travel time prediction, a Kalman filtering-based prediction model was proposed by Chien and Kuchipudi (2003) for predicting travel times with data collected from electronic toll tags. The proposed model was tested using either historical aggregated data or real-time data, and results indicated that the model based on real-time data suffers from the problem of data unavailability, while the model based on historical aggregated data cannot provide accurate results under congestion situations. Barceló et.al. (2010) proposed a Kalman filter approach for travel time prediction using

data collected from Bluetooth detectors. The model was developed using real-time data combined with historical data, and tested by applying the predicted model to a dataset collected from a 40-km-long section of motorway with one detector deployed at each end of this motorway section. The test results show high quality of the model performance (e.g. mean absolute relative error is 0.0354, correlation coefficient (R^2) between the two series, i.e. predicted and measured travel times, is 0.9863), however the test results and the method that was used to evaluate the performance suggest that the unavailability of the real-time Bluetooth data wasn't considered in their proposed prediction model.

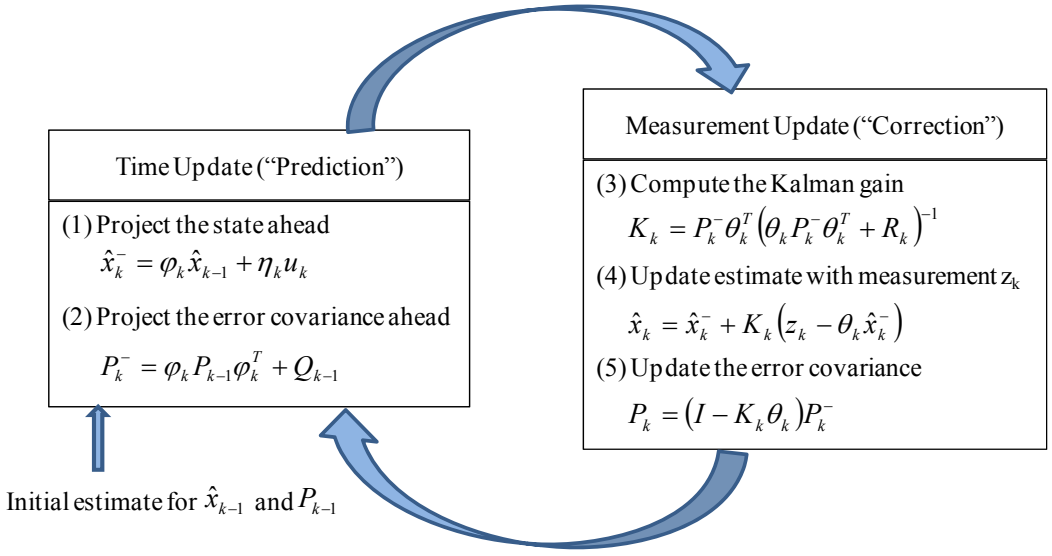


Figure 2.3: The complete Kalman filter equations and operation process (Welch and Bishop, 2006)

2.2.6 Artificial Neural Networks Models

Artificial Neural network (ANN) models are widely used to predict traffic conditions because they are able to model non-linear and dynamic processes well. However, ANN typically requires a large amount of training data and determining the optimum architecture is complicated. Many extensions on the basic concept of ANN have been tried to improve the prediction accuracy and reduce computational effort with some success. However, compared to other prediction methods, the main limitation of ANN is the difficulty associated with interpreting the calibration coefficients. It is this lack of physical interpretation of these coefficients that has resulted in the use of the term "black box" to refer to ANN models (Tan et al., 2004; Jiang et al, 2005; van Lint et al., 2005; Bucur et al., 2010).

2.2.7 Bayesian Combination Models

Bayesian combination models allow different models to be combined without the use of the prediction error in previous time interval (as is done with the Kalman Filter). This is advantageous as it overcomes the limitation associated with the real-time application of the Kalman Filter model which is the data (required for computing the prediction error) won't be available immediately at the end of the current time interval. The Bayesian approach determines the weighting factors for combining models through estimating the probability that a model is correct, and essentially, it is a way to estimate the best-fit likelihood of a model to a certain data set D, which is used to calibrate the models. The key in using the Bayesian combination model is choosing appropriately the data set D with which to calibrate the models (Fei et al, 2011, van Hinsbergen et al., 2008).

2.2.8 Traffic Theory-based Models

The prediction models described above all belong to the category of data driven models which do not consider various factors relating to traffic characteristics and travel behaviors. There have been some prediction methods based on the traffic theory, such as the traffic simulation models and the shockwave analysis.

Traffic Simulation Models

Traffic simulation models can be broadly classified into three types: macroscopic, microscopic, and mesoscopic. A macroscopic simulation model expresses the average behavior of the vehicles on a road network, and variables such as mean speed, density and flow are simulated. The macroscopic models were considered as a rough simulation method which cannot accurately represent system behavior (Chang, 1999), so few macroscopic simulation models have been applied for prediction purposes. Traditionally, macroscopic models have been used primarily for representing travel behavior within the context of regional planning level models. However, recently, Kurzhanskiy and Varaiya (2010) proposed a dynamic traffic model used to actively manage traffic based on macroscopic simulation. A microscopic simulation model simulates individual cars and the interactions between these cars. Micro-simulation models are computational intensive and typically run much slower than real time, which makes them unsuitable for real-time applications. Mesoscopic simulation models lie between macroscopic and microscopic models. Computation load is reduced from the level associated with microscopic models by reducing the detail with which vehicle behavior is modeled.

Although there are several limitations associated with using simulation models to predict traffic conditions, some researchers (van Hinsbergen et al., 2007) have concluded that apart from traffic simulation models only a few methods are able to be used for network wide predictions.

Shockwave Analysis

Shockwave analysis as part of the traffic flow theory is defined as boundaries in time-space domain that represent discontinuity in flow and /or density (May, 1990). In other words, a shockwave is the boundary between two different states of traffic flow and/or density. Figure 2.4 illustrates a roadway section with no entrance or exit. One lane is closed due to an incident. Assuming the capacity of this road section is reduced by one third, and the demand exceeds the capacity of the two remaining lanes, then a queue begins to form and grow upstream.

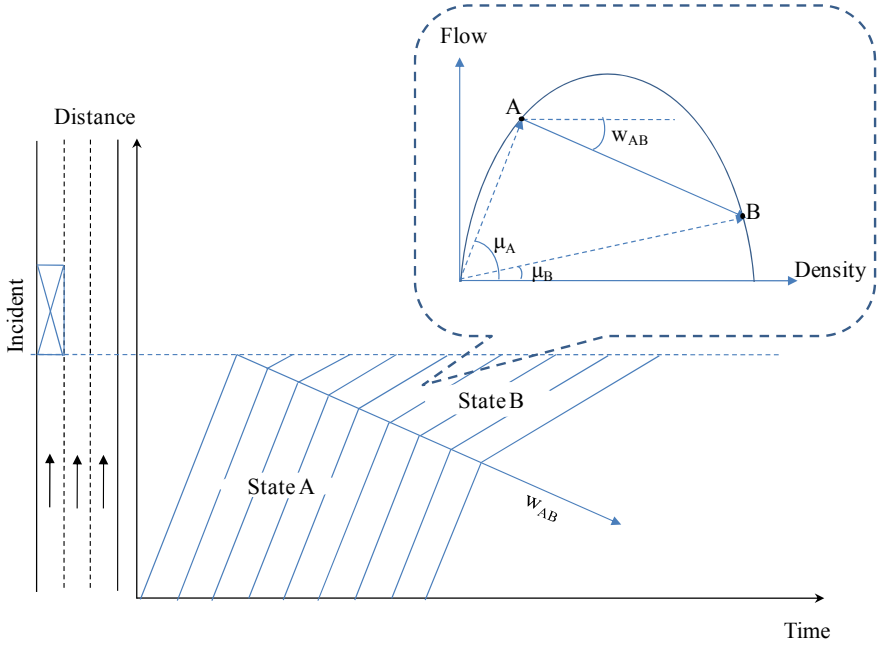


Figure 2.4: Illustration of shockwave occurrence in traffic stream

Vehicles approaching the incident zone will join in the queue, and in doing so, will travel from uncongested state A (μ_A represents speed of state A in Figure 2.4) to congested state B (μ_B represents speed of state B in Figure 2.4). The boundary between State A and State B is the shockwave, which propagates upstream with speed w_{AB}. The equation to calculate w_{AB} is given by:

$$w_{AB} = \frac{V_A D_A - V_B D_B}{D_A - D_B} = \frac{\Delta F}{\Delta D} \tag{2.30}$$

Where V_A and V_B are average speed of vehicles in state A and B, D_A and D_B are average densities in state A and B. ΔF and ΔD are change of flow rate and change of density between Sate A and State B respectively.

If the existence and speed of shockwaves can be determined by some methods, then future traffic states can be estimated and travel times can be estimated simultaneously.

2.3 Summary

To sum up, the accurate travel time estimation in real-time is a prerequisite for providing accurate and reliable short-term travel time prediction. Specific to the real-time travel time estimation based on Bluetooth data, detecting travel time outliers dynamically is challenging, especially when traffic conditions are changing quickly. Developing a reliable travel time outlier filter which is able to provide accurate travel time estimation under different traffic conditions (e.g. stable and non-stable traffic states, recurrent and non-recurrent traffic congestion) is necessary, and it is one of the objectives in this research. Another challenge of short-term travel time prediction using Bluetooth data is dealing with the unavailability of real-time data, because the Bluetooth data can only be obtained after the vehicles have finished their entire trips. Kalman filter technique is attractive in the field of short-term travel time prediction, because the real-time estimations are updated continuously whenever new measurements are available, which enables the predictor to quickly respond to traffic fluctuations. Therefore, the prediction model proposed in this research is on the basis of Kalman filter theory.

Chapter 3

Determining the Optimal Spacing of Bluetooth Detectors[1]

The travel time of an individual vehicle measured by Bluetooth detectors can only be obtained after the vehicle has traversed the monitored road segment and has been detected at the downstream Bluetooth detector. Consequently there is a time lag that exists between the time when vehicles enter the segment and the time that their travel time can be measured. This is a commonly accepted limitation of AVI data, and is one of the challenges of using this type of traffic data for real-time travel time estimation/prediction (Waller et al. 2006; Chen and Chien 2001). This time lag produces errors in estimated travel times, particularly when traffic conditions are changing. This error becomes large when the travel time between two successive Bluetooth detectors becomes large (either because the Bluetooth detectors are spaced far apart and/or because traffic is congested). Long freeway routes can be divided into shorter segments by placing additional Bluetooth detectors. However to-date, no study has quantitatively analyzed the optimal spacing between Bluetooth detectors with respect to maximizing the accuracy of real-time travel time estimations.

Some studies explored the optimal placement of traffic detectors (Ban et al. 2009; Edara et al.2008; Oh and Choi 2004), but most of them have focused on point detectors (e.g. Loop detector), rather than the point-to-point detectors (i.e. AVI detectors). A discrete optimization approach for locating AVI detector was proposed by Sherali et al. (2006), but this approach was developed to maximize the benefit (i.e. a factor that quantifies the quality of information, e.g. variability, obtained from the measured travel times) of measuring travel times on the entire transportation network given constraints (i.e. maximum number of available detectors), rather than optimize the accuracy of travel time estimation/prediction in real-time.

Specific to Bluetooth detectors, Haghani et al. (2010) suggested the detectors must be deployed on highway segments that are at least 1 mile long (1.6 km) to achieve the best performance. But this recommendation is based on ensuring measurement errors associated with uncertainty of the vehicle location at the time of detection (typically the detection zone diameter is around 100m) do not exceed a given threshold. This study did not examine the optimal detector spacing.

[1] The content of this Chapter are contained in a paper submitted for publication in the ASCE Journal of Transportation Engineering.

Edara et al. (2008) note that the placement of detectors for maximizing the accuracy of travel time estimations/predictions will vary by location based on specific conditions, and it's not necessary to place the detectors so they are evenly spaced. A study that focused on the AVI detectors spacing (Sherali et al. 2006) indicated that detectors should be placed at locations where the roadway geometry changes (e.g. an entering ramp, an exit ramp, or a change in the number of lanes) to better capture the variation of traffic conditions. We concur with the findings from Haghani et al. (2010); Edara et al. (2008) and Sherali et al. (2006); however, in this chapter we are focused on the average spacing between two consecutive Bluetooth detectors rather than on the specific locations where the detectors are to be placed.

The remainder of this chapter is organized as follows. It begins with a practical investigation of the difference between measured travel time from Bluetooth detectors and the travel time ground truth in real-time applications. Then, the impacts that the detector spacing have on the real-time estimation errors are analyzed. A proposed model is presented to generalize the analysis results, and the optimal average spacing of Bluetooth detectors as a function of route length is recommended for real-time applications. Conclusions and recommendations are provided in the last section.

3.1 Difference between Bluetooth Measurements & Travel Time Ground Truth

Travel time of an individual vehicle can only be measured by Bluetooth detectors after the vehicle has passed through the entire road segment (i.e. has arrived at the downstream detector). We call this travel time the arrival travel time (ATT). However, for real-time applications, such as posting travel time on a variable message sign (VMS), we are actually interested in disseminating the travel time that vehicles will experience at the time they enter the road segment, and this travel time is called departure travel time (DTT). Definitions of ATT and DTT can be found in Section 1.1.1.

3.1.1 Field Data Collected by Bluetooth Detectors

To examine the practical difference between ATT (i.e. measured travel time) and DTT (i.e. true travel time2) quantitatively, freeway travel times measured by Bluetooth detectors are used in this study. The field data were collected from a freeway segment (Queen Elizabeth Way from Royal Windsor Drive to Highway 427, Ontario, Canada) with a length of 18.6 km (Figure 3.1). The data set used in this study was collected during a time period from August 21, 2009 to August 26, 2009.

[2] We denote DTT as the true travel time to distinguish it from the estimated travel time (ATT). DTT are still subject to measurement errors, sampling errors and outliers.

Figure 3.1: Map of the study freeway segment

Sample travel time data for the study road segment are obtained by processing the raw data using BluSTATs software, which is provided by Traffax Inc. (2009). The outliers are also estimated by the BluSTATs software based on the following filtering algorithm (BluSTATs 2009):

For each record,

1. Calculate the 25th (ρ25) and 75th (ρ75) percentile of the 30 closest data points to the focus record.
2. Calculate the Inter-quartile Range (IQR) as the difference between the 25th and 75th percentile (IQR=ρ75−ρ25). Calculate a lower bound as LB=ρ25−2IQR. Calculate an upper bound as UB=ρ75+2IQR. These bounds approximate a three standard deviation boundary.
3. If the focus record falls below the lower bound, or above the upper bound, mark the focus record as an outlier.

3.1.2 Results of Comparison between ATT and DTT

The average ATT and average DTT during 5 minutes time interval are estimated (not including the outliers) using Equation 1.1 and 1.2. Comparison between the average ATT and the average DTT for each day (August 21, 2009 to August 26, 2009) is shown in Figure 3.2, in which we can see that the measured travel time lags behind the true travel time, and this phenomenon is more obvious when the travel time becomes large. Due to this time lag, directly using the average ATT (i.e. measured travel time) as an estimate of the average DTT (i.e. true travel time) will result in large errors, especially when traffic conditions are changing rapidly (e.g. Figure 3.2 (f)).

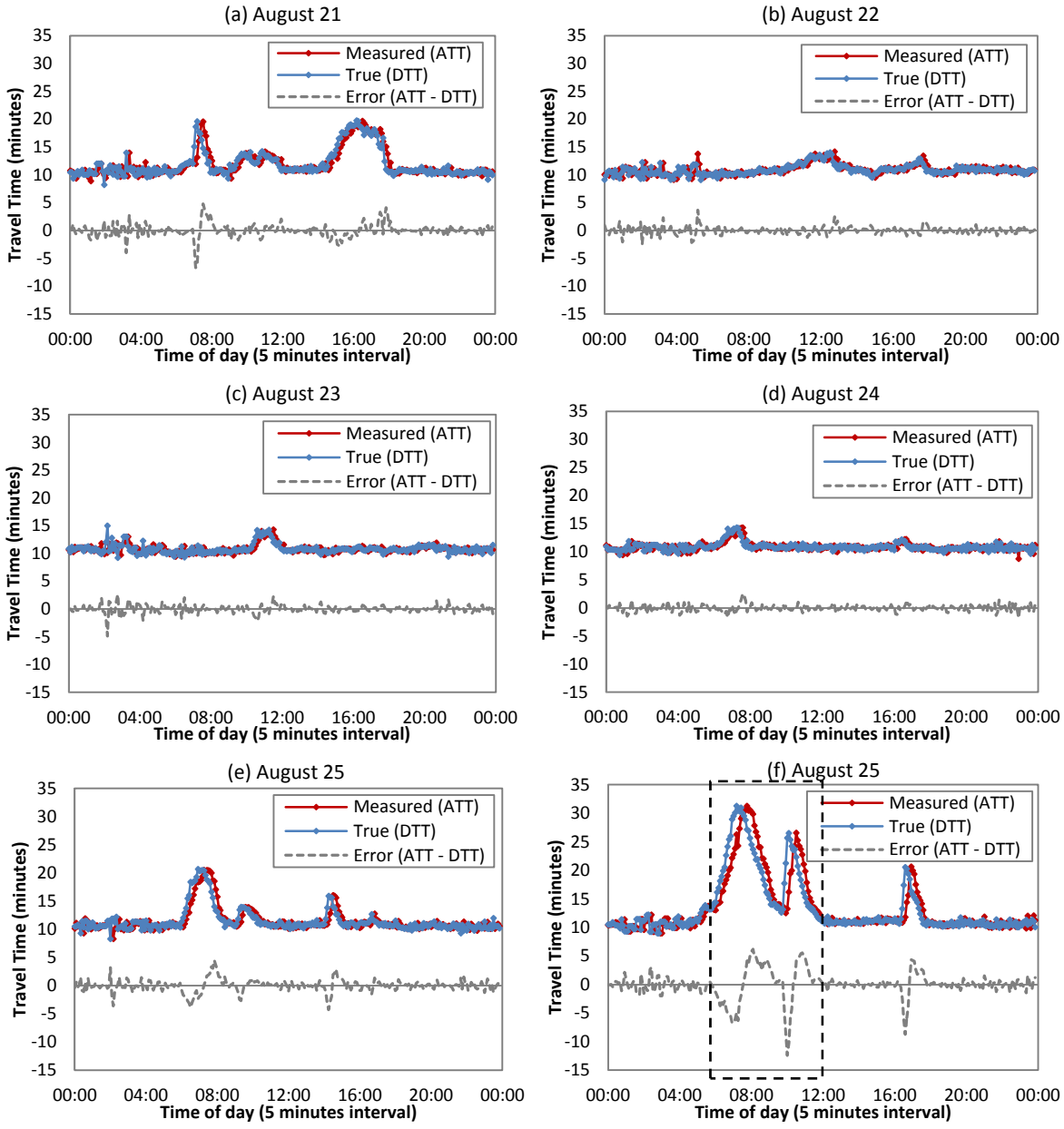


Figure 3.2: Comparisons between average ATT and average DTT

If ATT is directly used as an estimate of DTT, the absolute relative error (ARE) for time interval k can be calculated by Equation 3.1:

$$ARE_k = \frac{|ATT_k - DTT_k|}{DTT_k} \tag{3.1}$$

The mean absolute relative error (MARE) is computed as

$$MARE = \frac{1}{K}\sum_{k=1}^{K}\frac{|ATT_k - DTT_k|}{DTT_k} \quad (3.2)$$

Where, K is the total number of the time intervals within the study period.

Table 3.1 shows the results of mean absolute relative error (MARE), 95% confidence interval (95% CI) of the MARE, 90th percentile of the ARE (90th P ARE), and standard deviation of the ARE (Std. ARE) calculated based on data from the 6 days (24 hours within each day), and these measures of performance are computed in two different ways:

- Based on all the intervals (i.e. all traffic states).
- Based only on the intervals for which traffic is in congested state (where "congested" is defined as the average travel speed $<$ 80 km/h).

Table 3.1: Errors of using ATT as an estimate of DTT

	All states	Congested states
MARE	5.7%	15.2%
95% CI of MARE	(5.4%, 6.0%)	(13.7%, 16.7%)
90^{th} P ARE	13.1%	28.0%
Std. ARE	6.2%	10.1%

The results shown in Table 3.1 indicate that the measured travel time (ATT) is statistically different from the true travel time (DTT), and the errors become larger when traffic is in congestion states. A detailed illustration of the comparison on the basis of data collected from 6:00 AM to 12:00 noon on August 26 is shown in Figure 3.3.

Figure 3.3 shows that the time to identify a traffic state based on the measured travel time lags behind the time when this traffic state really happens, and this lag time is approximately equal to the measured travel time. For example, vehicles that enter the segment between 7:10 and 7:15 AM require an average 32 minutes to traverse the segment. However, the travel time of those vehicles cannot be measured until they arrive at the downstream detector, i.e. a time lag of approximate 35 minutes. The absolute relative errors are computed for the period from 6:00 AM to 12:00 AM (as shown in Figure 3.3), and the MARE on the basis of 72 time intervals within this time period is 16.57%, the 90th percentile ARE is 27.45%, and the maximum ARE is 48.39%. This result indicates that the error caused by using measured travel time as an estimate of true travel time during a time period when traffic conditions are changing substantially is very large. Moreover, due to this time lag, the pattern of travel time variation is changed (as shown in Figure 3.3), and this will further influence

the accuracy of travel time prediction, because many prediction models (especially short-term prediction) rely on the identified traffic pattern in real-time.

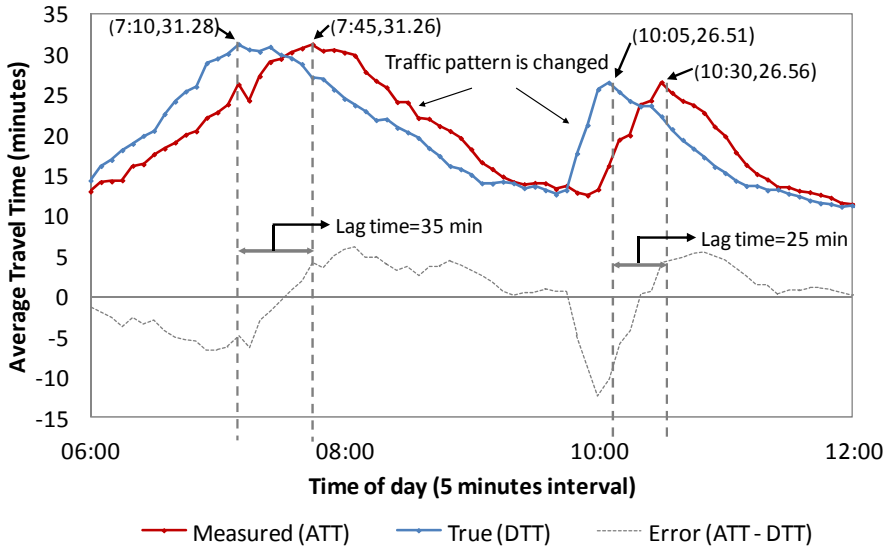


Figure 3.3: Detailed illustration of comparison between average ATT and average DTT (August 26, 2009)

It is quite obvious that the error caused by the time lag relates to the length of freeway segment, and the above analysis is based on a dataset collected from a freeway segment with a length of 18.6 km, which seems too long to satisfy the requirements of real-time travel time estimation/prediction. Previous studies have suggested dividing the long route into shorter segments to reduce the affects of the time lag that exists in AVI measurements. However, no study has quantitatively analyzed the impacts of detectors spacing on real-time travel time estimation errors, and that will be discussed in the following section.

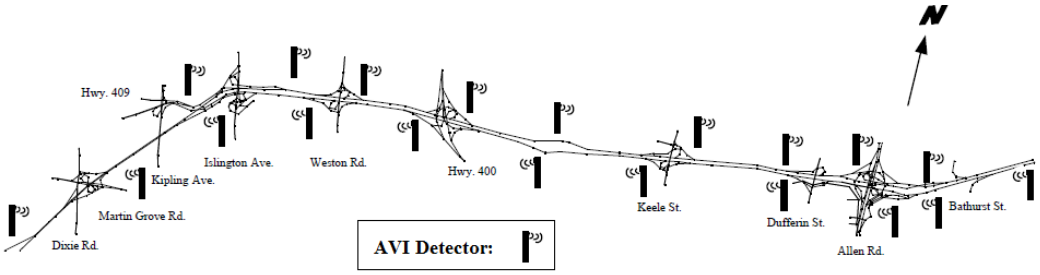
3.2 Impacts of Detectors Spacing on Real-time Travel Time Estimation Errors

The analysis requires travel time data from a freeway route with the following characters: (1) Bluetooth detectors are deployed with an average spacing approximately 1.5 km; (2) individual vehicle travel times are available between each possible pair of detectors (not just adjacent pairs); (3) the route experiences recurrent and non-recurrent congestion; (4) the sample size is sufficiently large such that sampling errors are not significant. However, field data satisfying that requirements were

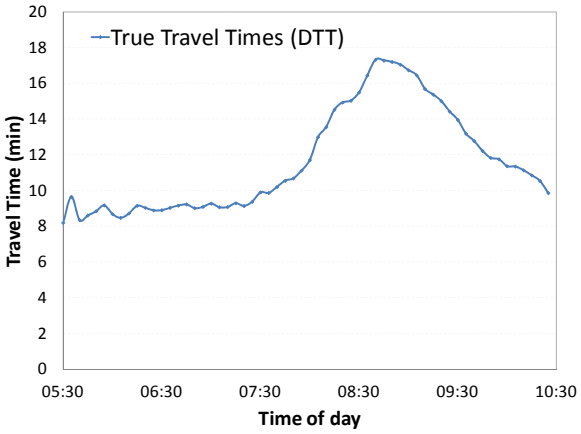
not available at the time of this study. Consequently, the data for analyzing the impacts of the spacing between the detectors on real-time estimation errors were generated using a simulation model.

3.2.1 Simulation Data

The simulation network used in this study was originally created for evaluating an automatic incident detection algorithm that made use of AVI data (Hellinga and Knapp 2000). The network is modeled after eight interchanges along a 12 km freeway route of Highway 401 in Toronto, Canada. As illustrated in Figure 3.4 (a), the eastbound and westbound freeway directions are both divided into 10 segments approximately 1.2 km in length with AVI roadside detectors at both ends of each segment.



(a) Simulation network (Hellinga and Knapp 2000)



(b) Average travel time data from the simulation model

Figure 3.4: Simulation network and simulated data

The network was simulated using the Integration traffic simulation model. The origin destination traffic demand was constructed to replicate the buildup of the AM peak from 5:30 AM to 10:30 AM. A total of 101,142 vehicle trips were simulated during this 5 hour time period. The network

experiences severe recurring congestion at several locations. The effects of non-recurring congestion were captured by simulating incidents in which the incident's location, duration, time of day of occurrence and severity (capacity reduction) were varied. For each simulation run, the travel time between each pair of AVI detectors was recorded for each vehicle. Figure 3.4 (b) shows the travel time over the 12 km freeway route as a function of trip start time averaged across all the traffic conditions simulated. The figure illustrates the average temporal variation in the route travel times.

3.2.2 Impact Analysis

To analyze the impacts of the spacing between detectors, we assume the 12 km freeway route is divided into a different number of segments. On the basis of the original network that was divided into 10 segments, the freeway route also can be divided into 5, 4, 3, and 2 segments respectively by combining different segments together, and the entire route can be treated as a single segment. There are different ways of dividing the entire route into shorter segments corresponds to 6 different average segment lengths. The simulation data include the travel time of vehicles traversing each freeway segment, and the time when vehicles enter or depart from this segment. Therefore, the average ATT and average DTT of vehicles traversing each freeway segment l during time period k can be calculated by Equations 3.3 and 3.4 as follows:

$$\overline{ATT}_{l,k} = \frac{1}{n_A}\sum_{j=1}^{n_A} tt_{l,j}\,, \qquad t_{l,j} \in k \tag{3.3}$$

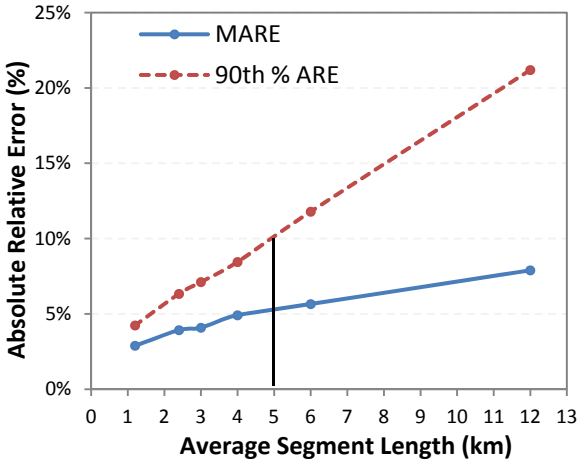
$$\overline{DTT}_{l,k} = \frac{1}{n_D}\sum_{j=1}^{n_D} tt_{l,j}\,, \qquad t_{l-1,j} \in k \tag{3.4}$$

Where, $\mathrm{tt_{l,j}}$ which is equal to $t_{l,j} - t_{l-1,j}$ is the time taken for vehicle j to traverse the road segment l, $t_{l-1,j}$ is the time at which vehicle j passed the upstream boundary of segment l, and $t_{l,j}$ is the time at which vehicle j passed the downstream boundary of the segment l; n_A is the number of vehicles passing the downstream boundary of road segment l during time period k; n_D is the number of vehicles passing the upstream boundary of road segment l during time period k.

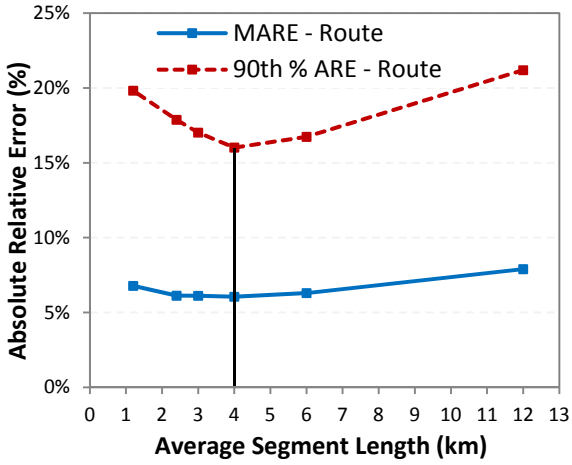
For each individual freeway segment l, the measured/estimated travel time by Bluetooth detectors is the average ATT (calculated by Equation 3.3), and the true travel time is the average DTT (calculated by Equation 3.4). The errors (MARE, 90th P ARE) between the estimated travel time and the true travel time are calculated for each of the 6 different average segment lengths, and the results

are shown in Figure 3.5 (a). These results indicate that the estimation errors caused by the time lag that exists in the AVI measurements increase approximately linearly with increases of the spacing between detectors. Although the mean errors for all the different average segment lengths considered (1.2 km to 12 km) are all lower than 10%, the 90th percentile errors vary substantially from 4.2% to 21.2% with the increase of the average segment length.

(a) **Variation of the real-time estimation errors over the average segment length (single segment)**



(b) **Variation of the real-time estimation errors over the average segment length (entire route)**



(c) **Ratio of the estimation errors over the number of segments that the entire freeway route is divided into**

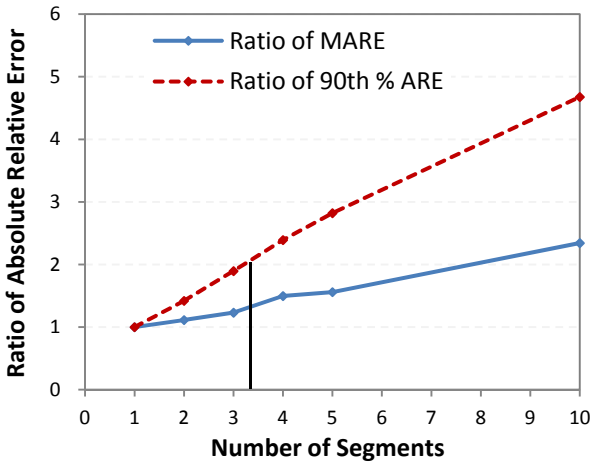


Figure 3.5: Illustration of the analysis results based on the simulation data

It is not appropriate to consider only the average prediction error (i.e. MARE), as in practice it is also necessary to consider the distribution of these errors since large errors, which may occur

infrequently, may also be unacceptable. With this consideration, the mean prediction error is not a good measure for determining the optimal average spacing between Bluetooth detectors. Consequently, a maximum limitation of the spacing between detectors (5 km) is suggested on the basis of a criterion of 90th percentile error < 10%. Combining this maximum limitation with the minimum limitation proposed in literature (Haghani et al. 2010), an optimal range of the spacing between Bluetooth detectors is expected to be 2-5 km.

This optimal detector spacing (i.e. 2-5 km) is on the premise that only a single freeway segment is analyzed. In practice, the average travel time of vehicles traversing the entire freeway route i during time period k is often estimated by adding all the average ATT of the freeway segments together as shown by Equation 3.5:

$$\overline{ATT}_{i,k} = \sum_{l=1}^{N_i} \overline{ATT}_{l,k} \tag{3.5}$$

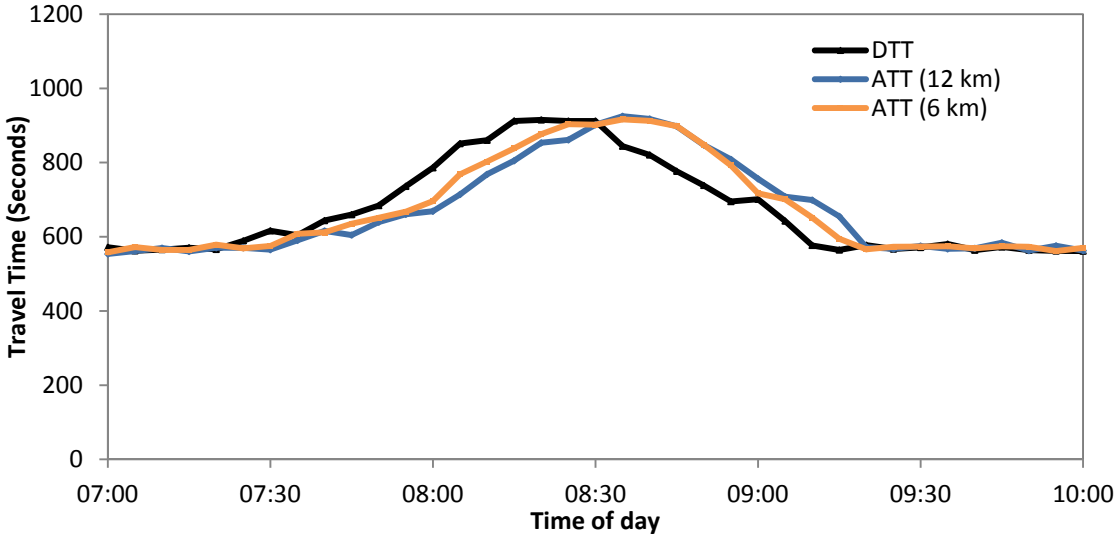
Where, N_i is the number of freeway segments that the entire freeway route i is divided into.

The "true" travel time taken for vehicles to traverse the entire freeway route is equal to the average DTT calculated by Equation 3.4 when the entire freeway route is considered as 1 segment. Therefore, the errors between the estimated travel times (i.e. travel time calculated by Equation 3.5) and the "true" travel times for the entire route are calculated as a function of average segment length and the results are shown in Figure 3.5 (b).

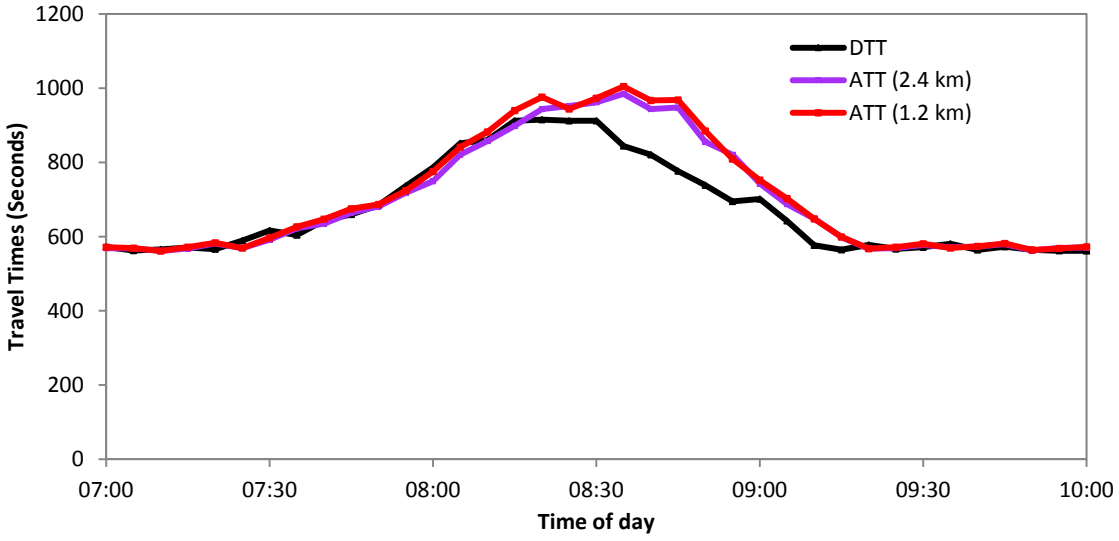
The results of Figure 3.5 (b) show that the estimation errors are no longer monotonically growing with increase of the average length of freeway segment. An inflection point can be found, and it is more obvious for 90th percentile ARE. This inflection point represents an optimal spacing between detectors (e.g. 4 km) when the objective is to provide real time travel time estimation for the entire freeway route (e.g.12 km). The inflection point arises because errors are introduced when travel times from multiple segments are aggregated.

An example of the variation of ATT over the average segment length from one scenario of the simulation data is shown in Figure 3.6. Figure 3.6 (a) shows comparisons between DTT (black line) and two cases for ATT. The blue line (ATT 12 km) represents ATT if the entire route is considered as one segment. The yellow line (ATT 6km) represents ATT if the route travel time is obtained as the sum of the ATT times from 2 segments with average segment length about 6 km. Figure 3.6 (b)

shows comparisons between DTT (black line) and two additional cases, namely ATT computed as the sum across 5 segments (ATT 2.4km) and across 10 segments (ATT 1.2km with the average segment length about 2.4 km and 1.2 km respectively.



(a)



(b)

Figure 3.6: ATT varies with average length of segment

From the results shown in Figure 3.6 (a), we can see that when the entire route (12 km) is divided into two segments (average length is 6 km), the errors between DTT and ATT are reduced. However, when placing the detectors very close together (as shown in Figure 3.6 (b), the error between DTT and ATT becomes lager when congestion is dissipating (e.g. from 8:30 – 9:30 am).

The reason that errors increase when detector spacing is very small can be illustrated using an example shown in Figures 3.7-3.9. Figures 3.7-3.9 are time space diagrams illustrating a hypothetical freeway route that is considered as one segment, two segments, and four segments respectively. $j_1, j_2 \ldots j_{10}$ are hypothetical vehicle trajectories, and vehicles experienced non-recurrent congestion because of an incident on this freeway route. For time period k, the true travel time (DTT) is the time experienced by vehicle j_8 (the only vehicle to enter the freeway route in time period k). This vehicle joined the tail of the queue when congestion was beginning to dissipate. The travel time measured in time interval k (ATT) is the time experienced by vehicle j_2. The error between the true travel time (DTT) and the measured travel time (ATT) is denoted as Error 1 and is illustrated in the Figure.

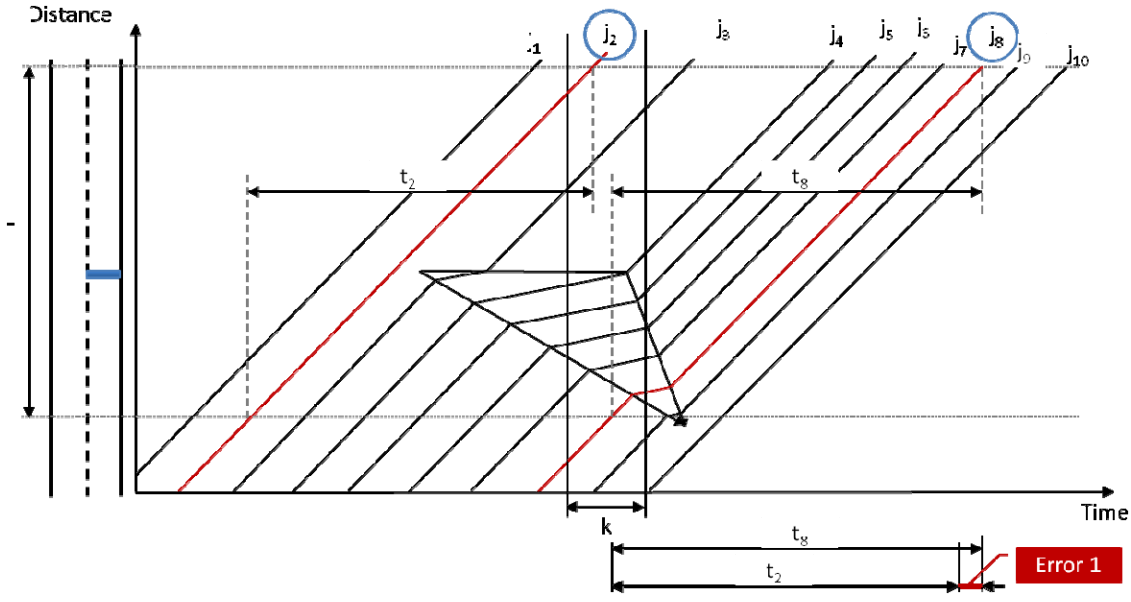


Figure 3.7: A time space diagram used to illustrate errors between DTT and ATT (one segment)

In Figures 3.8 and 3.9, the entire freeway route is divided into two and four segments respectively. The measured route travel time (ATT) is a summation of the measured travel time for each segment. In the same way, the errors (i.e. Error 2 and Error 3) between ATT and DTT are illustrated, and it can be seen that the error becomes larger as the number of segments within the route increase. This

example indicates that when congestion is dissipating, increasing the number of detectors along a route may not decrease the error between ATT and DTT.

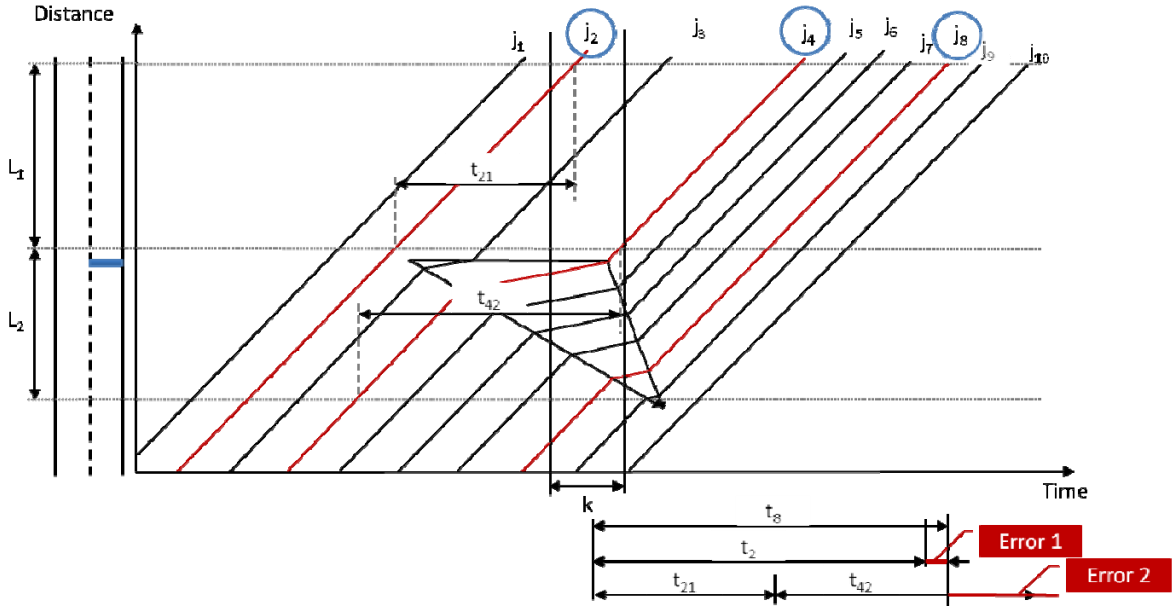


Figure 3.8: A time space diagram used to illustrate errors between DTT and ATT (two segments)

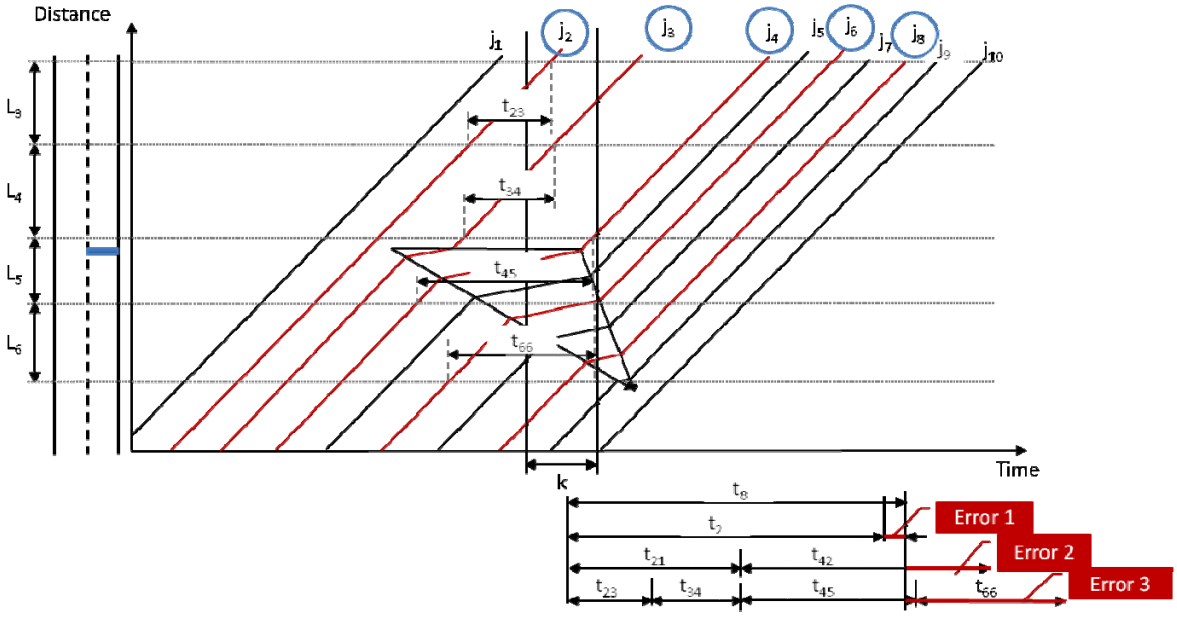


Figure 3.9: A time space diagram used to illustrate errors between DTT and ATT (four segments)

Using the "trajectory method" (Coifman 2000; Izadpanah et al. 2011) instead of the simple summation method (Equation 3.5) for aggregating the travel times from multiple segments may reduce the estimation errors, however doing so in real time will require predicting travel times for different prediction horizons, and the estimation errors still exist because of the prediction errors. Thus, the estimation errors caused in the process of travel time aggregation cannot be avoided. Of course, the optimal spacing indicated in Figure 3.5 (b) (i.e. 4 km) varies with different lengths of the entire freeway route. Consequently the results shown in Figure 3.5 (b) are not general results and therefore are not applicable if the freeway route is not 12 km in length.

In order to generalize the result and provide guidance for deployment of Bluetooth detectors for real-time applications, the ratio of the estimation errors (i.e. the error shown in Figure 3.5 (a) divided by the corresponding error shown in Figure 3.5 (b)) is computed, and the change of this ratio over the number of segments is shown in Figure 3.5 (c). The results of Figure 3.5 (c) indicate that the ratio of the estimation errors increases linearly with increase of the number of segments, and it can be seen that the 90th percentile error is already doubled when the number of segments increases to 3.

There is a practical need for guidance on the optimal average spacing between Bluetooth detectors as a function of the route length. In order to obtain such a guidance based on the results obtained previously, a generalized model is proposed and presented in the next section.

3.3 Generalization Model for Determining the Optimal Average Detectors Spacing

3.3.1 Proposed Generalization Model

The variation of the estimation errors over the average segment length (shown in Figure 3.5 (a)) and the variation of the ratio of the estimation errors over the number of segments (shown in Figure 3.5 (c)) are all approximately linear. Therefore, we can model these two relationships using linear functions as follows:

Model 1:

$$E = A \cdot L_a + B \tag{3.6}$$

Where, E is the 90th percentile estimation error of a single segment (i.e. absolute relative error, %); L_a is the average length (km) of freeway segment ($L_a = L/N$); L is the total length of the entire

freeway route (km); N is the number of freeway segments that the entire freeway route is divided into. A and B are coefficients of the linear function.

Model 2:

$$R = C \cdot (N - 1) + (D + 1) \tag{3.7}$$

Where, R is the ratio of the estimation error if travel times from single segments are aggregated; C and D are coefficients of the linear function.

Then, the estimation error when travel times from multiple segments are aggregated (E') is equal to the product of E and R, which can be substituted by parameters defined in Equation 3.6 and 3.7 as follows:

$$E' = E \cdot R = [A \cdot L/N + B][C \cdot (N - 1) + (D + 1)] \tag{3.8}$$

The coefficients A, B, C and D from the regression functions are obtained based on previously estimated 90th percentile errors. We use the 90th percentile error instead of the mean ARE because we wish to avoid large errors.

The regression results are shown in Table 3.2. For Model 1 the results show that the intercept, B = 0.024, is statistically significant. This might be considered to be counterintuitive as we might expect that as the segment length approaches zero, the error resulting from the use of ATT as an estimate for DTT should also approach zero. However, when the segment length becomes very short, measurement error (i.e. resulting from the uncertainty of the vehicle location at the time of detection) becomes more significant relative to the true travel time and therefore E does not approach zero.

For Model 2a, the regression results indicate that the intercept (i.e. $D = 0.0733$) is not statistically significant at the 95% level (p-value is greater than 0.05), and therefore the intercept was set equal to zero and the model recalibrated Model 2b).

The regression statistics shown in Table 3.2 indicate that these two models (Model 1 and Model 2b) explain over 99% of the variation in the dependant variable and all coefficients have logical values and are statistically significant. The goodness of fit plots for Model 1 and Model 2b are shown in Figure 3.6 (a) and (b).

Table 3.2: Summary of the regression results

Model 1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
B	0.0240	0.0010	25.1417	0.0000	0.0213	0.0266
A	0.0156	0.0002	97.5353	0.0000	0.0152	0.0161
Model 2a						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
D	0.0733	0.0575	1.2750	0.2713	-0.0863	0.2328
C	0.4088	0.0134	30.5930	0.0000	0.3717	0.4459
Model 2b						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
C	0.4213	0.0096	43.9277	0.0000	0.3966	0.4460

Regression Statistics			
	Model 1	**Model 2a**	**Model 2b**
R^2	0.9996	0.9957	0.9974
Standard Error	0.0014	0.0953	0.1010
Observations	6	6	6

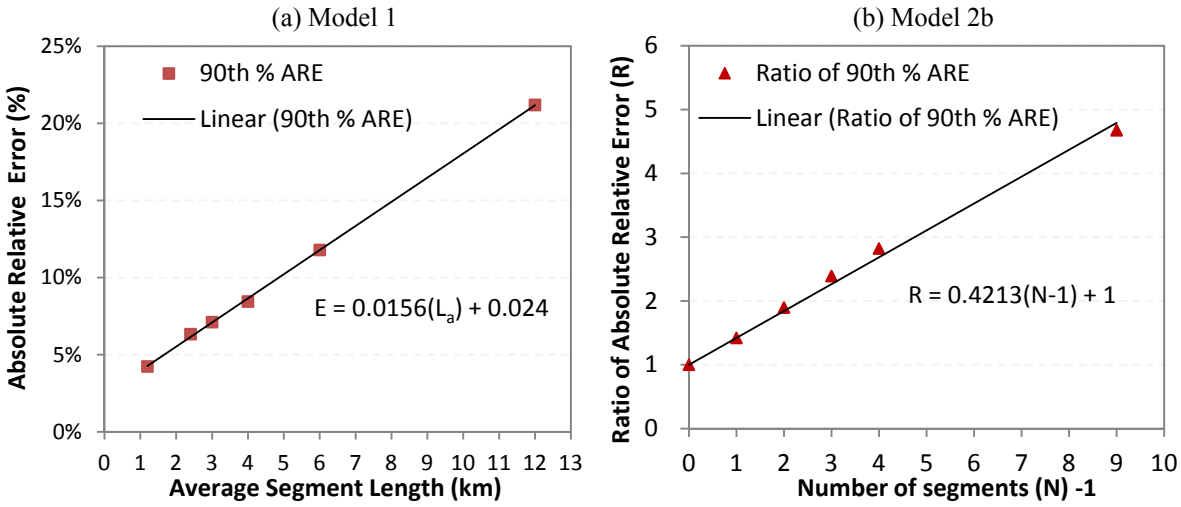


Figure 3.10: Goodness of fit for the regression models

3.3.2 Generalization Results

The values of the coefficients A, B and C are applied to Equation 3.8. The estimation error E' as a function of average segment length for different freeway route lengths is shown in Figure 3.7. The results of Figure 3.7 indicate that the optimal spacing between detectors varies with the length of the

freeway route. The optimal average spacing between detectors can be computed by taking the derivative of Equation 3.8, setting it equal to zero, and solving for N. Given that N must be an integer; we obtain the optimal number of segments as:

$$N_{opt} = \begin{cases} N_1 = integer\left(\sqrt{\dfrac{A(1-C)}{BC}} \cdot L\right); & if\ E'(N_1) \leq E'(N_2) \\ N_2 = N_1 + 1 \qquad ; & if\ E'(N_1) > E'(N_2) \end{cases} \tag{3.9}$$

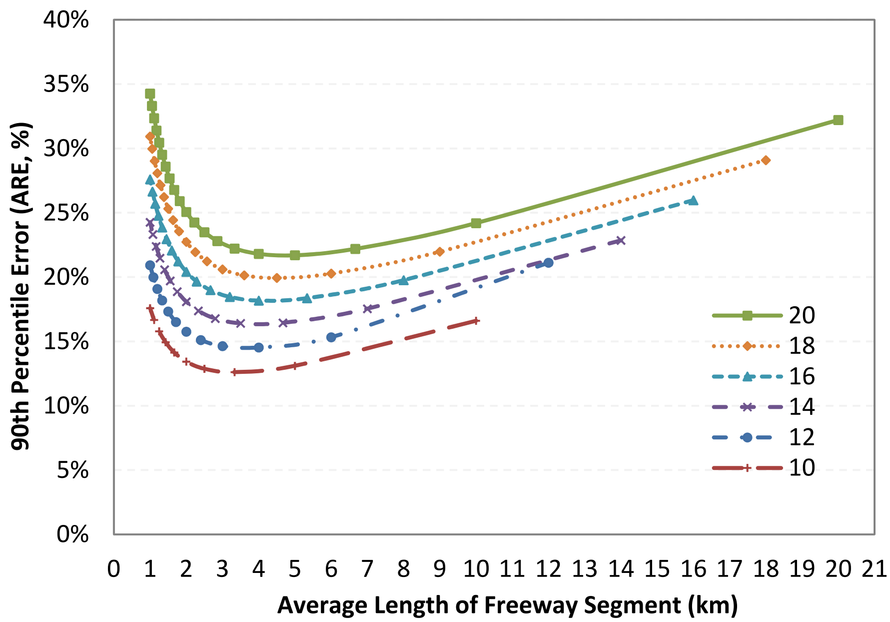


Figure 3.11: Estimation error E' as a function of average segment length for different freeway route lengths

Using the coefficients in Table 3.2, i.e. $A = 0.0156$, $B = 0.024$, $C = 0.4213$, the optimal number of segments (N_{opt}), the optimal spacing between detectors ($L_{a\_opt} = L/N_{opt}$) and the optimal number of detectors to be deployed ($N_{D\_opt} = N_{opt} + 1$) are obtained and shown in Table 3.3.

The results shown in Table 3.3 provide guidance for determining the optimal average spacing between Bluetooth detectors in real-time applications. The recommended optimal spacing is an average spacing; the detectors do not need to be evenly spaced. The roadway geometry, location of utilities, signs, obstructions, and traffic conditions should be considered for selecting specific location of individual Bluetooth detectors.

Table 3.3: Recommended average Bluetooth detectors spacing as a function of route length

Length of the freeway route L (km)	Optimal number of segments N_{opt}	Optimal average detectors spacing L_{a_opt} (km)	Optimal number of detectors N_{D_opt}
2	1	2	2
4	2	2	3
6	2	3	3
8	3	2.7	4
10	3	3.3	4
12	3	4	4
14	4	3.5	5
16	4	4	5
18	4	4.5	5
20	4	5	5

3.4 Summary

In this chapter, the travel time estimation errors caused by the time lag that exists in Bluetooth measurements are investigated. The difference between Bluetooth measured travel time (ATT) and true travel time (DTT) is quantified using field data collected from a freeway segment with a length of 18.6 km. The results show that the Bluetooth measured travel time lags behind the true travel time, and this problem is more severe when the travel time between two successive Bluetooth detectors becomes large. The real-time estimation error caused by using ATT directly as an estimate of DTT is not negligibly small (e.g. MARE = 5.7%, 90th percentile ARE = 13.1%), especially when traffic is in congestion state (e.g. MARE = 15.2%, 90th percentile ARE = 28.0%). Moreover, the temporal variation pattern of ATT is different from DTT which further degrades the accuracy of travel time prediction.

The impacts of the detector spacing on the real-time estimation errors are analyzed using simulated data for a 12 km long urban freeway that experiences substantially recurrent and non-recurrent traffic congestion. Using the idea that in practice, we wish to avoid large estimation errors, we find the optimal detector spacing by limiting the 90th percentile estimation error rather than the mean estimation error. The results suggest that a maximum spacing between detectors of 5 km in order to maintain the 90th percentile error < 10%.

The analysis results are generalized into a proposed model which can be used to find the optimal average spacing between Bluetooth detectors as a function of the freeway route length (i.e. route for

which travel time is to be posted). The proposed model is applicable for urban freeways which experience moderate to severe recurrent and non-recurrent congestion.

Chapter 4

Real-time Travel Time Estimation – Dynamic Outlier Filtering[3]

Similar to other AVI technologies (e.g. electronic toll tags, license plate recognition, etc.), the travel time data collected by Bluetooth detectors typically contain outliers as shown in Figure 4.1. The individual observations must be filtered in order to remove the outliers.

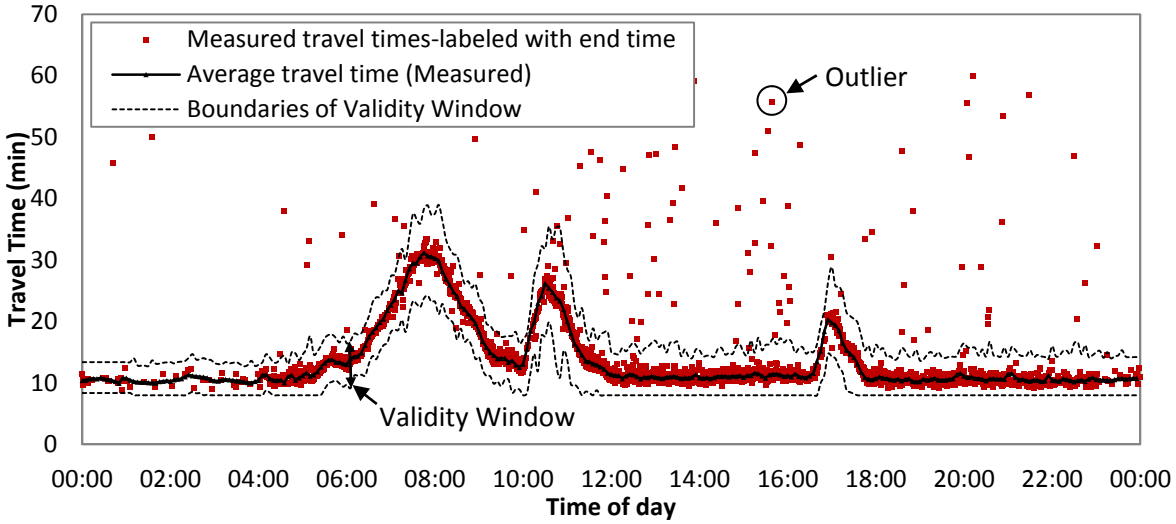


Figure 4.1: Sample of travel time data collected by Bluetooth detectors (from an 18.6 km long freeway segment on QEW, Ontario, Canada)

These outliers represent measured travel times which are not representative of the traffic stream for which travel time measurements are desired. Outlier observations can arise from a number of sources including: (1) vehicles making an enroute stop or taking a detour between two consecutive Bluetooth detectors; (2) Bluetooth devices which are not within an automobile (e.g. the device may be in a public transit vehicle, on a pedestrian, cyclist, etc.); (3) vehicles in special purpose lanes; (4) vehicles on parallel roadways; (5) vehicles on off-ramps.

Existing filtering algorithms (Traffax Inc 2009; Dion and Rakha 2006; Robinson and Polak 2006; Clark et al. 2002; SwRI 1998; Mouskos et al. 1998; Fowkes 1983) developed based on measurements from AVI systems can be applied to data collected by Bluetooth detectors; however only a few of these algorithms are suitable for real-time detection which is the focus of this thesis. Dion and Rakha

[3] The content of this Chapter are contained in a paper that has been accepted for presentation at the 2014 TRB Annual Meeting.

(2006) investigated three existing real-time filtering algorithms (i.e. TransGuide, TranStar and Transmit). They identified the limitations of these algorithms and proposed a low-pass adaptive filtering algorithm to address the problems of reliable estimation of travel times in real-time using AVI data. Dion and Rakha noted that the main challenge of detecting outliers in real-time is distinguishing outliers from sudden changes in actual travel times, especially when the AVI systems have limited sampling rates.

The existing real-time outlier detection methods are data driven algorithms which do not explicitly consider the characteristics of traffic flow. Consequently, when travel times change rapidly, for example congestion forms as a result of an incident, these algorithms frequently fail to detect the sudden changes in travel times. This chapter presents a traffic flow theory based extension that can be used to enhance the performance of existing data driven outlier detection algorithms. The proposed method can be applied to different basic filtering algorithms as an extension to solve the problem of distinguishing outliers from sudden changes in actual travel times.

The remainder of this chapter is organized as follows. It starts with a description of two existing real-time filtering algorithms - TransGuide algorithm and Dion & Rakha's (D&R) algorithm. Then, these algorithms are evaluated through application to a data set of observed freeway travel times obtained from Bluetooth detectors deployed in the Region of Waterloo, Canada. Problems with the existing algorithms are discussed and a filtering model extension based on traffic flow theory is presented. Evaluation of the proposed filtering model extension is then performed by incorporating the proposed extension into the TransGuide and D&R algorithms respectively. A summary of this chapter is provided in the last section.

4.1 Evaluation of the Existing Filtering Algorithms

As described in section 2.1.3, the TransGuide and D&R algorithms (D&R1 and D&R2) are all designed for real-time estimation and forecasting of roadway travel times using AVI data. The outliers are removed before computing the average travel time, and the estimated/predicted travel times are updated in a fixed time period. The main differences between these three algorithms relate to their ability to respond to sudden changes in travel times under low sampling conditions. In general, the D&R algorithms provide more smoothed travel time estimates due to the utilization of exponential smoothing. This was considered as an advantage of the D&R algorithm over the benchmark algorithms in the reference (Dion and Rakha, 2006), however no quantitative

evidence/basis was given for this conclusion, and therefore there was no evidence that the more smoothed the travel time estimates, the more accurate the travel time estimates.

The TransGuide and D&R2[4] algorithms were applied to a set of freeway travel times collected by Bluetooth detectors deployed along a 3.1 km section of suburban freeway (Highway 401 eastbound indicated in Figure 4.2) in the Region of Waterloo, Canada.

Figure 4.2: Map of the study area

This section of freeway (shown in Figure 4.2) consists of 3 lanes in each direction and has an AADT of approximately 145,000. Travel time data were collected over a period of approximate 12 month (February 2012 through February 2013) for both directions. On average, there are approximate 13,700 travel time observations per day (or 48 observations per 5 minute period) representing a sampling rate of approximately 9%. The variation of the sample size by time of day is shown in Figure 4.3.

The purpose of this chapter is to examine the performance of outlier detection under rapid changes in travel times. Consequently, the entire data set was screened to identify days which experienced significant recurrent and/or non-recurrent congestion. Travel time data of 10 days were selected from the dataset of eastbound direction (i.e. 401 from H8 to H24), and the travel time observations from the 10 days can be found in Appendix A. For the purpose of illustrating their performance, the TransGuide and D&R2 algorithms were applied to these data and the results are presented in this Chapter for two representative days (Oct.5th and Nov.5th, 2012).

[4] The D&R1 algorithm was not applied because it is an earlier and less robust version of the D&R2 algorithm.

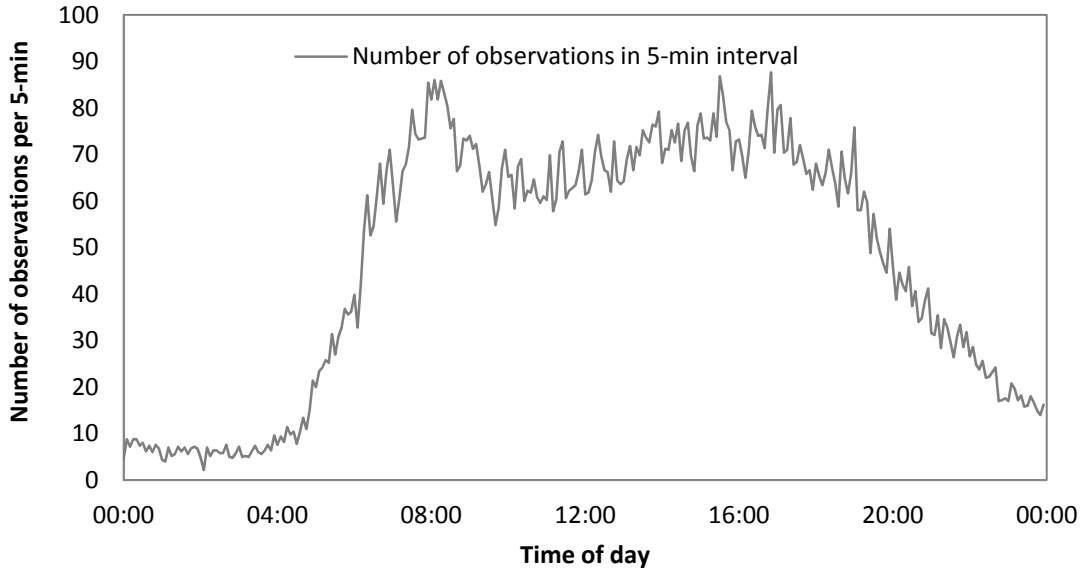


Figure 4.3: Variation of average number of measured travel times in 5-min interval

4.1.1 TransGuide Algorithm

Figures 4.4-4.7 illustrate the applications of the TransGuide algorithm to two different days (24 hour periods), Oct. 5th and Nov. 5th in 2012 based on different parameters (i.e. k = 2 minutes or 5 minutes and travel time threshold δ = 20% or 50%).

It can be seen clearly from the results in Figures 4.4-4.7, that the algorithm performs poorly when a travel time threshold (δ) = 20% is used. When the threshold parameter is increased to 50%, the results are improved for the data collected on Oct.5th (Figure 4.4 (b) and Figure 4.5 (b)), but a large amount of the valid data are incorrectly labelled as outliers and excluded from the validity window at the beginning of the traffic congestion (14:00 -14:30) if the rolling-average interval is 5-minutes (Figure 4.5 (b)). This occurs because the change of travel time between consecutive intervals will increase with the length of time interval increase, and when this change is larger than 50% of the travel time estimated in previous interval, it cannot be detected by TransGuide algorithm with δ = 50%. This suggests that more frequently updating the estimation may be able to avoid this kind of problem however it only applies to data with a higher sampling rate.

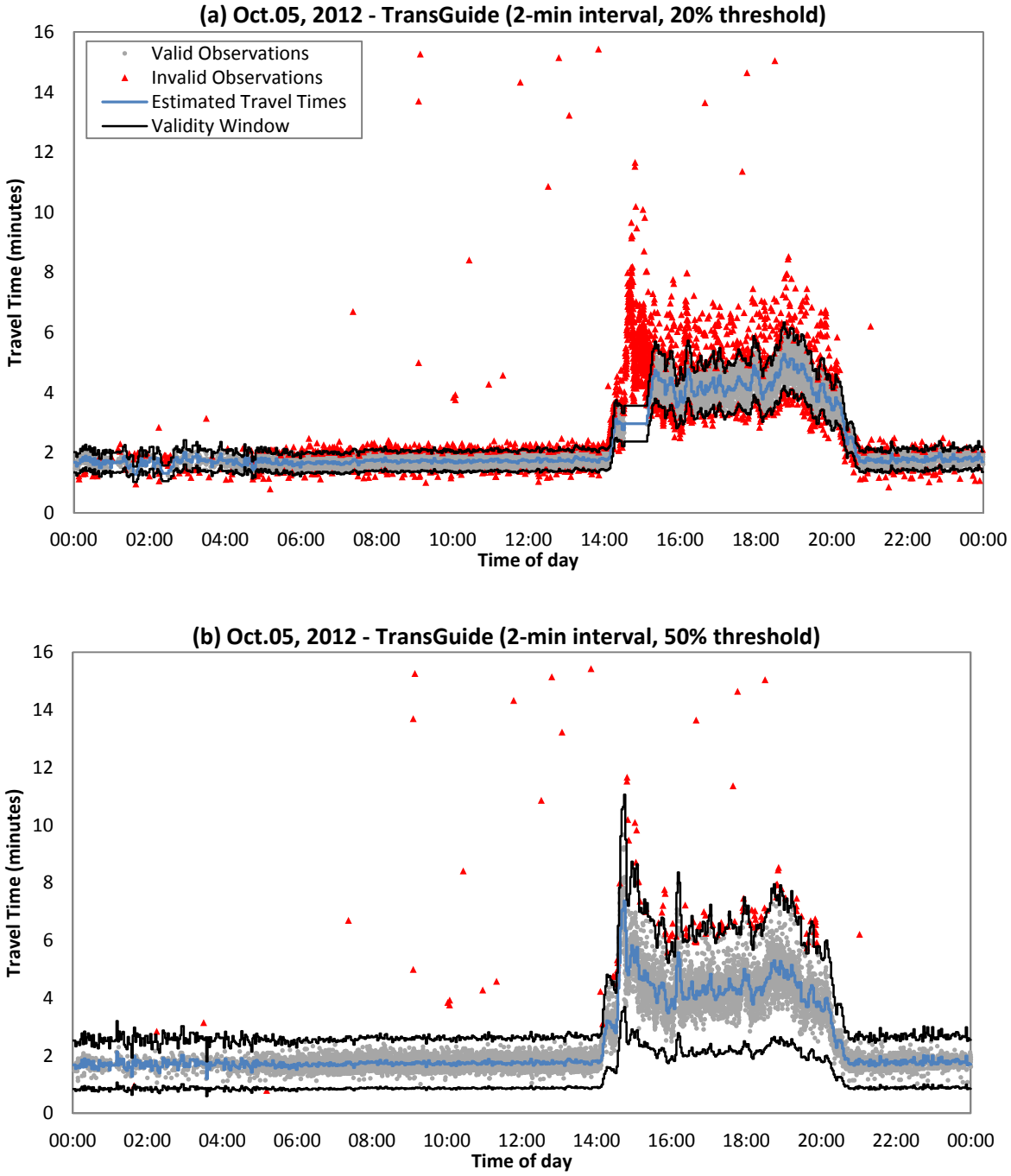


Figure 4.4: Applications of TransGuide algorithm on data collected by Bluetooth detectors (2-min interval, Oct. 05 2012)

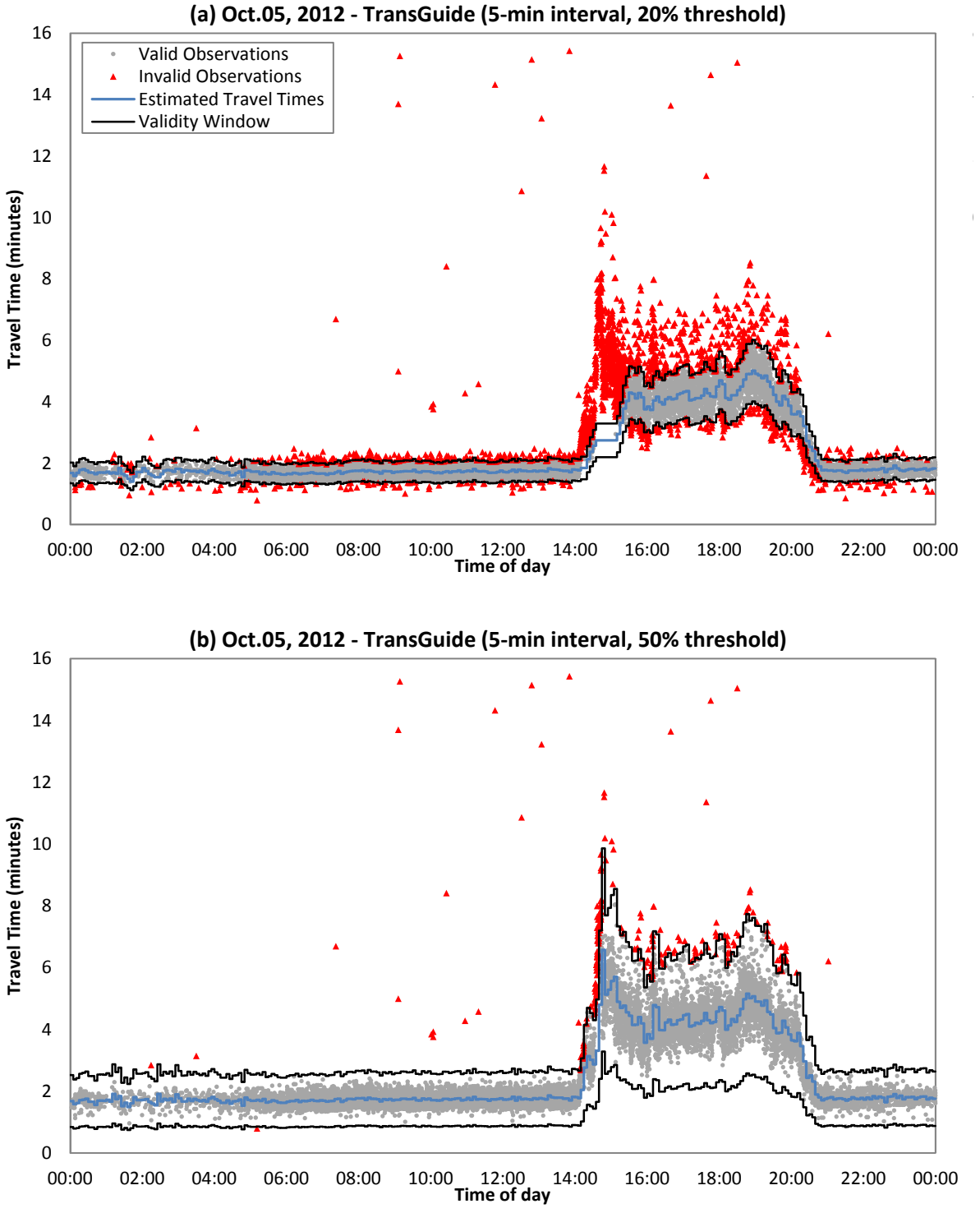


Figure 4.5: Applications of TransGuide algorithm on data collected by Bluetooth detectors (5-min interval, Oct. 05 2012)

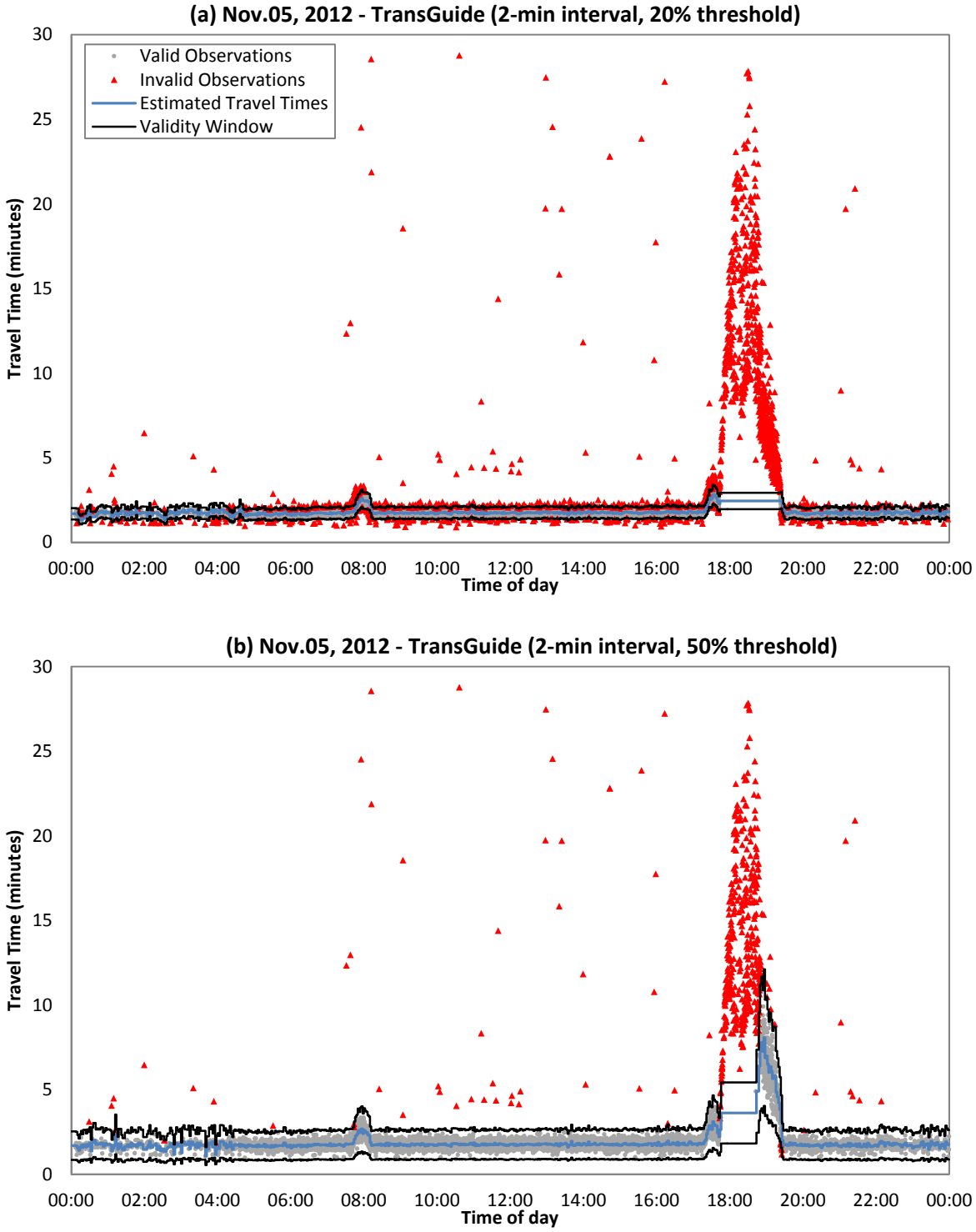


Figure 4.6: Applications of TransGuide algorithm on data collected by Bluetooth detectors (2-min interval, Nov. 05 2012)

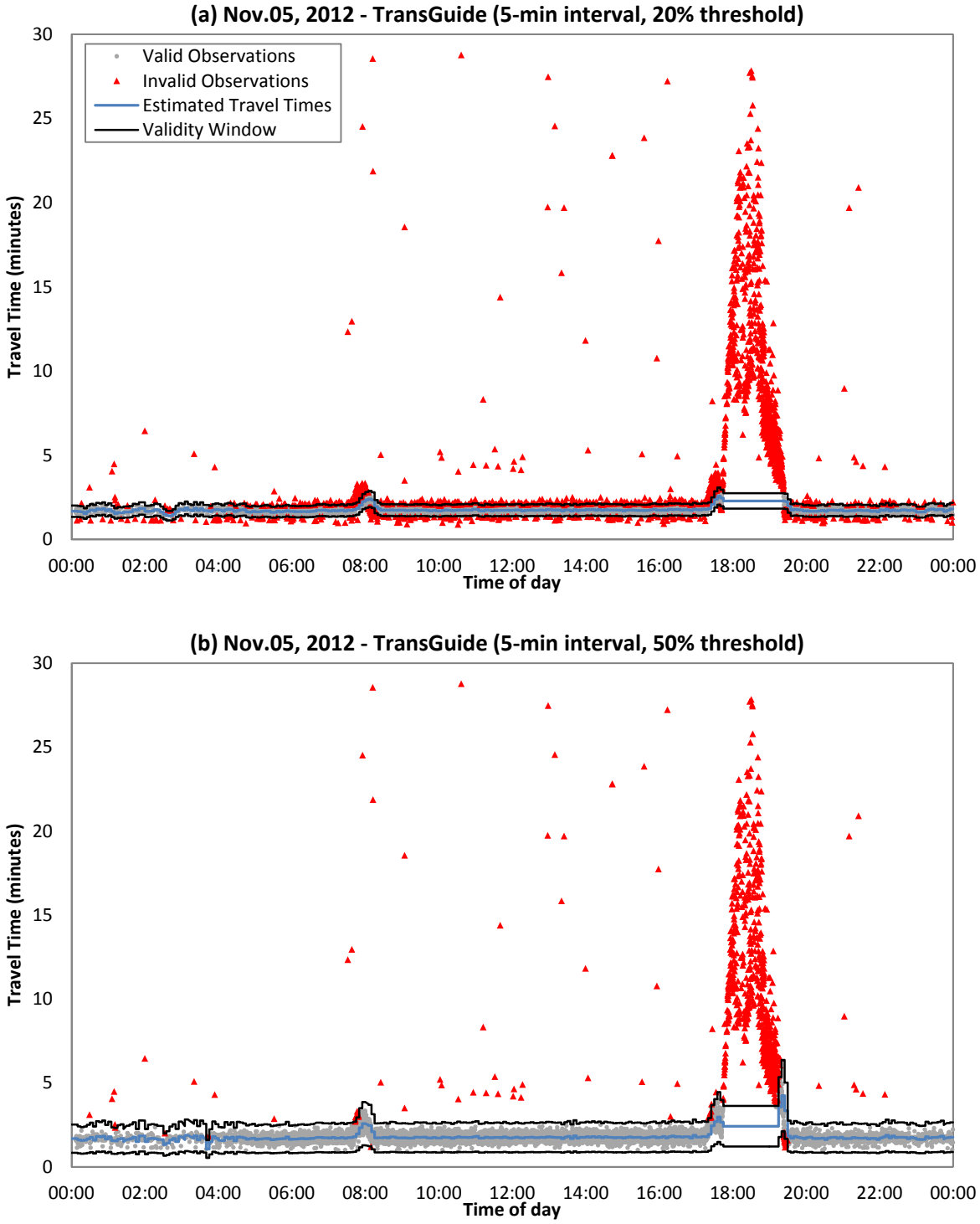


Figure 4.7: Applications of TransGuide algorithm on data collected by Bluetooth detectors (5-min interval, Nov. 05 2012)

Similar performance issues occur in the application to the data collected at Nov.5th (Figure 4.6 and 4.7). On this occasion the TransGuide algorithm isn't able to track the sudden changes in travel times for both the 2-min and 5-min interval durations even when the threshold parameter is set to 50%. This is because the change of actual travel time between two consecutive intervals exceeds the 50% threshold during the period around 17:50, and further increasing the threshold parameter is not a practical solution because it would also result an increased number of outliers incorrectly labelled as valid data. These results illustrate that the TransGuide algorithm is not able to perform reliably when travel times change rapidly.

4.1.2 D&R Algorithm

The D&R2 algorithm was applied to the same set of data and the results (using parameter values of $\lambda = 3, \beta = 0.2, \beta_\sigma = 0.2$) are shown in Figures 4.8-4.9. From these results we can observe that the D&R2 algorithm performs much better than the TransGuide algorithm in the aspect of tracking sudden changes of travel times, and the overall performance of D&R2 algorithm is better than TransGuide. However, further investigation shows that the underlying problem has not been entirely solved.

Consider Figure 4.10-4.12 which illustrates the application of the D&R2 algorithm to the Oct. 5th data using 2-min and 5-min interval, and the Nov. 5th data using 5-min interval. The case of Nov.5th data using 2-min interval is not illustrated here because no serious issues arise for the D&R2 algorithm for those data. From Figures 4.10 (a), 4.11 (a) and 4.12 (a), we can see that the validity window becomes extremely large at some intervals, and this is mainly because three consecutive data points are observed above the validity window, and then the sample variance is set to $0.01(tt_{i,k-1})$ according to Equation 2.15. With this large sample variance, the validity window is expanded so as to track the sudden changes of traffic conditions. However, when the validity window is expanded to be unrealistically large (as in the illustrated cases), the outlier detection algorithm treats almost all observations as valid, and frequently incorrectly labels outliers as valid data. An example of invalid observations being included in the validity window, and therefore labeled as valid data, can be observed in Figure 4.10 (b).

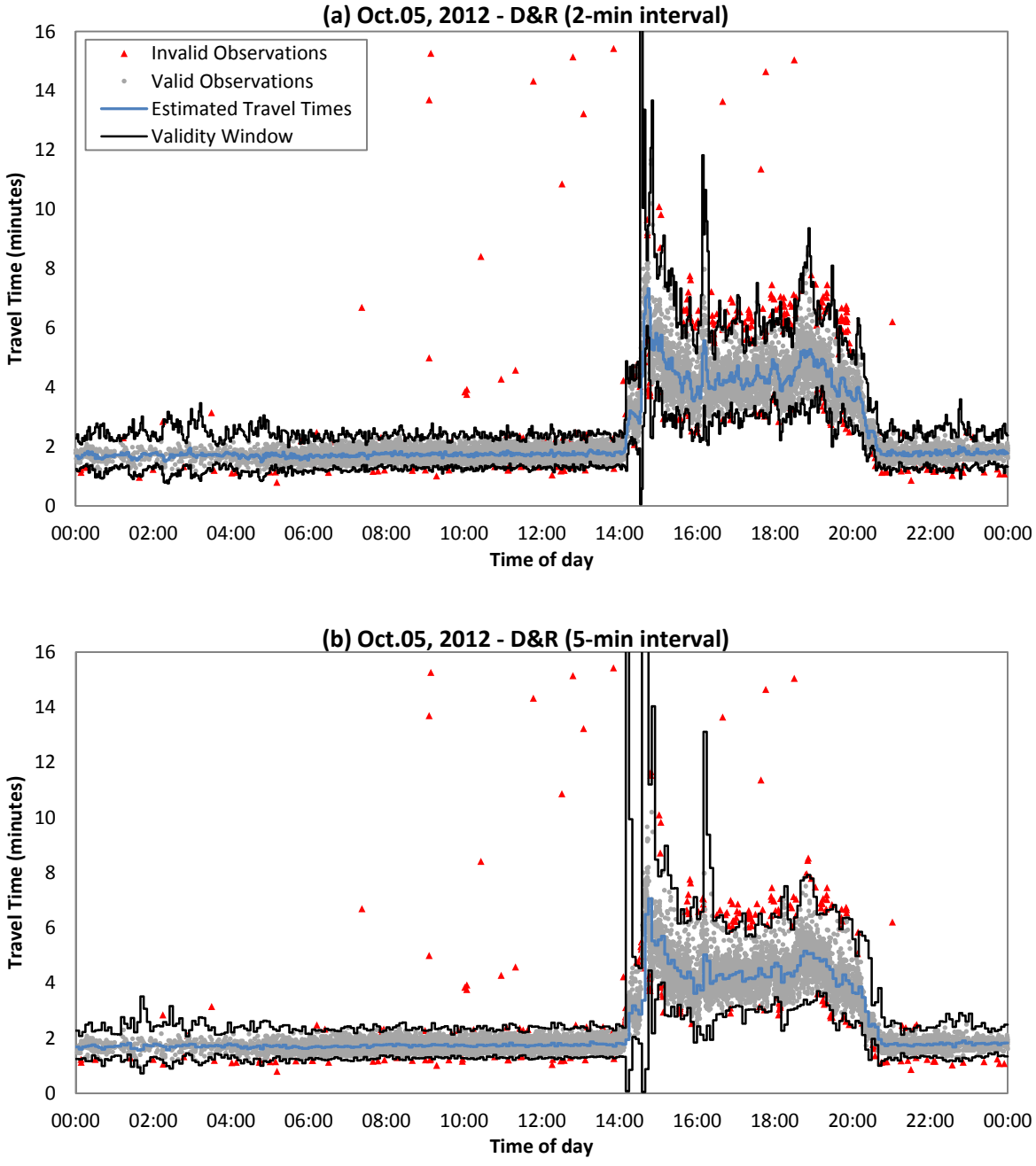


Figure 4.8: Applications of D&R2 algorithm on data collected by Bluetooth detectors (Oct. 05 2012)

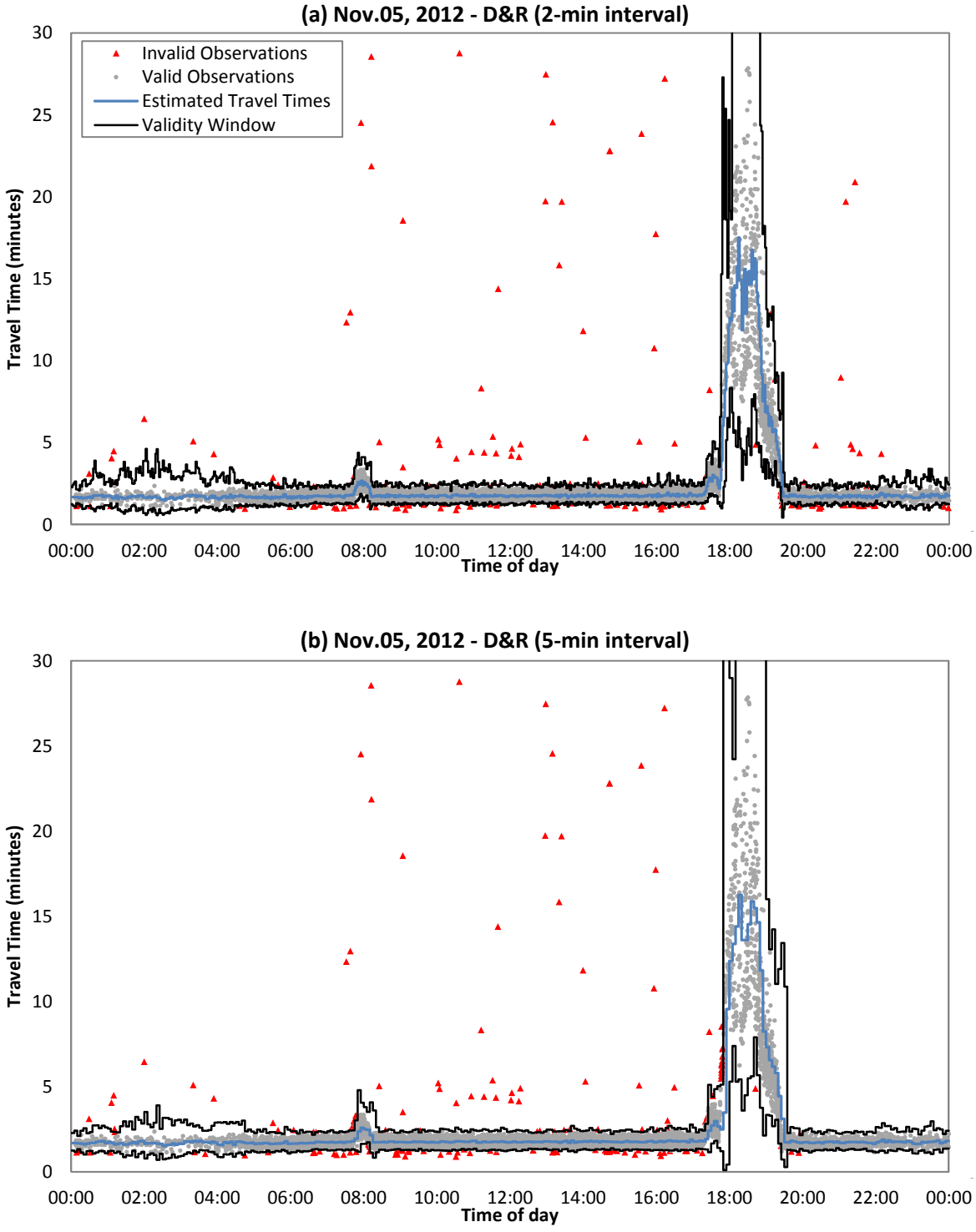


Figure 4.9: Applications of D&R2 algorithm on data collected by Bluetooth detectors (Nov. 05 2012)

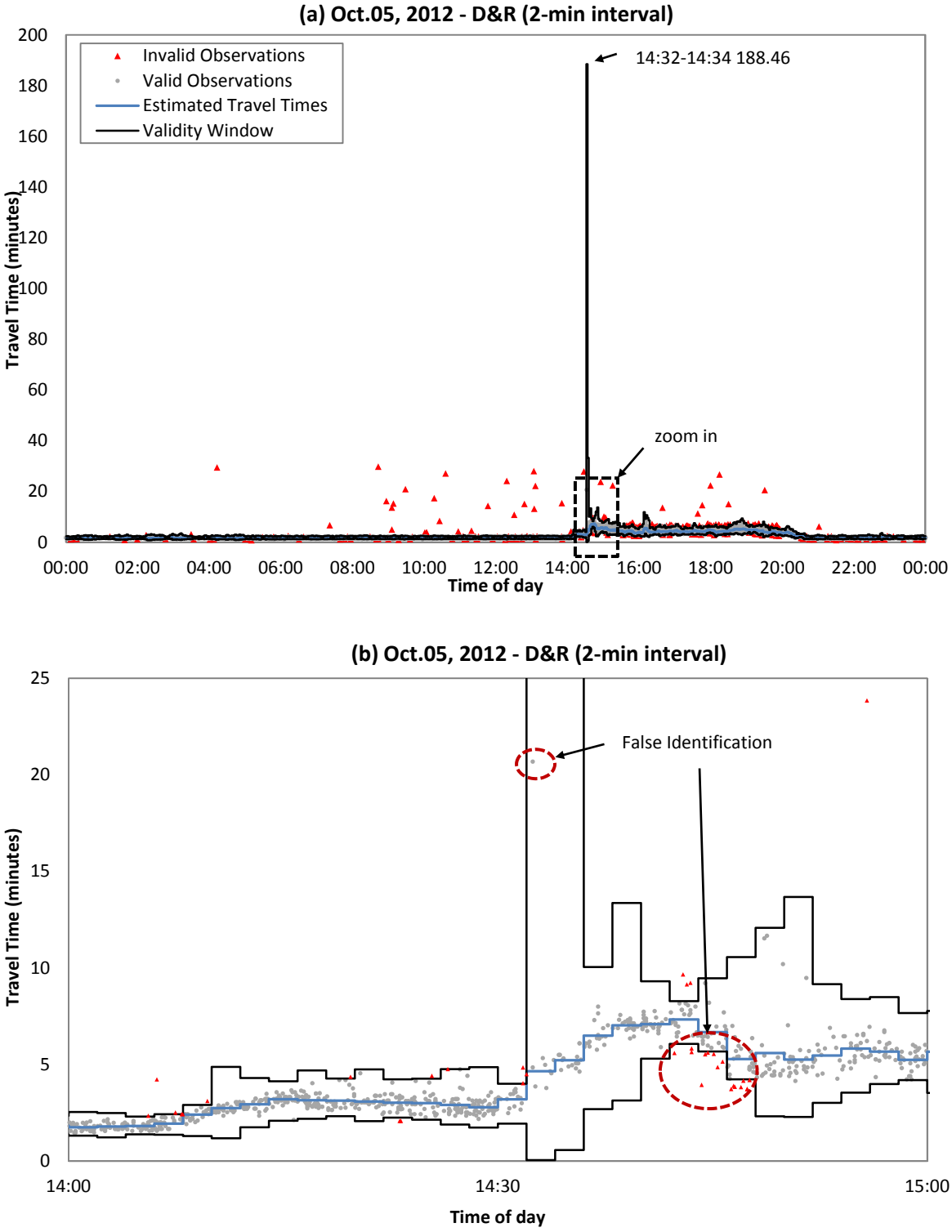


Figure 4.10: Illustrations of D&R2 algorithm details (2-min interval, Oct. 05 2012)

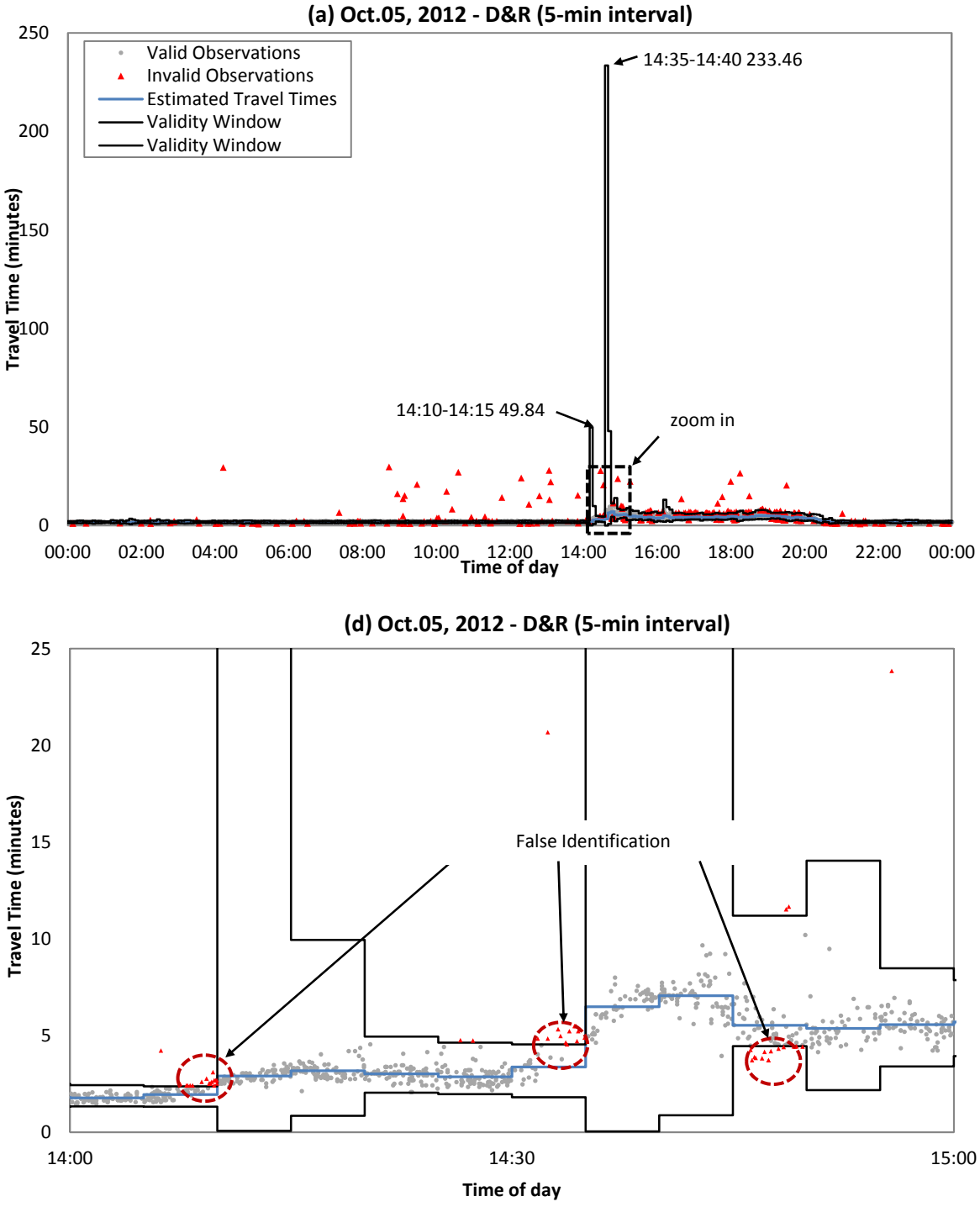


Figure 4.11: Illustrations of D&R2 algorithm details (5-min interval, Oct. 05 2012)

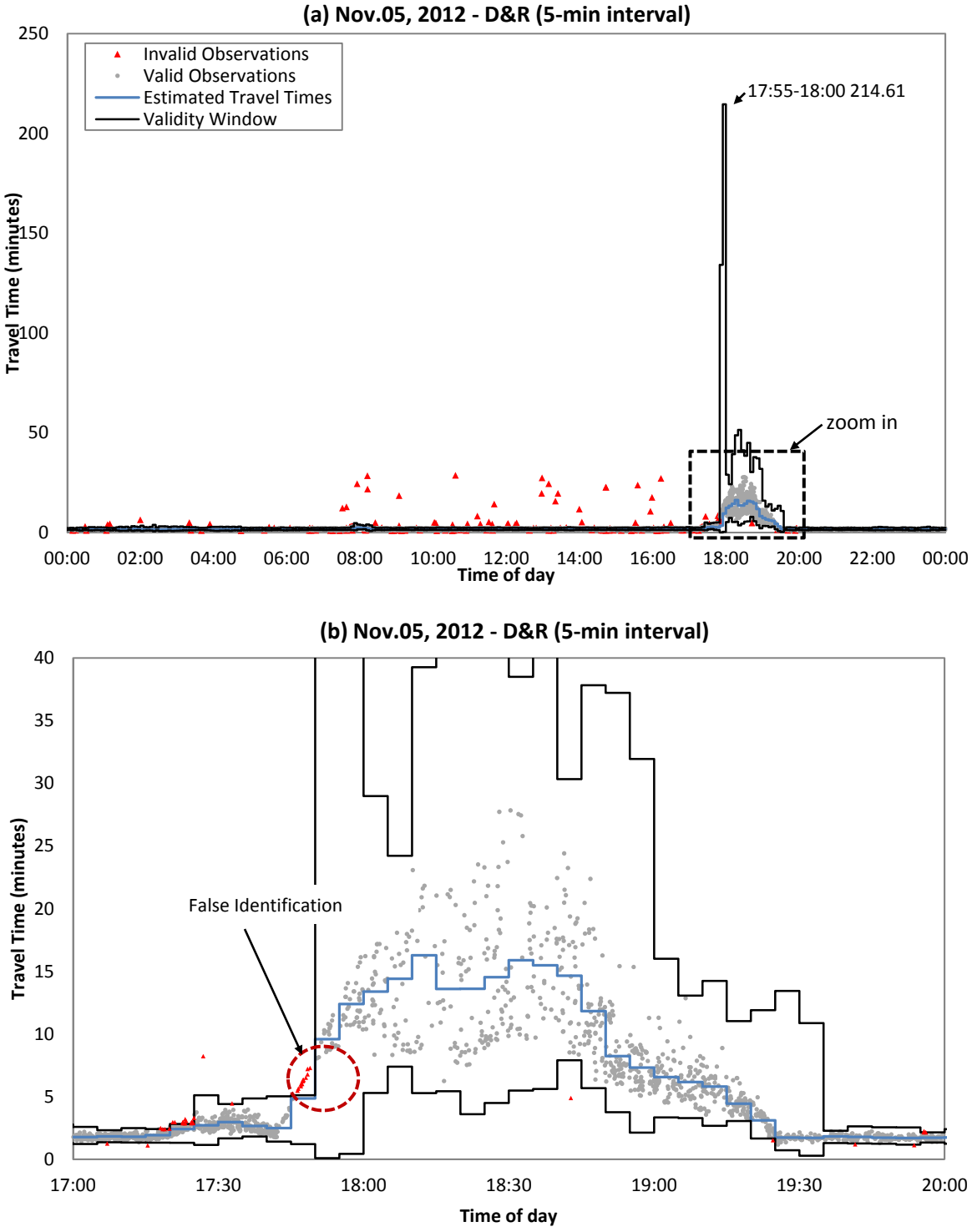


Figure 4.12: Illustrations of D&R2 algorithm details (5-min interval, Nov. 05 2012)

Another problem, shown in Figures 4.10 (b), 4.11 (b) and 4.12 (b), is that a large amount of valid observations lie outside of the validity window during some specific time intervals and are therefore incorrectly labeled as outliers. The reason is that D&R2 algorithm assumes the expected average travel times and standard deviations of observed travel times from consecutive intervals remain constant. However, in reality the difference of these two values between consecutive intervals is substantial when traffic conditions are changing. Although this problem is addressed in D&R2 by the "three consecutive points" criteria, the adjustments to the validity window take effect in the next interval, meaning that the observations incorrectly labeled as outliers in the current interval remain incorrectly labeled.

The results suggest that though the D&R2 algorithm provides better performance than the TransGuide algorithm, the algorithm still suffers performance limitations when travel times change rapidly. It is likely that performance could be improved by calibrating the various parameters; however doing so is not trivial, in large part because there is no objective way to quantitatively evaluate the performance of the filtering model using field data as the truth is not known. Consequently, this limits the robustness and transferability of the model and undermines confidence in the use of the model for real-time applications, particularly for those periods when travel times are changing rapidly – the periods of most importance for traffic control and traveler information.

Accordingly, a simple, robust and transferable method that can be used to solve the problem of existing filtering algorithms in terms of tracking sudden changes in travel times is required. We propose to use a model based on traffic flow theory to better explain different traffic situations.

4.2 Proposed Traffic Flow Filtering Model

The proposed method can be applied as an extension to an existing data driven filtering model (e.g. TransGuide algorithm and D&R1) to enhance the model's ability to track sudden changes in travel times. This extension is developed on the basis of traffic flow theory rather than purely responding to observed data characteristics. The proposed model uses the concept of shockwaves to impose a boundary of the validity window as illustrated in Figure 4.13.

In a free flow traffic state (i.e. state A shown in Figure 4.13 (a)), if the free flow speed is known, then the minimum travel time over this road section can be estimated ($tt_{i,min} = L_i/V_f$). When congestion occurs, as is shown in Figure 4.13 (b), a congested traffic state (i.e. state B) is formed in which vehicles travel with speed V_B. The travel time over section i would be computed as $tt_i =$

$L_{iA}/V_A + L_{iB}/V_B$ in which L_{iA} and L_{iB} are the length of the highway section traversed by the vehicle when operating in traffic state A and B, respectively. Of course, in practice, L_{iA} and L_{iB} are unknown. Traffic state B will propagate upstream with speed w (i.e. shockwave speed) until the entire road section is congested (shown in Figure 4.13 (c)). In this case, vehicles would travel at a speed that equals V_B over the entire section. If the speed of the most congested state (V_{min}) is able to be calibrated without considering the special case that the entire road section is closed, then the maximum travel time over this road section can be estimated as $tt_{i,max} = L_i/V_{min}$.

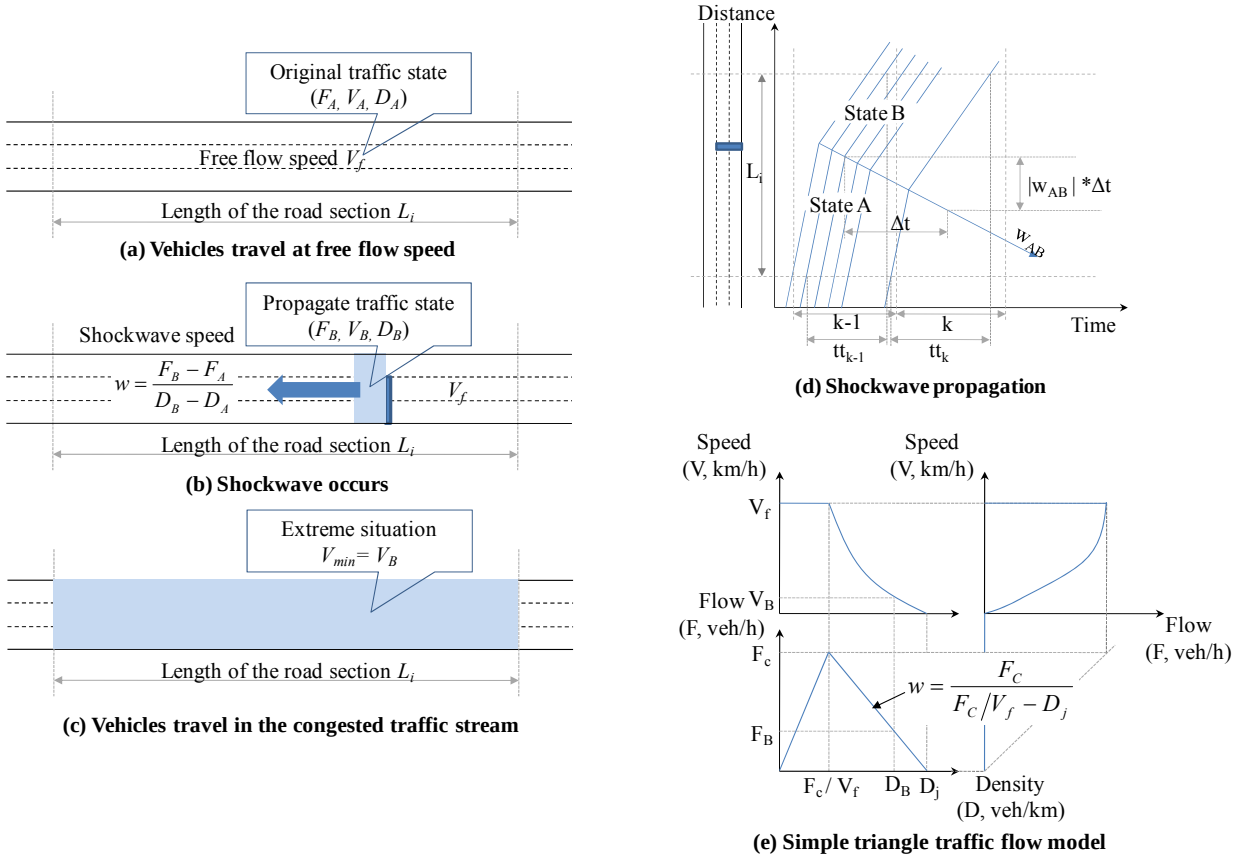


Figure 4.13: Illustration of the traffic characteristics and traffic flow model

Based on the above analysis of travel behaviors, the boundaries of the validity window determined based on a data driven model can be updated as follows:

$$\overline{tt}_{i,lower,k} = max\left(tt_{i,lower,k}, tt_{i,min}\right) \quad (4.1)$$

$$\overline{tt}_{i,upper,k} = min\left(tt_{i,upper,k}, tt_{i,max}\right) \quad (4.2)$$

In addition, the change of travel time over a short time period is restricted by the speed of shockwave propagation. As illustrated in Figure 4.13 (d), travel time increases from $tt_{i,k-1}$ to $tt_{i,k}$ as a result of the shockwave propagating upstream and increasing the proportion of the link occupied by state B. Therefore the increase of travel time during a time period Δt is a function of the speeds in traffic states A and B (V_A and V_B) and the speed of shockwave propagation (w_{AB}). Travel time during interval k ($tt_{i,k}$) can be approximately estimated by Equation 4.3.

$$tt_{i,k} = tt_{i,k-1} + \left(\frac{|w_{AB}| \cdot \Delta t}{V_B} - \frac{|w_{AB}| \cdot \Delta t}{V_A} \right) \tag{4.3}$$

In practice, A and B might be any two different traffic states, for which estimating accurate values for V_A, V_B, F_A, F_B is difficult. However, if state A and state B are assumed as two extreme traffic situations (i.e. free flow state and the most serious congested traffic state as they are defined in Figure 4.13 (a) and (c)), the maximum shockwave speed can be calculated, and consequently the maximum change of travel times between two consecutive intervals can be estimated.

Based on the above assumption, the speeds V_A and V_B in Equation 4.3 are substituted by V_f and V_{min} respectively, and shockwave speed w_{AB} is substituted by the maximum shockwave speed w which can be estimated on the basis of a simple triangle traffic flow model (Figure 4.13 (e)).

Then, the maximum change of travel times between interval $k-1$ and k ($\Delta tt_{i,k-1}$) can be computed as follows:

$$\Delta tt_{i,k-1} = \frac{(V_f - V_{min}) \cdot |w|}{V_{min} V_f} \cdot \Delta t \tag{4.4}$$

With this maximum change of travel times between interval $k-1$ and interval k, the upper bound and lower bound of travel time within interval k can be further updated as follows:

$$\overline{\overline{tt}}_{i,lower,k} = min\left(\overline{tt}_{i,lower,k}, max\left(tt_{i,k-1} - \Delta tt_{i,k-1}, tt_{i,min}\right)\right) \tag{4.5}$$

$$\overline{\overline{tt}}_{i,upper,k} = max\left(\overline{tt}_{i,upper,k}, min\left(tt_{i,k-1} + \Delta tt_{i,k-1}, tt_{i,max}\right)\right) \tag{4.6}$$

To sum up, this valid range is designed to be incorporated into a basic validity window ($tt_{i,lower,k}, tt_{i,upper,k}$) that is determined on the basis of a data driven model. This basic validity window can be determined by TransGuide, D&R1 or any other existing outlier detection algorithm which is unable to track the sudden changes in travel times.

The challenge for implementing the proposed method is to estimate the traffic states (i.e. free flow and congestion). Free flow speed V_f, roadway capacity F_c and jam density D_j can be estimated based on historical data. However, the value of V_{min} varies depending on the level of traffic congestion, making it difficult to select an appropriate fixed value. We propose a method for the real-time calibration of V_{min} (Figure 4.14).

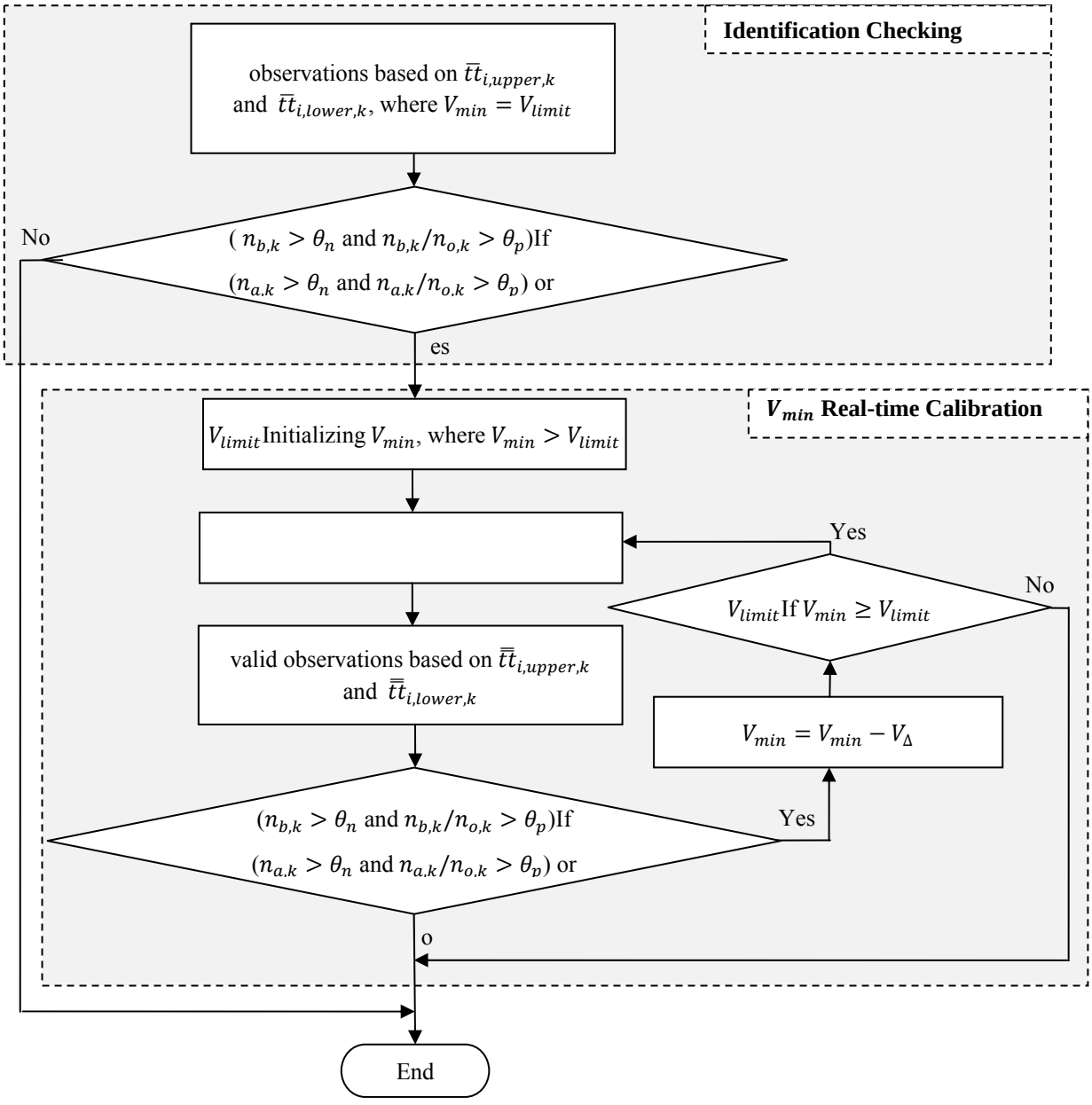


Figure 4.14: Real-time calibration of the V_{min}

As a first step, the method checks if the absolute number and relative fraction of travel time observations in the current period fall outside of the current validity window. This is achieved through the condition:

$$\text{if } (n_{a,k} > \theta_n \text{ and } n_{a,k}/n_{o,k} > \theta_p) \text{ or } (n_{b,k} > \theta_n \text{ and } n_{b,k}/n_{o,k} > \theta_p) \tag{4.7}$$

where, $n_{a,k}$ and $n_{b,k}$ are computed using Equations 4.8 and 4.9 ($\tau_{j,k,a} = 1$ if the measured travel time tt_k is identified as an outlier which is above the upper bound of the validity window, otherwise $\tau_{j,k,a} = 0$; $\tau_{j,k,b} = 1$ if the measured travel time tt_k is identified as an outlier which is below the lower bound of the validity window, otherwise $\tau_{j,k,b} = 0$), which are the number of invalid observations above and below the upper bound and lower bound of the validity window in time interval k respectively; $n_{o,k}$ is the total number of observations within interval k; V_{limit} is a limit value of V_{min} as we don't consider the special situation that the entire road section is closed. Real-time calibration of V_{min} is triggered when the condition in Equation 4.7 is true.

$$n_{a,k} = \sum_{j=1}^{n_k} \tau_{j,k,a} \tag{4.8}$$

$$n_{b,k} = \sum_{j=1}^{n_k} \tau_{j,k,b} \tag{4.9}$$

Once the real-time calibration is triggered, an initial value of V_{min} is given, and $\Delta tt_{i,k-1}$ is calculated based on Equation 4.4. Then $\overline{tt}_{i,lower,k}$ and $\overline{tt}_{i,upper,k}$ are updated using Equations 4.5 and 4.6, respectively. Valid observations are identified based on $\overline{\overline{tt}}_{i,lower,k}$ and $\overline{\overline{tt}}_{i,upper,k}$, and the condition defined in Equation 4.7 is used again to determine whether or not the current value of V_{min} is appropriate. If the condition is true, then the value of V_{min} is reduced, otherwise end the calibration process, and keep the current value of V_{min}. Another condition of $V_{min} \geq V_{limit}$ is used to limit the value given to V_{min}.

The values of parameters (θ_n, θ_p and V_{limit}) relate to the traffic and road conditions in a specific road section, such as traffic flow, road geometry (e.g. whether or not the studied section is close to an on-ramp or off-ramp), and the area of the road section (e.g. urban or rural). Therefore they can be estimated off-line based on the historical data, and be updated periodically (e.g. monthly or yearly) depending on the specific situation.

One way to determine the values of the parameters is to do a statistical analysis using historical data. We can select a sample dataset from historical data for which the existing filtering algorithm has satisfactory performance. For example, consider the application of the TransGuide algorithm to data collected on Oct 5th with parameters k=2 minutes and δ=50% (as illustrated in Figure 4.5 (b)). The cumulative relative frequency of the number of outliers within each interval $(n_{o,k} - n_{v,k})$ and the fraction of the total observations labeled as outliers within each interval $\left((n_{o,k} - n_{v,k})/n_{o,k}\right)$ are computed using 2-minute and 5-minute intervals as shown in Figure 4.15.

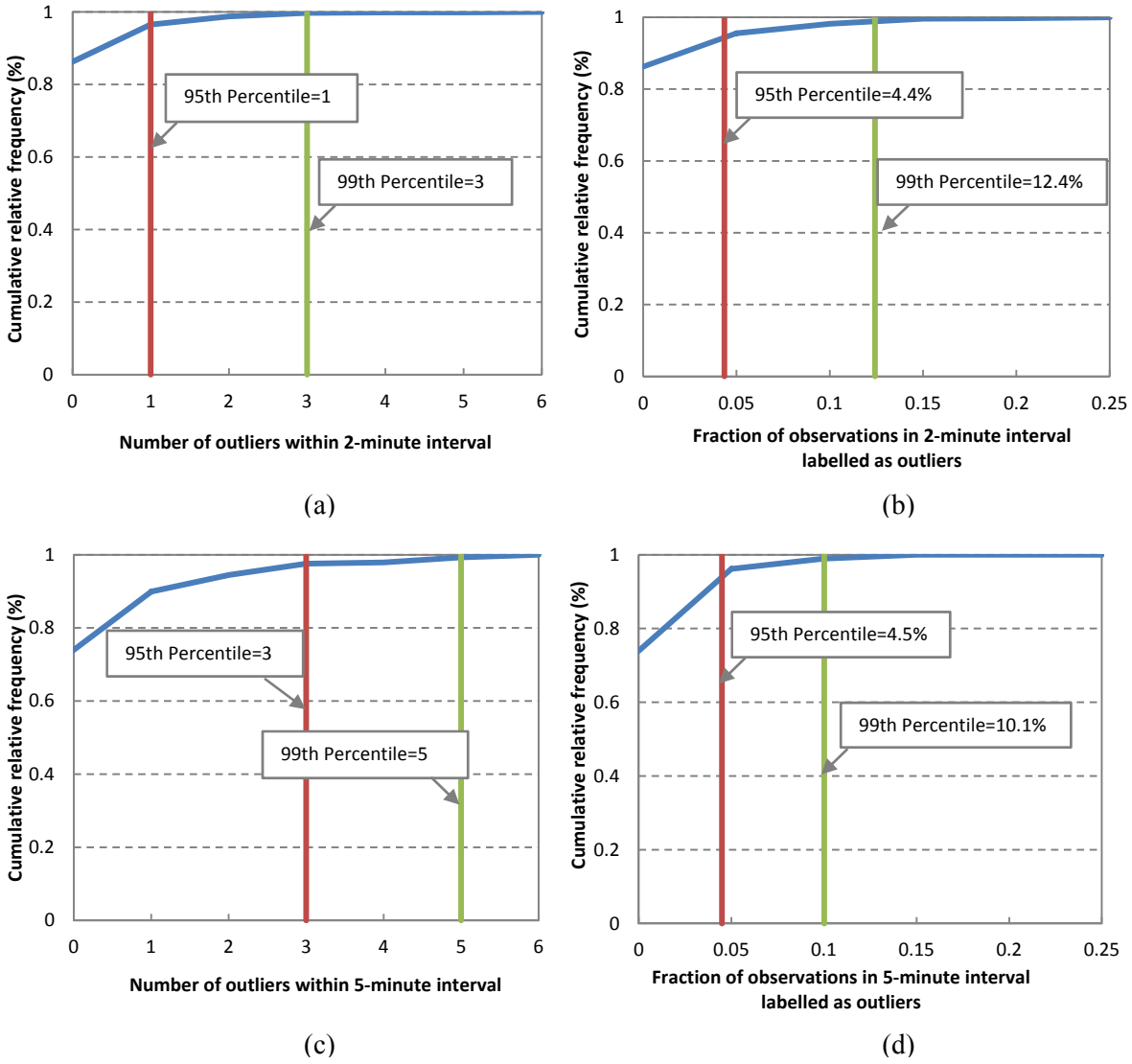


Figure 4.15: Using historical data to select parameter values for proposed model

The results from Figures 6a and 6b indicate that for an interval duration of 2 minutes, less than 1 percent of the intervals had more than 3 outliers, and less than 1 percent of the intervals had more than 12.4% of the total travel times observed in the interval labelled as outliers. For an interval duration of 5 minutes, less than 1 percent of the intervals had more than 5 outliers, and less than 1 percent of the intervals had more than 10.1% of the total travel times labelled as outliers. We assume that there is little possibility that the number of outliers ($n_{a,k}$ or $n_{b,k}$) and the fraction of observations labelled as outliers ($n_{a,k}/n_{o,k}$or $n_{b,k}/n_{o,k}$) will exceed the 99th percentile values obtained from the historical data for this roadway segment, and therefore the recommended values of $\theta_n$and θ_p are θ_n= 3, θ_p= 12.4% for 2-min interval, and θ_n= 5, θ_p= 10.1% for 5-min interval. Of course, in practice, the sample dataset should include results from more than 1 day.

The value of V_{limit}, a minimum speed of vehicles along the road section of interest can be found from the historical data or this value can be selected on the basis of engineering judgment. In the following test, V_{limit} is set to 5 km/h.

4.3 Test and Validation of the Proposed Method

The proposed method is incorporated into the TransGuide and D&R1 algorithms but not the D&R2 algorithm because the proposed extension is designed as an alternative to the modifications made in D&R2 algorithm. The modified algorithms are noted as TransGuide-M and D&R-M in the following descriptions. Applications of TransGuide-M and D&R-M are performed using the same data that were used to test the original TransGuide and D&R algorithms previously.

The following parameter values were selected for the proposed model: free flow speed is set to 110 km/h, road capacity is assumed as 2200 vehicles/h/lane, and jam density is assumed as 125 vehicles/km/lane. $\theta_n$and θ_p are set equal to 3 and 12.4% for 2-min interval, and 5 and 10.1% for 5-min interval. V_{limit} is set to 5 km/h, and V_Δ is set to 10km/h.

For application of the TransGuide-M algorithm, a threshold of 50% is used as it can be seen clearly from Figure 1 that the 20% threshold is not suitable for our data set. The results, illustrated in Figure 4.16 and 4.17 for the two days of data, show that the deficiency in the original TransGuide algorithm in terms of tracking sudden changes of traffic conditions is solved (especially for Nov.5th shown in Figure 4.17).

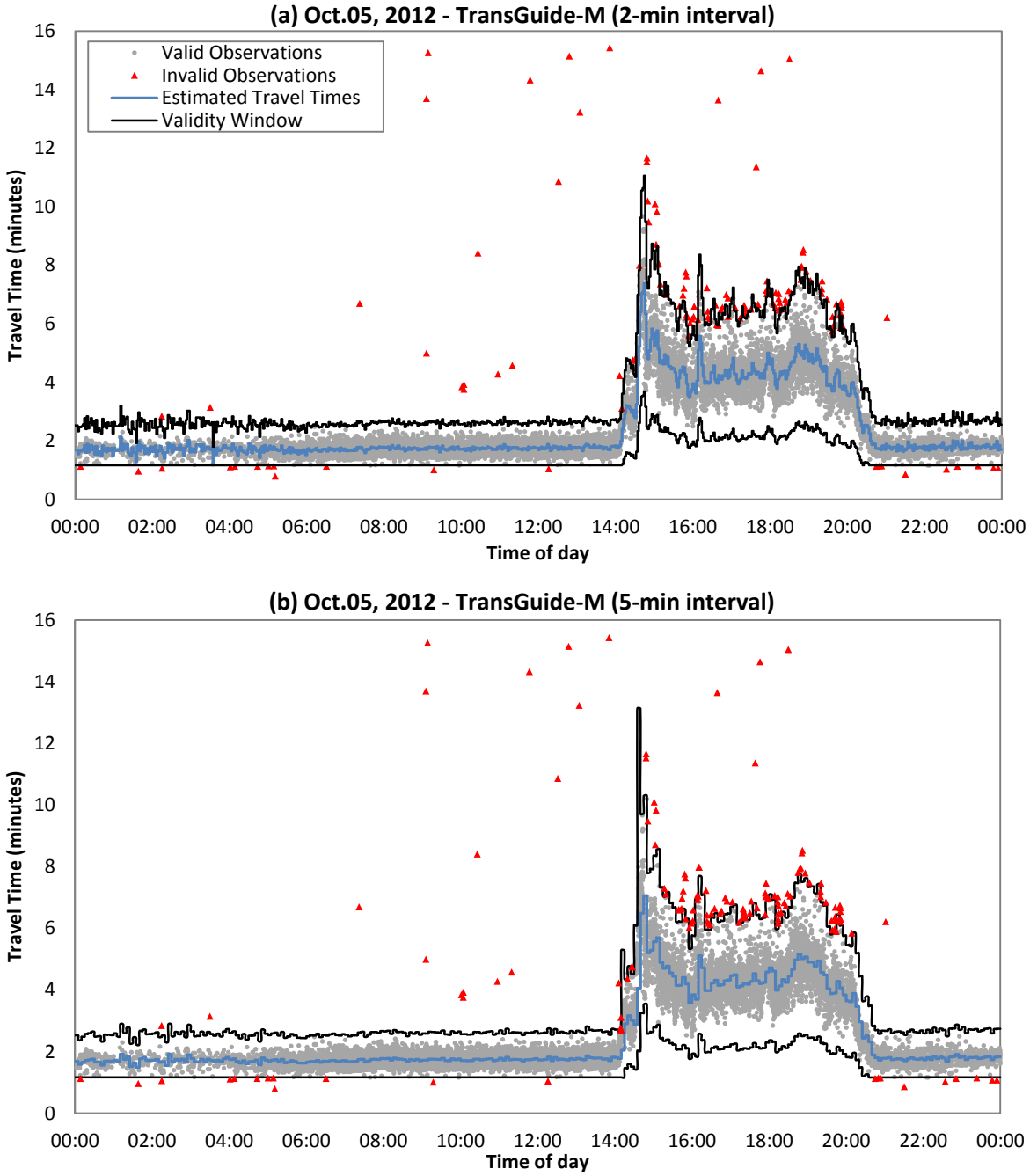


Figure 4.17: Applications of the proposed extension combined with TransGuide (Oct. 05 2012)

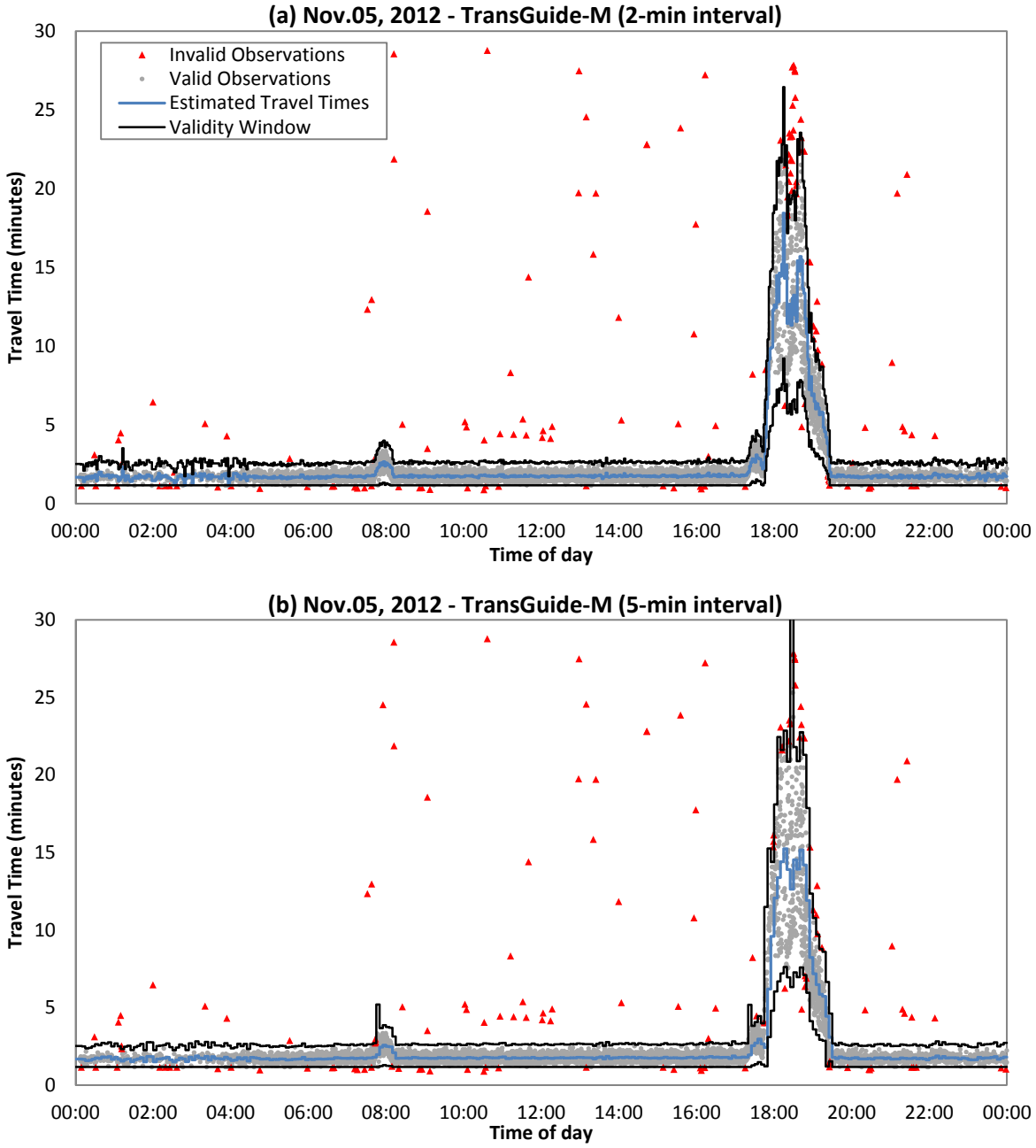


Figure 4.18: Applications of the proposed extension combined with TransGuide (Nov. 05 2012)

The benefit of the proposed outlier detection algorithm extension is demonstrated by comparing the performance of the original TransGuide algorithm with the performance of the modified TransGuide algorithm (Figures 4.18 - 4.20, where the figure (a) shows the performance of the original algorithm and the figure (b) shows the performance of the modified algorithm). For all comparisons, the

deficiencies in performance of the original TransGuide algorithm have been overcome using the proposed extension indicating that the proposed extension makes the TransGuide algorithm much more robust. It can be observed that the modified algorithm still mislabels some of the travel time observations; however, the proportion of incorrectly labelled observations appears to be sufficiently small that there is no substantive impact on the accuracy of estimated mean travel times.

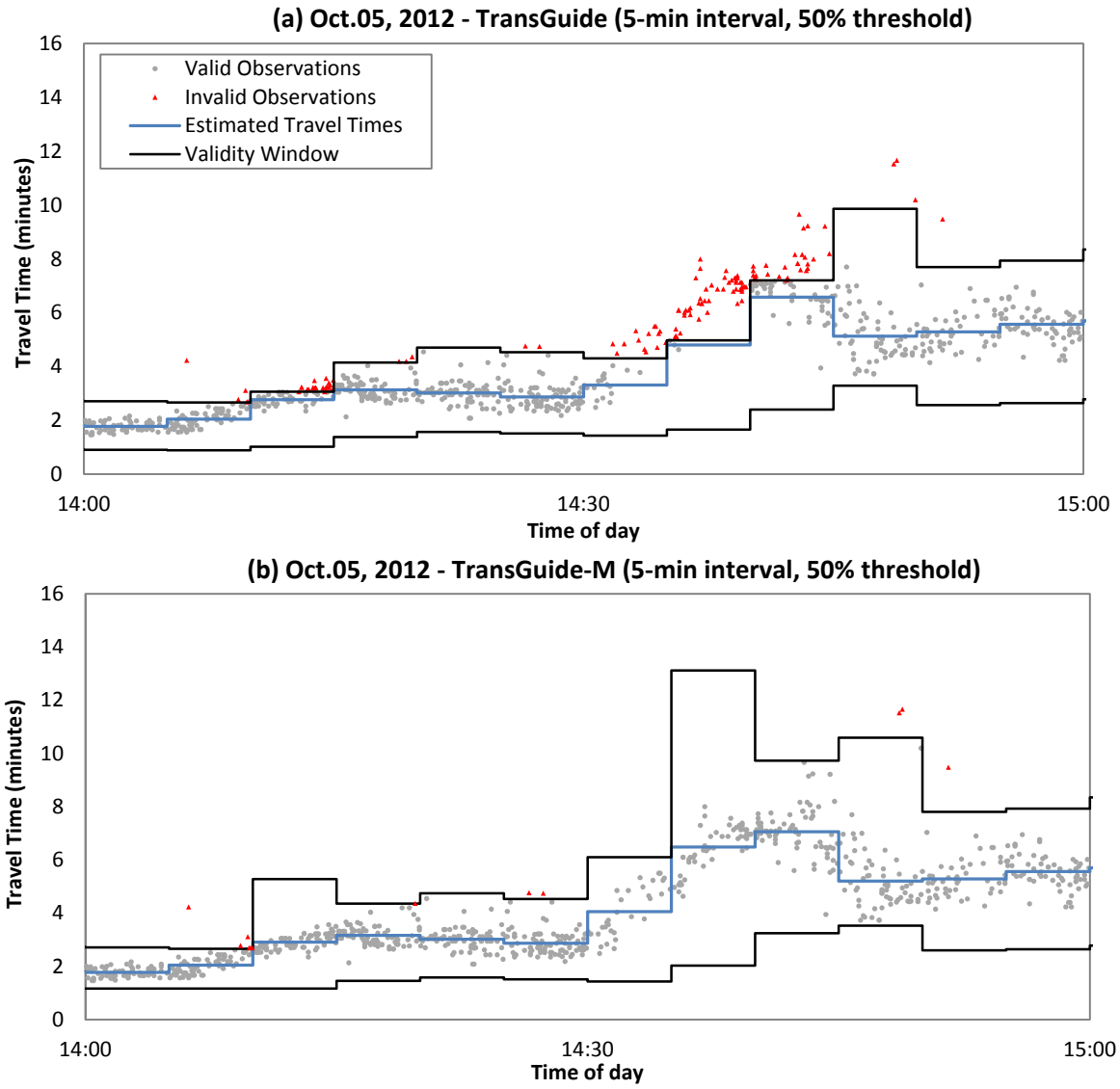


Figure 4.19: Comparisons between original TransGuide and modified TransGuide algorithms based on data collected from Oct. 05 2012 (5-min interval)

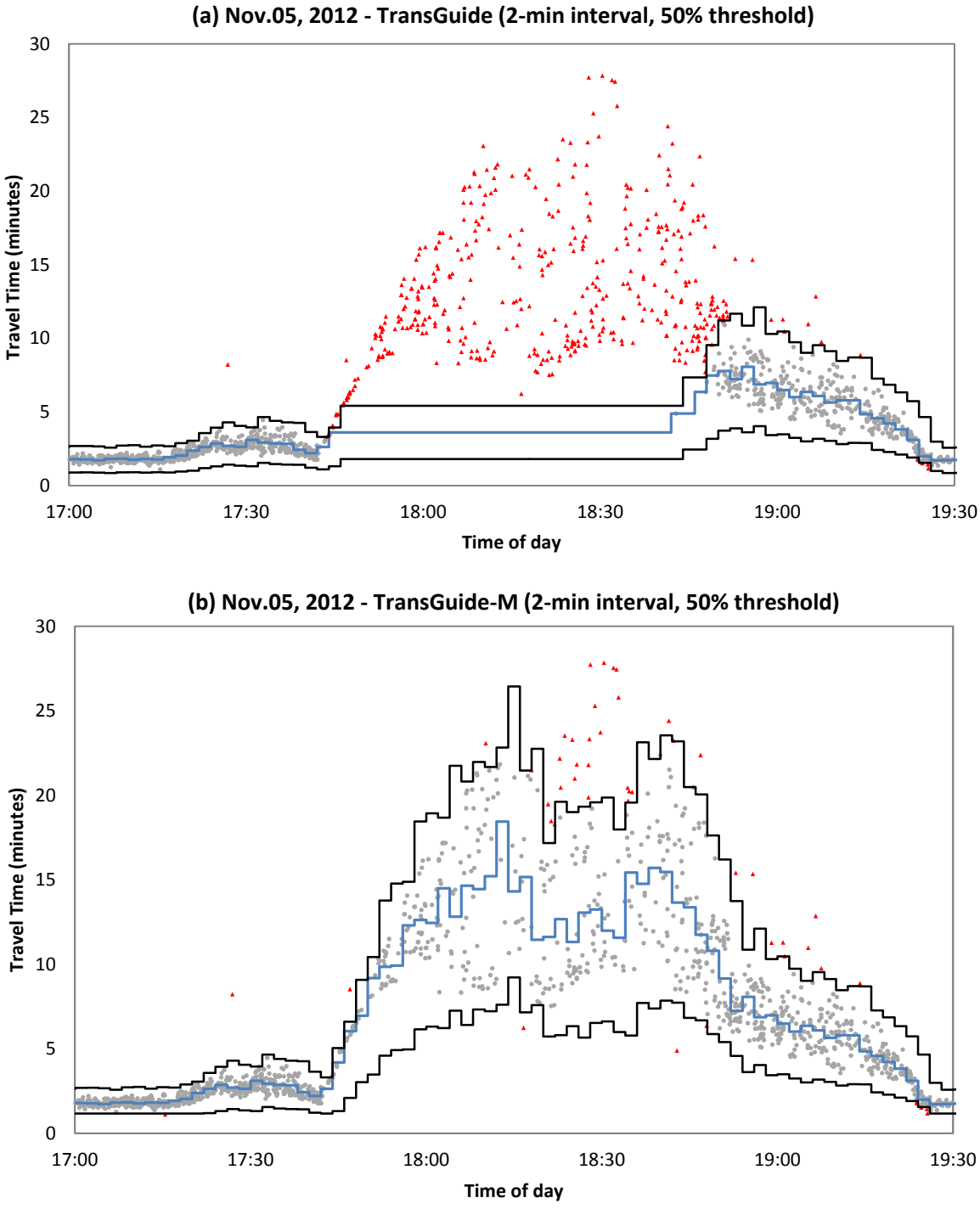


Figure 4.19: Comparisons between original TransGuide and modified TransGuide algorithms based on data collected from Nov. 05 2012 (2-min interval)

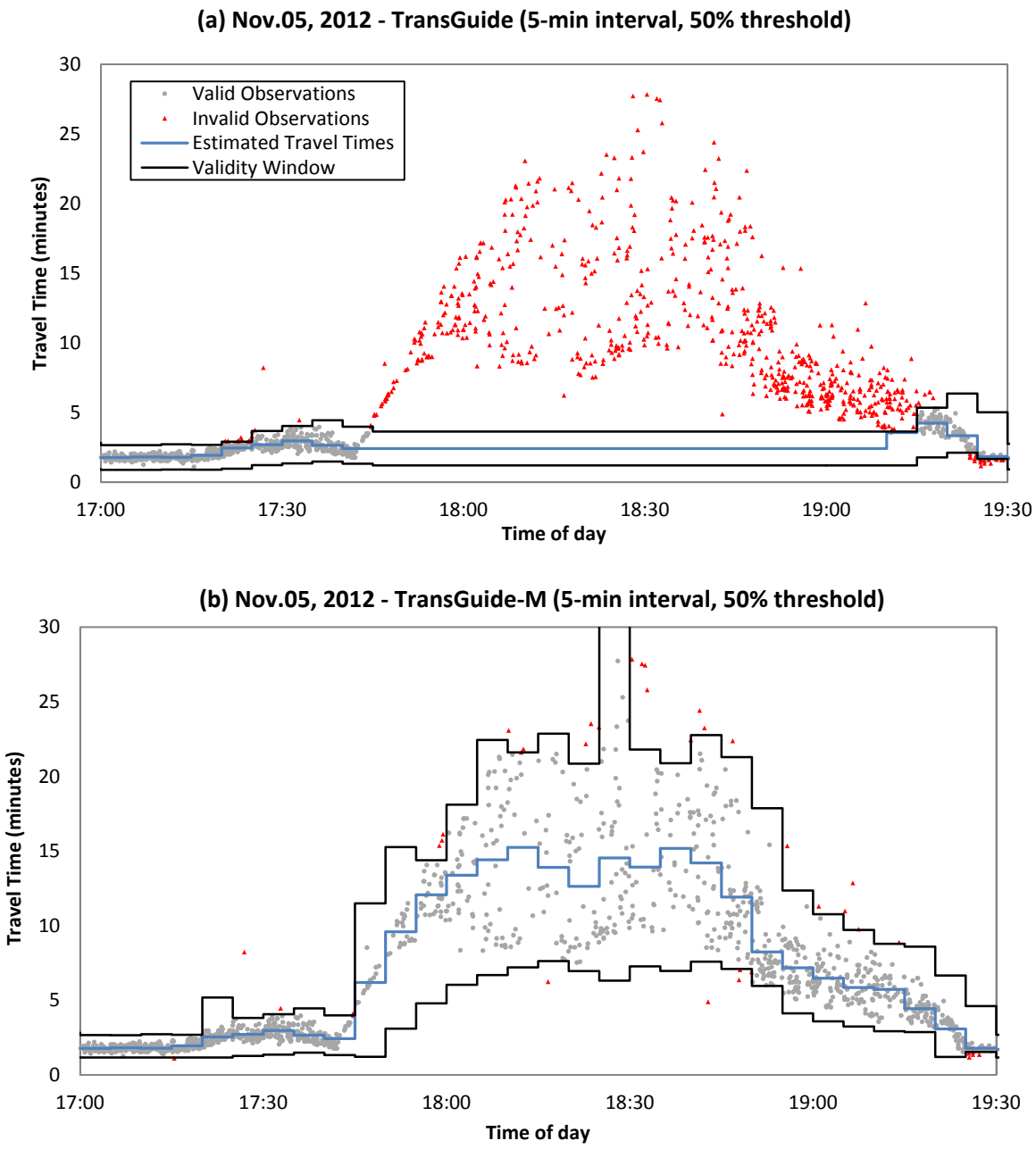


Figure 4.20: Comparisons between original TransGuide and modified TransGuide algorithms based on data collected from Nov. 05 2012 (5-min interval)

The same methodology is used to evaluate the effect that the proposed extension has on the D&R algorithm. Figures 4.21 and 4.22 show the overall results, and Figures 4.23-4.25 show the detailed results.

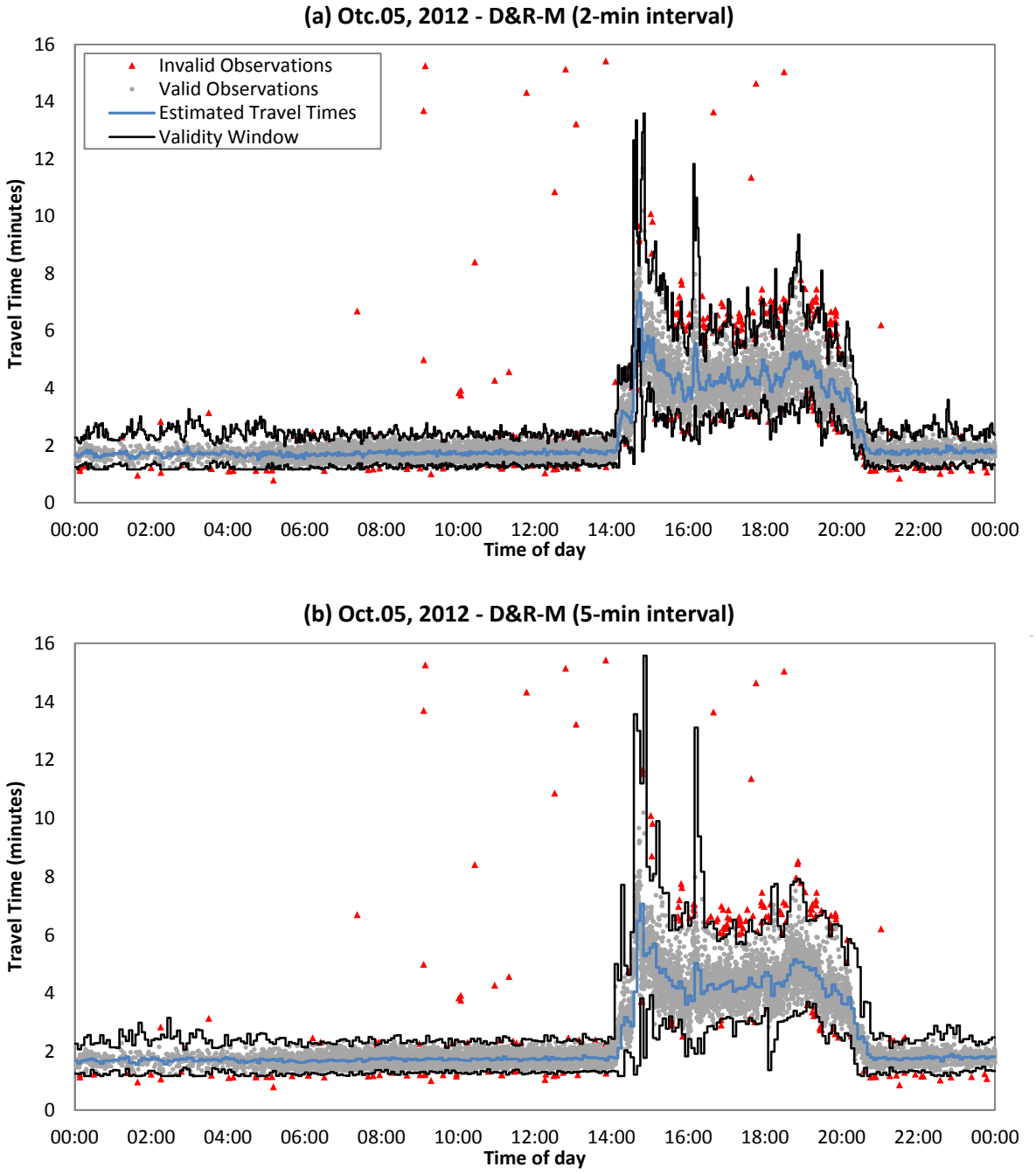


Figure 4.21: Applications of the proposed extension combined with D&R1 (Oct. 05 2012)

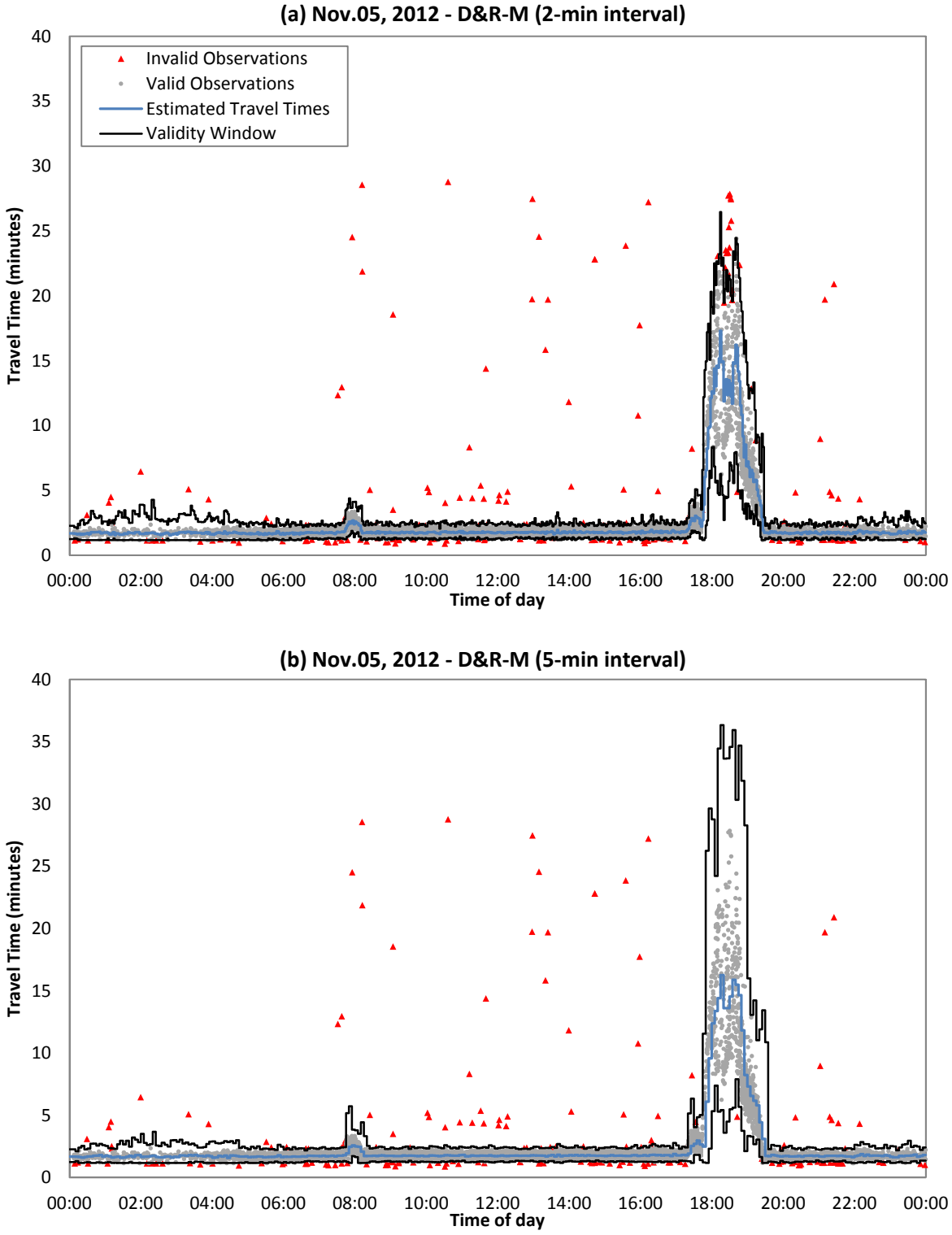


Figure 4.22: Applications of the proposed extension combined with D&R1 (Nov. 05 2012)

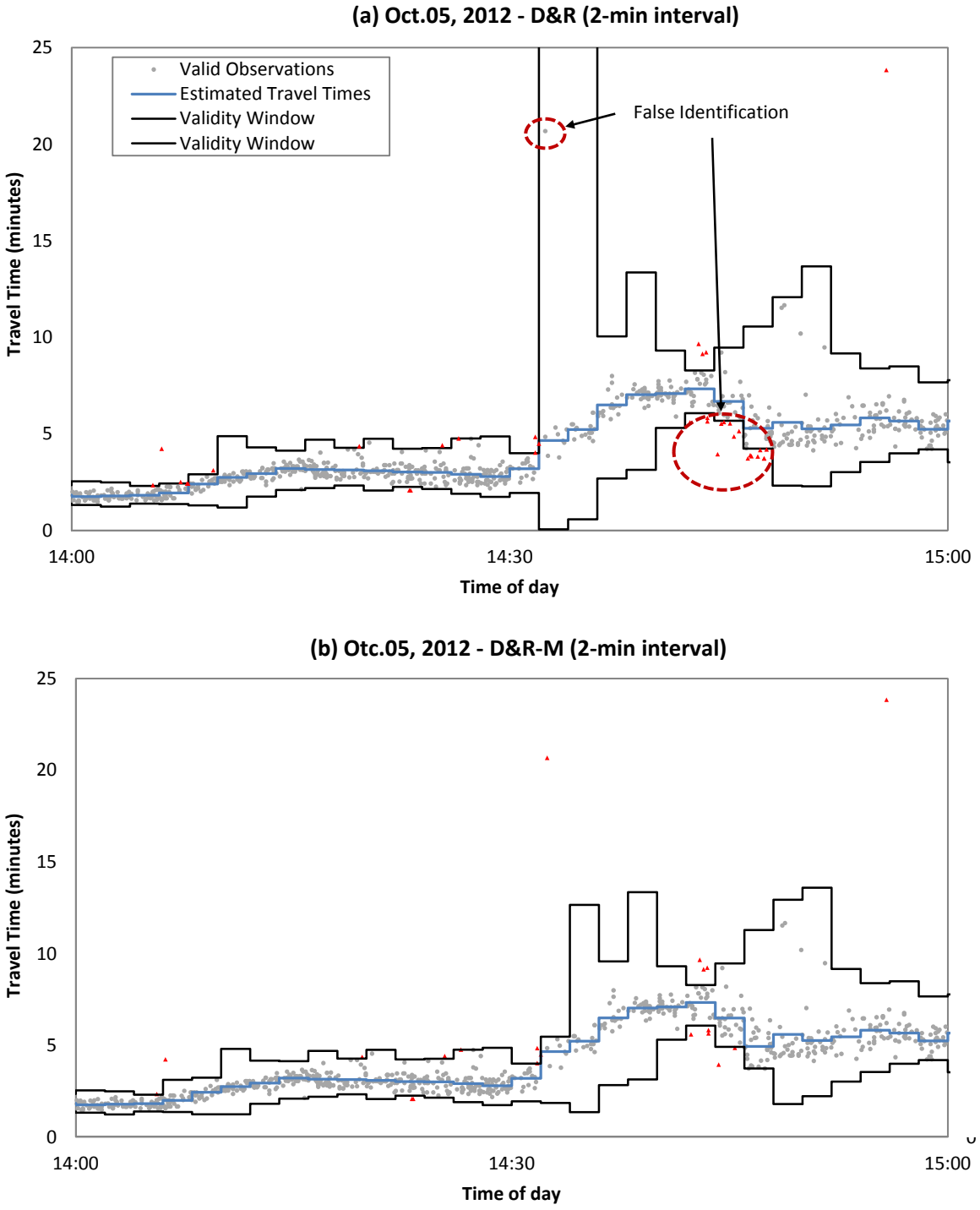


Figure 4.23: Comparisons between D&R2 and modified D&R algorithms based on data collected from Oct. 05 2012 (2-min interval)

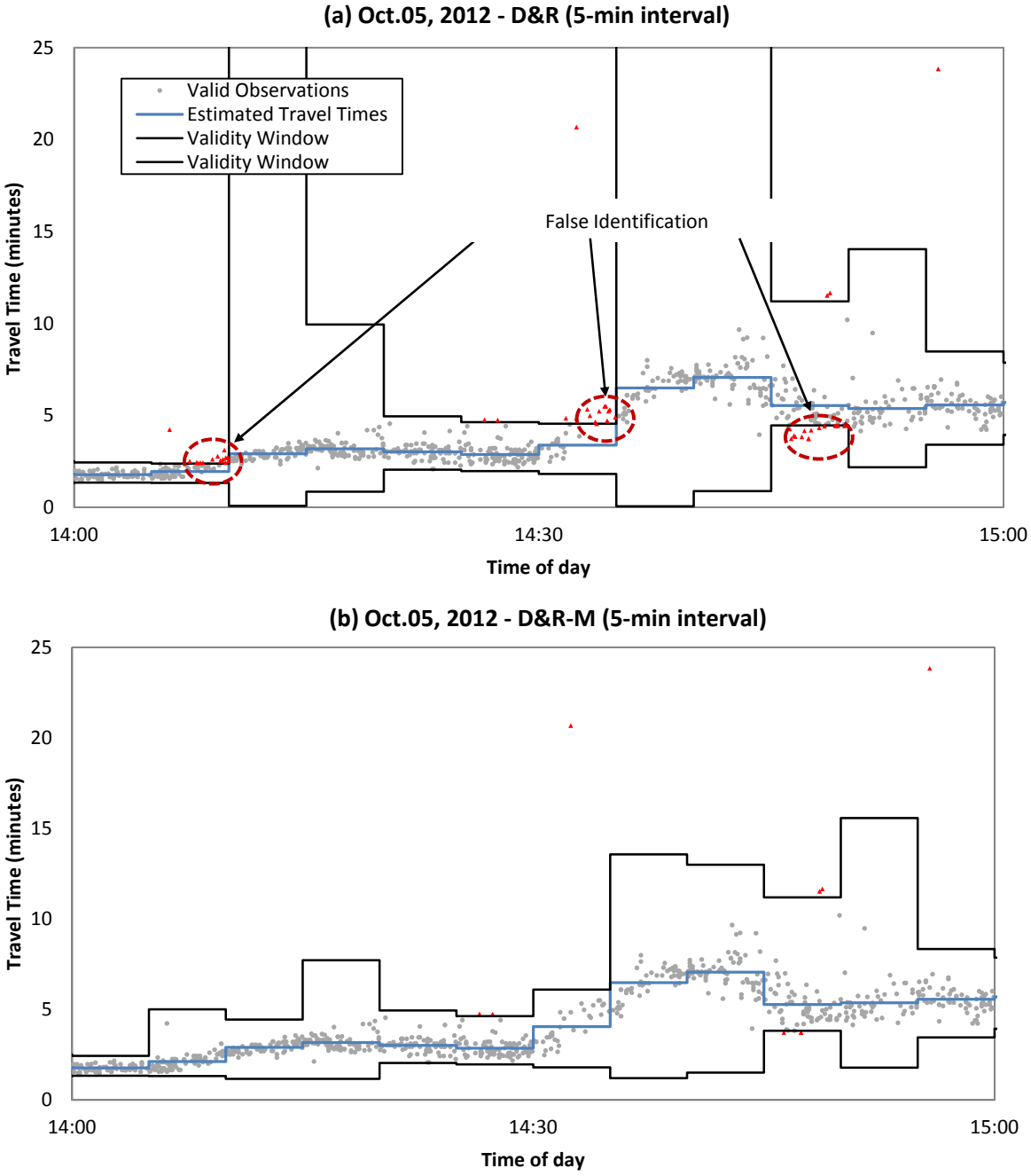


Figure 4.24: Comparisons between D&R2 and modified D&R algorithms based on data collected from Oct. 05 2012 (5-min interval)

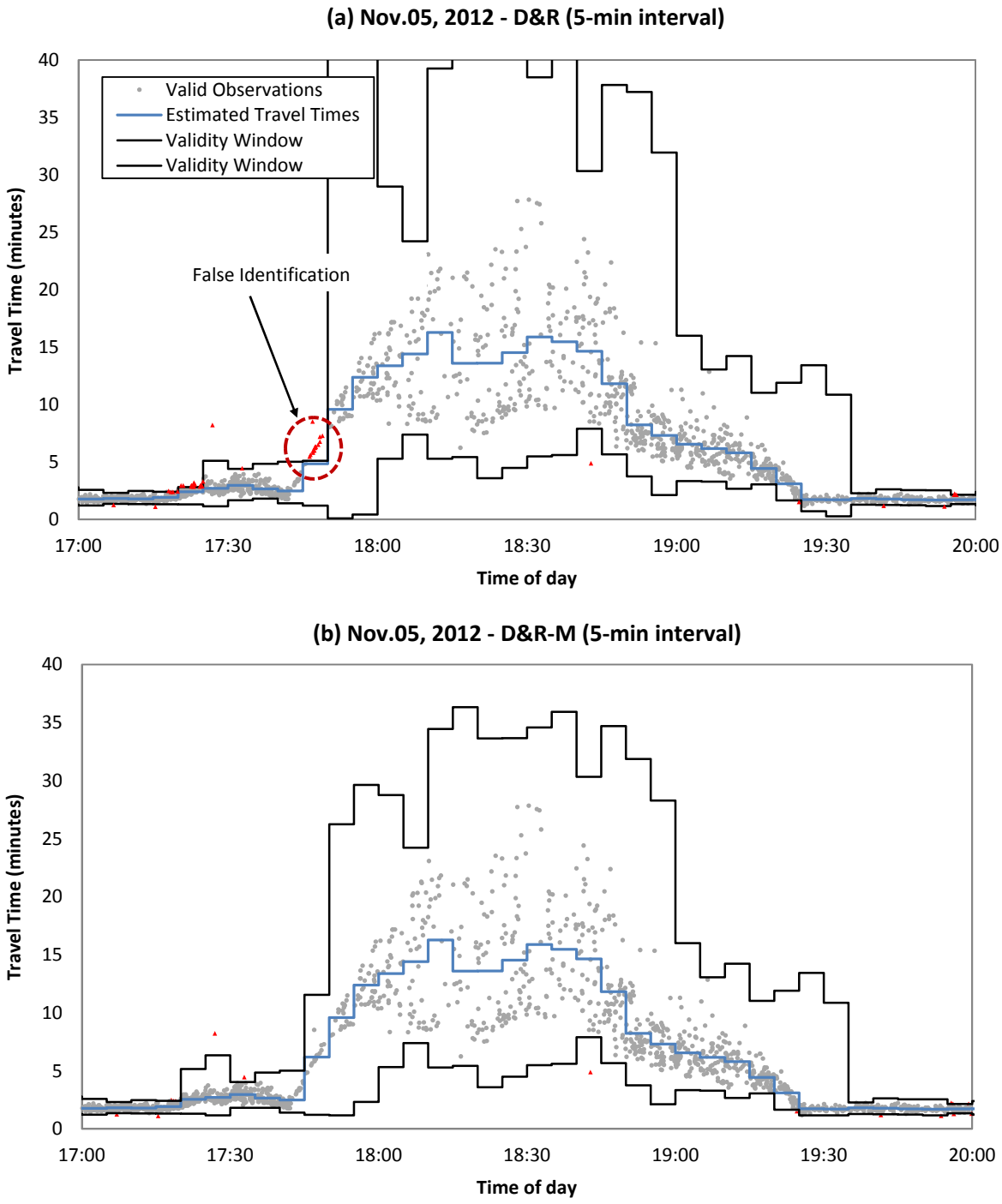


Figure 4.25: Comparisons between D&R2 and modified D&R algorithms based on data collected from Nov. 05 2012 (5-min interval)

The results indicate that using the proposed extension instead of the extension developed by Dion and Rakha (i.e. D&R2), can increase the responsiveness of the filtering model to rapid changes of travel times, and at the same time avoid the problem of an unrealistically large validity window. The problem of large amount of valid data being excluded from the validity window for some specific intervals is solved as well, due to the utilization of real-time calibration and correction.

4.4 Summary

In this chapter, a traffic flow theory based travel time outlier filter enhancement is proposed as an extension to existing data driven outlier detection algorithms.

The proposed method improves the performance of existing data driven outlier detection algorithms for periods when travel times are changing rapidly, as when congestion is forming or dissipating. The parameters in the proposed method can be determined based on historical data and the model incorporates self-calibration to increase robustness and transferability of the model.

The proposed model is suitable for off-line and real-time applications in which outliers are to be identified from individual measured travel times. The performance of the proposed method was illustrated by incorporating the model into the TransGuide and D&R1 algorithms and applying the models to Bluetooth data collected over a 3.1 km section of suburban freeway.

The improved performance of the proposed model is demonstrated through depiction of the data labeling results (i.e. valid versus outlier). A quantitative assessment of the improvement in performance was not possible using field data because the truth (i.e. which observations are actually outliers) is not known.

The proposed filtering method is used as one component of the short-term travel time prediction framework developed in this study, and it is discussed in Chapter 6.

Chapter 5
Selecting Historical Data for Short-term Travel Time Prediction

Irrespective of the nature of the model, almost all the prediction models described in the literature have incorporated historical data in some manner to improve the accuracy of prediction results.

Rice and Van Zwet (2004) developed a linear regression model with varying parameters combining current traffic situation with historical data. Chien and Kuchipudi (2003) performed travel time prediction on the basis of Kalman filter techniques using real-time data combining with aggregated historical data from automatic vehicle identification (AVI) systems. Guin (2006) proposed an ARIMA time series model to predict future travel times using historical travel time data, and he concluded that the current traffic data are able to provide good travel time prediction for the near future, while historical data being used in a time series model provide more accurate predictions for the longer term. A Bayesian inference-based dynamic linear model was proposed by Fei et al. (2011) to predict short-term freeway travel times, where the median of historical travel times was employed to recognize the primary travel time patterns. Based on an expert system which assigns weights to historical and real-time traffic information dynamically according to "expert" knowledge, Lee et al. (2009) found that the combined model that assimilates historical and current information has better performance than the two separate models in terms of the accuracy of prediction results.

Historical data have long been considered as an important input to travel time prediction models because the time series of traffic states (e.g. travel time, traffic flow) collected from different days at the same site in the same situation have similar time-varying traffic patterns (Chen et al., 2012).

5.1 Methods of Selecting Historical Data

5.1.1 Simple Aggregation (SA)

Most commonly, the traffic pattern from historical data is estimated by aggregating travel times from past consecutive days, and the patterns for work days and weekends/holidays are separately estimated. According to the number of the days being aggregated, weekly/monthly aggregated historical data are usually used.

This method of aggregating travel times from past consecutive days is attractive because it is intuitive, simple to implement and easy to understand. However, this method can only provide a

primary pattern, and the large day-to-day variations in travel times caused by variations in capacity (supply) and demand cannot be captured. In the prediction models that combine current information with historical information, the day-to-day variation in travel times is expected to be captured by current data observed in previous time interval(s), however, the real-time travel time observations usually suffer from the problems of missing data, data unavailability or sampling error caused by low sample rate, and these problems usually occur when travel time change substantially. Therefore, to better capture the day-to-day variation in travel times, it's necessary to obtain meaningful travel time patterns from historical data by selecting only those historical travel time data that are most similar to the conditions of the current day.

5.1.2 Nearest Neighbor (KNN)

K Nearest Neighbors (KNN) technique is one way to find a sub-set, which contains only those data that are most similar to the conditions observed so far on the current day from the entire set of historical data. Traffic predictions based on KNN were investigated in previous studies (Smith et al., 2002; Wild, 1997), where the KNN is used to find K days (K is an optimal number of days in the selected sub-set to minimize the prediction error) in the past that are most similar to the present day in some appropriate sense. Once the historical days have been identified, then the conditions observed after the current time during these historical days are used to predict the future conditions for today. The prediction based on KNN is actually a non-parametric regression method, in which the effectiveness of this method is heavily influenced by the quality of the historical database. If the similar traffic conditions are not present in the historical database it's difficult to predict future traffic conditions accurately (Smith et al., 2002).

The key in using the KNN technique is finding a suitable measurement to represent the similarity (normally termed the "distance") between the traffic conditions of the present day and conditions in the past days. Traffic conditions are represented by sequences of traffic state values, and each value represents an average level of the traffic states over a road segment during a time interval, so the traffic condition is actually a type of time series data (e.g. travel time, traffic flow, average speed, etc.). A general method used to measure this "distance" in previous studies (Smith et al., 2002; Clark, 2003; Rice et al., 2004; Chen et al., 2012) is to compute the Euclidean distance between two time series. For example, we have a discrete time series of travel times, $a = \{\hat{tt}_{k-N_k+1}, \dots, \hat{tt}_{k-1}, \hat{tt}_k\}$ measured today, and each measurement in the time series is obtained at a constant time interval - say every 5 minutes. N_k is the number of the data points contained in the time series. The same discrete

time series exists in each of the historical days $h_d = \left\{\left(\hat{tt}_{k-N_k+1}\right)_d, \ldots, (\hat{tt}_{k-1})_d, (\hat{tt}_k)_d\right\}$, where $d = \{1,2, \ldots, D\}$ and D is the total number of days in the historical data. The Euclidean distance $Dist_d(a,\ h_d)$ between time series a and h_d can be computed by Equation 5.1:

$$Dist_d(a,\ h_d) = \sqrt{\sum_{n=1}^{N_k} [\hat{tt}_{k-n+1} - (\hat{tt}_{k-n+1})_d]^2} \tag{5.1}$$

By using this distance measurement (Equation 5.1), we can identify the degree of similarity between the time series of today with each of the historical time series and select the K most similar ones.

5.2 Performance Comparison between SA and KNN

Aggregation of historical data based on a subset selected using KNN method is expected to provide better estimation of traffic pattern than the simple aggregation of historical data from past consecutive days. To verify this assumption, two naïve models based on the subsets of historical data selected using different methods (i.e. SA & KNN) were applied to predict travel times.

The naïve models used in the tests are described as follows:

Naïve 1:

$$\hat{tt}_{k+1} = \hat{tt}_k + \Delta\overline{tt}_k \tag{5.2}$$

Where, $\hat{tt}_k$ represents travel time of today within time interval k; $\hat{tt}_{k+1}$ is the predicted travel time for time interval $k+1$; $\Delta\overline{tt}_k$ represents the average of change in travel time observed at time interval k in historical data. The value of $\Delta\overline{tt}_k$ is computed differently for the SA and KNN methods.

For the SA (Simple aggregation) methods, $\Delta\overline{tt}_k$ is designated as $\Delta\overline{tt}_{k-SA}$ and is computed as:

$$\Delta\overline{tt}_{k-SA} = \frac{\sum_{d=1}^{D_s}(\hat{tt}_{k+1} - \hat{tt}_k)_d}{D_s} \tag{5.3}$$

Where $(\hat{tt}_{k+1} - \hat{tt}_k)_d$ is the trend of travel time in historical day d.

For the example in Figure 5.1, $\Delta\overline{tt}_{k-SA}$ is computed as $[(10.6 - 6.1) + (11.7 - 8.5)]/2 = (4.5 + 3.2)/2 = 3.85$. For the KNN method, $\Delta\overline{tt}_k$ is designated as $\Delta\overline{tt}_{k-KNN}$ and is computed as:

$$\Delta\bar{tt}_{k-KNN} = \frac{\sum_{d=1}^{D_s}(\hat{tt}_{k+1} - \hat{tt}_k)_d / Dist_d}{\sum_{d=1}^{D_s} 1/Dist_d} \quad (5.4)$$

Where, $Dist_d$ represents the dissimilarity error (i.e. "distance") between the time series of observations from the current day and the time series from historical day d; D_s is the total number of historical days selected in the sub-set. If for the example data in Figure 5.1, $Dist_{d1} = 2.5$ and $Dist_{d2} = 1.3$ then, $\Delta\bar{tt}_{k-KNN} = \frac{[(10.6-6.1)/1.3+(11.7-8.5)/2.5]}{(1/1.3+1/2.5)} = (3.45 + 1.28)/1.17 = 4.04$.

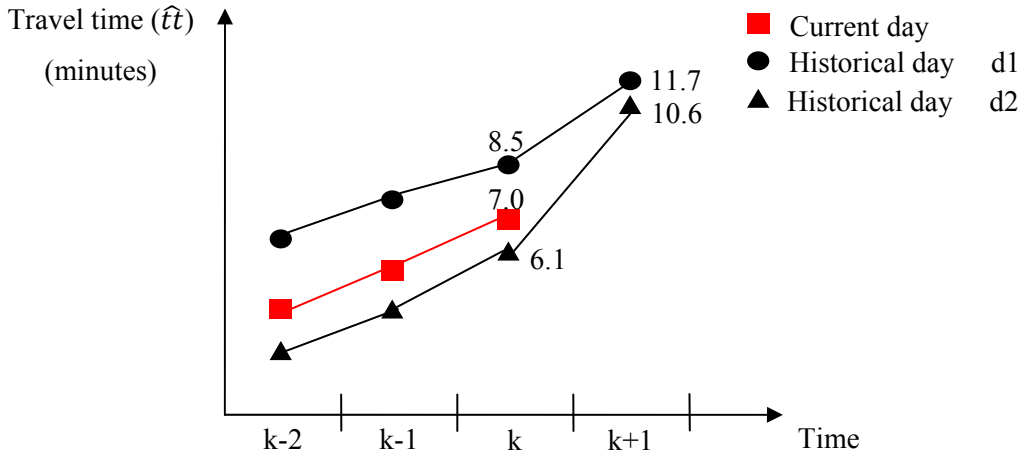


Figure 5.1: An example for illustrating the calculation of the average change in travel time

Naïve 2:

$$\hat{tt}_{k+1} = \hat{tt}_k \times R_{\bar{tt}_k} \quad (5.5)$$

Where, $R_{\bar{tt}_k}$ represents a trend of travel time observed at time interval k in historical data as well, but it is a ratio of the travel time between two consecutive intervals.

For the SA method:

$$R_{\bar{tt}_{k\_SA}} = \frac{\sum_{d=1}^{D_s}(\hat{tt}_{k+1}/\hat{tt}_k)_d}{D_s} \quad (5.6)$$

For the KNN method:

$$R_{\bar{tt}_{k\_KNN}} = \frac{\sum_{d=1}^{D_s}(\hat{tt}_{k+1}/\hat{tt}_k)_d / Dist_d}{\sum_{d=1}^{D_s} 1/Dist_d} \quad (5.7)$$

The above two naïve models follow the logic that the trend of travel time in the corresponding interval of selected historical data is a good indicator of the trend of travel time for the current day over the next time interval. The average difference and average ratio of travel times between the current time interval and the previous time interval are used to represent the trend of travel time in the two naïve models respectively.

The accuracy of the prediction is quantified in terms of the difference between the predicted travel time and the observed travel time, and is measured by the following two error indices:

The Mean Absolute Relative Error (MARE):

$$MARE_{m,t} = \frac{1}{D}\sum_{d=1}^{D} ARE_{m,d,t} \tag{5.8}$$

$$ARE_{m,d,t} = \frac{\left|\hat{a}_{m,d,t} - a_{d,t}\right|}{a_{d,t}} \tag{5.9}$$

Where: $ARE_{m,d,t}$ is the absolute relative error using method m to select the historical data, i.e. m=SA or m=KNN, on current day d and prediction horizon t; $MARE_{m,t}$ is the mean absolute relative error when selecting historical data using method m for prediction horizon t; $\hat{a}_{m,d,t}$ is the travel time predicted by method m for day d for time horizon t; and $a_{d,t}$ is the actual travel time observed on day d at time corresponding to the predicted travel time; and D is the total number of days compared.

The Mean Absolute Error (MAE):

$$MAE_{m,t} = \frac{1}{D}\sum_{d=1}^{D} AE_{m,d,t} \tag{5.10}$$

$$AE_{m,d,t} = \left|\hat{a}_{m,d,t} - a_{d,t}\right| \tag{5.11}$$

Where: $AE_{m,d,t}$ is the absolute error using method m to select the historical data for prediction horizon t; $MAE_{m,t}$ is the mean absolute error using method m to select the historical data for prediction horizon t.

Data used to do the following comparisons are nine years of travel time data (non-holiday weekdays) from 2001 to 2009 for a section of the A-12 motorway (24.5 km) in the Netherlands. The travel times are 15-minute average time determined using 15-minute aggregate loop detector data and the "trajectory" method. For the purposes of short-term travel time prediction, it would be

advantageous to use travel times aggregated over a shorter time horizon (e.g. 1-minute or 5-minutes). However, for the purpose of evaluating the methods of selecting historical data, the 15-minute aggregated data are sufficient.

We elect to carry out the travel time prediction for the afternoon peak hour (5:00 pm to 6:00 pm) of the total 255 non-holiday weekdays in 2009. Variation of the mean, 90th percentile and 10th percentile of the total 255 days are shown in Figure 5.2, and the cumulative distribution of travel time during peak hour is shown in Figure 5.3. The 255 days are considered as the "current" day respectively during the following tests, and approximately one year data (i.e. 260 days) before the "current" day are considered as the historical database. Travel time predictions are made for 15 minutes, 30 minutes, 45 minutes and 1 hour into the future (i.e. 5:15pm, 5:30pm, 5:45pm and 6:00pm).

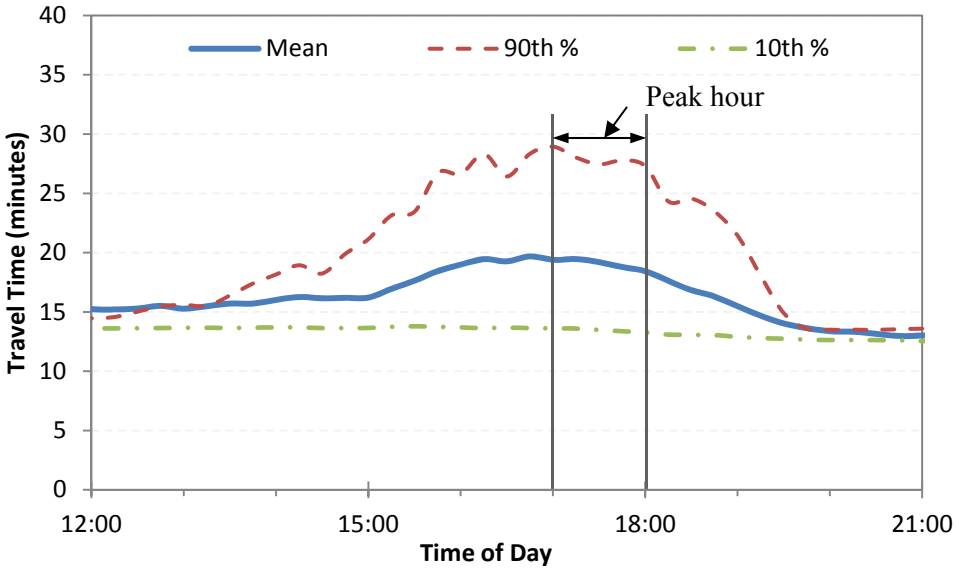


Figure 5.2: Variation of the travel times on A-12 from 12 noon to 9 pm in 2009

Travel time prediction is more challenging when conditions exhibit large variations. Moreover the true benefit of the KNN method for selecting an appropriate subset of historical data is expected to occur on those days when something unusual happens (e.g. non-recurrent congestion), because it is during these time periods of non-recurrent congestion that travel time predictions are most valuable. It isn't expected that the KNN method will perform much better than the simple average (SA) method in terms of selecting historical data for normal traffic conditions, such as free flow speed, or even a recurrent congestion.

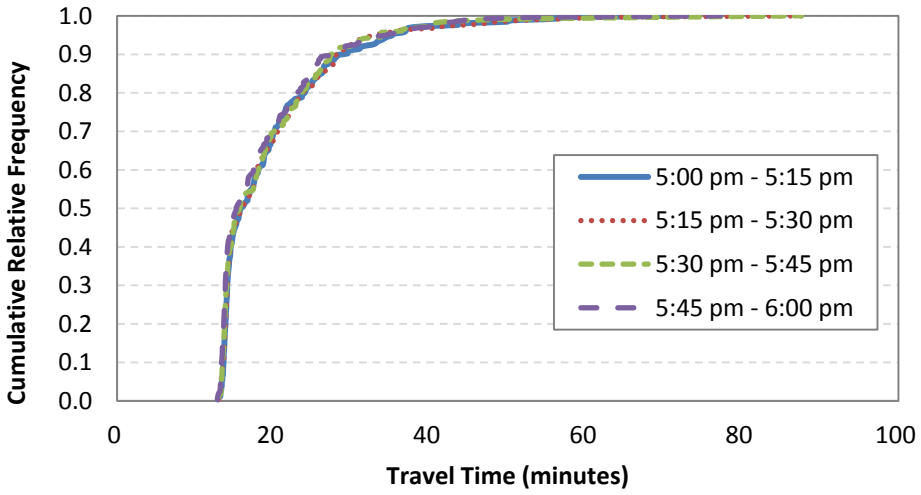


Figure 5.3: Cumulative distribution of travel time during peak hour

Tests are conducted based on travel time predictions during PM peak hours because these periods experience relative large day-to-day variations in the travel times, and therefore pose a greater challenge. Moreover the true benefit of the KNN method being used to select the similar historical data is expected on those days when something unusual happens,

Two options regarding the number of the days selected for the historical sub-set are tested, namely one single day and 5 days. For the simple average method, the one single day is the same weekday in previous week (i.e. if today is Tuesday, then the day from the Tuesday in the previous week is selected), and the 5 days are the previous 5 consecutive non-holiday weekday days (i.e. weekly average). For the KNN method, the single most similar day and the 5 most similar days identified by KNN method are selected respectively. The distance measurement (i.e. Equation 5.1) is computed over the period from 2 pm to 5 pm (i.e. twelve 15-minutes observations).

Table 5.1 provides the comparison results of travel time prediction accuracy by the two naïve models based on historical data selected by SA method and KNN method. The table is divided into two main sections. The top provide travel time prediction in terms of absolute relative error and the bottom in terms of absolute error. Each row labeled as "percent of improvement" (shaded) indicates the prediction accuracy improvement achieved by using KNN rather than SA to select the historical data. Positive values indicate KNN provides better performance.

Table 5.1: Results of comparisons between predictions based on SA and KNN methods

Prediction methods	# of days in the selected sub-set	Methods of selecting historical data	Absolute relative error (ARE)							
			MARE				90^{th} percentile ARE			
			Prediction horizon (min)				Prediction horizon (min)			
			15	30	45	60	15	30	45	60
Naïve 1	1	SA	14.5%	17.6%	22.8%	26.1%	37.1%	43.9%	57.1%	59.5%
		KNN	9.8%	11.6%	13.2%	17.3%	24.8%	30.3%	33.9%	41.2%
		Percent of improvement[5]	32.5%	33.7%	42.0%	33.9%	33.1%	31.0%	40.7%	30.8%
	5	SA	10.5%	11.8%	15.8%	18.5%	22.4%	25.4%	36.6%	43.6%
		KNN	7.7%	9.6%	12.1%	13.7%	20.9%	23.4%	31.1%	32.3%
		Percent of improvement	26.8%	18.6%	23.7%	26.3%	6.9%	7.8%	15.0%	26.0%
Naïve 2	1	SA	11.6%	14.6%	18.7%	21.2%	28.6%	37.2%	46.7%	48.2%
		KNN	9.7%	11.8%	13.0%	16.7%	24.7%	30.9%	32.4%	40.4%
		Percent of improvement	16.0%	19.7%	30.5%	21.2%	13.4%	17.0 %	30.6%	16.3%
	5	SA	9.3%	11.0%	14.6%	16.9%	21.4%	24.3%	33.4%	39.9%
		KNN	7.6%	9.5%	11.8%	13.3%	20.5%	22.7%	31.1%	32.0%
		Percent of improvement	18.6%	13.2%	19.0%	21.2%	3.9%	6.3%	7.1%	19.7%
Prediction methods	# of days in the selected sub-set	Methods of selecting historical data	Absolute error (AE)							
			MAE				90^{th} percentile AE			
			Prediction horizon (min)				Prediction horizon (min)			
			15	30	45	60	15	30	45	60
Naïve 1	1	SA	3.1	3.6	4.5	5.1	8.2	9.7	12.2	14.4
		KNN	2.4	2.7	2.9	3.8	6.8	6.7	7.6	9.9
		Percent of improvement	22.6%	23.7%	35.7%	25.6%	17.5%	31.3%	37.3%	31.2%
	5	SA	2.3	2.5	3.3	3.9	5.6	5.7	8.2	9.3
		KNN	1.9	2.2	2.7	3.1	4.9	5.4	7.0	7.9
		Percent of improvement	18.0%	15.4%	19.8%	21.0%	12.4%	4.3%	15.4%	14.8%
Naïve 2	1	SA	2.7	3.2	3.9	4.3	6.6	7.8	11.1	10.5
		KNN	2.4	2.8	2.9	3.6	6.3	6.8	7.6	9.7
		Percent of improvement	10.5%	11.1%	26.4%	16.2%	3.2%	12.4%	31.4%	7.5%
	5	SA	2.2	2.5	3.2	3.7	5.2	5.6	7.7	9.6
		KNN	1.9	2.1	2.6	2.9	5.0	5.2	6.7	7.6
		Percent of improvement	14.0%	13.8%	18.2%	19.5%	2.7%	7.3%	13.1%	20.4%

Figure 5.4 illustrates the difference of the prediction errors (mean error and 90^{th} percentile error) when the historical data are selected from different number of days (i.e. single day vs. 5 days).

Figure 5.5 illustrates the difference of the prediction errors (mean error and 90^{th} percentile error) when the predictions are made on the basis of different models (i.e. Naïve 1 model vs. Naïve 2 model).

[5] Percent of improvement = $(\mathrm{SA} - \mathrm{KNN})/\mathrm{SA} \times 100\%$

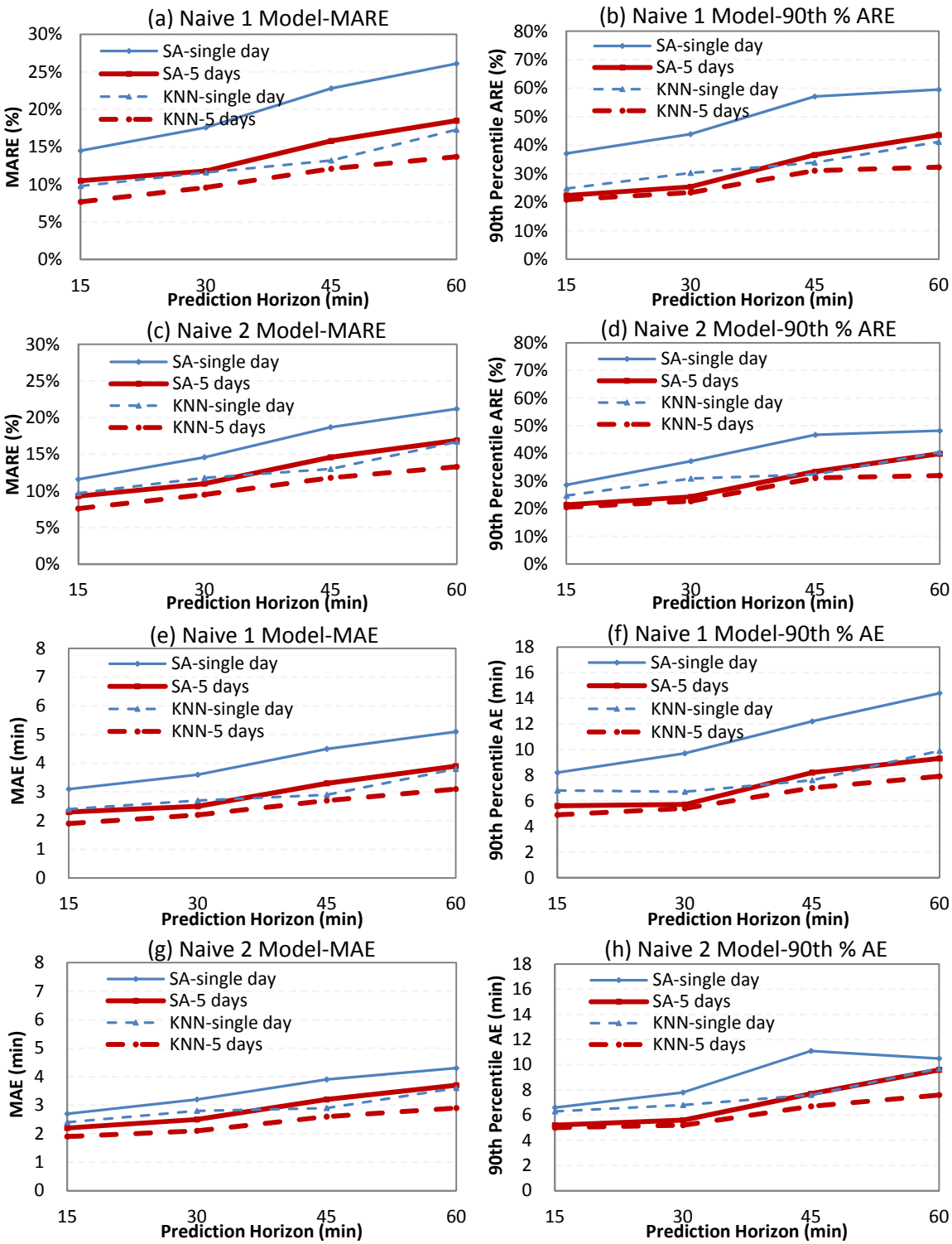


Figure 5.4: Difference of the prediction errors when historical data are selected from different number of days

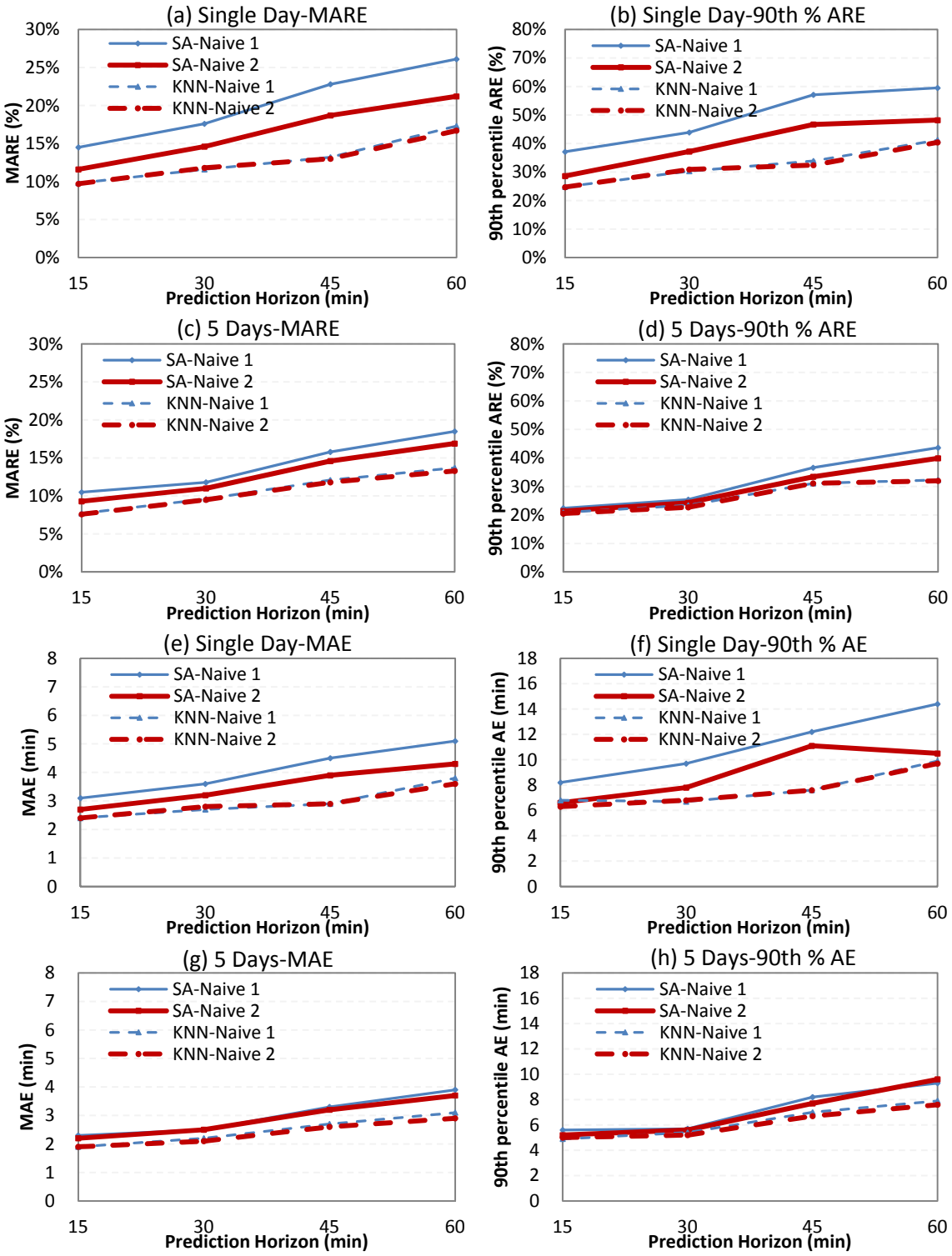


Figure 5.5: Difference of the prediction errors when the predictions are made on the basis of different prediction models

The following conclusions can be made on the basis of the results shown in Table 5.1, Figure 5.4 and Figure 5.5.

1. Historical data selected by KNN provide better prediction accuracy than historical data selected by SA, no matter what form of the prediction model (i.e. Naïve 1 or Naïve 2) is used.
2. The use of data from multiple historical days (5 days may not be an optimal number) provides better travel time prediction accuracy than data from a single historical day.
3. Naïve 2 model (i.e. trend represented by ratio) is superior to Naïve 1 model (i.e. trend represented by difference) when only a single day historical data are selected and simple average method is used. However, this superior performance becomes less obvious when multiple days are selected, especially when the historical data are selected using the KNN method rather than the simple average method.

The paired sample z-test is used to test whether or not the improvements in the MARE and MAE provided by the KNN method as compared to the simple average method are statistically significant, and whether or not the prediction results of Naïve 1 model are statistically different from the prediction results of Naïve 2 model when the KNN method is used to select 5 days historical data.

***Notation*:**

d_i	:	paired ARE/AE difference	$d_i = ARE_{SA,i} - ARE_{KNN,i}$ *or* $AE_{SA,i} - AE_{KNN,i}$ *Or* $d_i = ARE_{N1,i} - ARE_{N2,i}$ *or* $AE_{N1,i} - AE_{N2,i}$
$\bar{d}$	:	mean of the sample paired differences	$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$
n	:	sample size	
σ_d	:	standard deviation of the population paired differences	typically σ_d=S_d in practice
S_d	:	standard deviation of the sample paired differences	$S_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2}$

The mean of the paired ARE or AE differences ($\bar{d}$) and the confidence interval (CI) on a selected significant level $\alpha = 0.05$ (Equation 5.12) are computed for each prediction horizon. The statistical significance test for the improvement provided by the KNN method is shown in Table 5.2 and the Statistical significance test between the prediction results of Naïve 1 model and the prediction results of Naïve 2 model is shown in Table 5.3.

$$CI = \left(\bar{d} - z_\alpha \frac{\sigma_d}{\sqrt{n}} \quad , \; \bar{d} - z_\alpha \frac{\sigma_d}{\sqrt{n}}\right) \tag{5.12}$$

Table 5.2: Statistical significance test for the improvement provided by the KNN method

Prediction methods	Statistic values	Mean absolute relative error				Mean absolute error			
		Prediction horizon (minutes)				Prediction horizon (minutes)			
		15	30	45	60	15	30	45	60
Naïve 1	$\bar{d}$	2.8%	2.2%	3.8%	4.9%	0.4	0.4	0.7	0.8
	n	255	255	255	255	255	255	255	255
	S_d	9.5%	9.4%	13.6%	15.9%	1.7	1.9	2.8	3.1
	CI_{lower}	1.6%	1.1%	2.1%	2.9%	0.2	0.2	0.3	0.4
	CI_{upper}	4.0%	3.4%	5.4%	6.8%	0.6	0.6	1.0	1.2
Naïve 2	$\bar{\bar{d}}$	1.7%	1.5%	2.8%	3.6%	0.3	0.3	0.6	0.7
	n	255	255	255	255	255	255	255	255
	S_d	7.2%	8.8%	12.9%	15.1%	1.6	2.2	3.3	3.6
	CI_{lower}	0.9%	0.4%	1.2%	1.7%	0.1	0.1	0.2	0.3
	CI_{upper}	2.6%	2.5%	4.4%	5.4%	0.5	0.6	1.0	1.2

For the results shown in Table 5.2, if the lower bound of the confidence interval is greater than zero then the KNN method is statistically superior to the simple average method (SA). The tests are performed only for the predictions based on the historical dataset consisting of data from 5 days, because using 5 days rather than just a single day provides better travel time prediction accuracy.

As results shown in Table 5.2, at the 95% confidence level, the KNN method is superior to the SA method for all the four prediction horizons (i.e. 15, 30, 45 and 60 munities), and the superiority of the KNN method is more obvious for longer prediction horizons.

Table 5.3: Statistical significance test between the prediction results of Naïve 1 model and the prediction results of Naïve 2 model (KNN method, 5 days)

Statistic values	Mean absolute relative error				Mean absolute error			
	Prediction horizon (minutes)				Prediction horizon (minutes)			
	15	30	45	60	60	30	45	60
$\bar{d}$	0.07%	0.10%	0.26%	0.36%	0.02	0.03	0.08	0.11
n	255	255	255	255	255	255	255	255
S_d	0.58%	0.87%	1.79%	2.17%	0.19	0.34	0.79	0.90
CI_{lower}	0.00%	0.00%	0.04%	0.09%	-0.01	-0.01	-0.02	0.00
CI_{upper}	0.15%	0.21%	0.48%	0.62%	0.04	0.07	0.17	0.22

For the results shown in Table 5.3, if the lower bound of the confidence interval is greater than zero then the Naïve 2 model is statistically superior to the Naïve 1 model. The tests are performed only for the predictions based on the historical data which are selected using KNN method, and the historical

dataset consisting of data from 5 days, because it has been proved that the KNN method is significantly superior to the SA method, and using 5 days rather than just a single day provides better travel time prediction accuracy.

As shown by the results in Table 5.3, at the 95% confidence level, the difference of the prediction accuracy between the Naïve 1 model and the Naïve 2 model is only statistically significant when the prediction horizon is longer than 30 minutes, and the prediction accuracy is quantified using MARE. This result indicates that the Naïve 2 model is not statistically superior to the Naïve 1 model when the historical data are selected using KNN method and the historical dataset consists of data from 5 days, especially for the case that prediction horizon is equal to or shorter than 30 minutes.

Based on the above results, we come to the conclusion that selecting historical data using KNN provides more accurate prediction results as compared to using SA. However, three parameters must be determined before KNN method is applied, namely (1) size of the historical database; (2) length of the time window; and (3) number of days of data selected for aggregation (i.e. value of K). For example, in the above tests, the size of historical database is one year data (i.e. 260 weekdays), the length of the time window is 3 hours (i.e. time series from 2:00 pm to 5:00 pm), and the value of K is 5. The accuracy of the prediction results is a function of these three parameters, and the computational efficiency of the KNN method relates to these parameters as well. Therefore, off-line calibration of the parameters associated with KNN method for maximizing the prediction accuracy and computational efficiency is performed and results are discussed in the following section.

5.3 Calibration of Parameters Associated with KNN Method

Data used to do the following calibration are the same data described at Chapter 4. A set of data from the eastbound direction that consisting of data from 10 days (shown in Appendix A) were used as "current" days, and the rest of the data are considered as historical data in the following calibration.

Travel times are aggregated in 5 minutes interval. Naïve 1 model (as described in the previous section) is used to predict travel time in future 5 and 15 minutes. Mean absolute relative error (MARE), 90th percentile of the absolute relative error, and standard deviation of the travel time prediction errors are calculated for each test scenario. These measures of performance are calculated in two different ways:

- Based on all intervals.

- Based only on the intervals for which traffic is in a congested state ("congested" defined as the case when the average travel speed is equal to or less than 80 km/h)

A list of the tested values associated with each parameter is shown in Table 5.4. A total of 160 scenarios can be derived based on the combination of the parameters. For each scenario, prediction in future 5 minutes and 15 minutes are performed respectively.

Table 5.4: List of the tested parameter values

Parameters	Tested values
Size of the database	30 days, 60 days, 90 days, and 210 days
Length of time window	10 min, 30 min, 60 min, and 90 min
Value of K	1,2,….,10

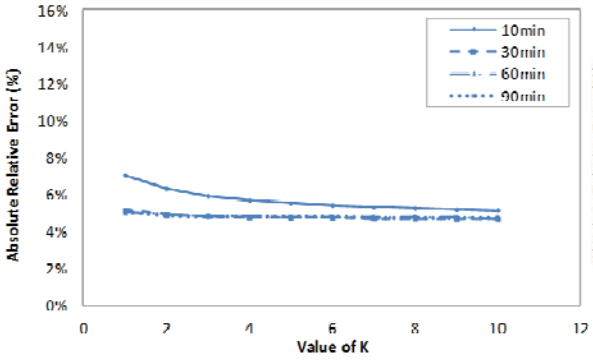
Figure 5.6 and 5.7 show the prediction accuracy results for 5 and 15 minute prediction horizons based on historical database of 210 days, with different lengths of time window and different values of K.

From Figure 5.6, we can observe that the prediction results of using 10 minutes time window are notably worse than the prediction results of using a longer length of time window (e.g. 30 min, 60 min or 90 min). When using a time window of 10 minutes, the distance calculation (Equation 5.1) is based on only two points and therefore isn't able to represent the trend of the data series. The results of using 30, 60 and 90 minutes time window are not notably different from each other for 5 minutes prediction horizon.

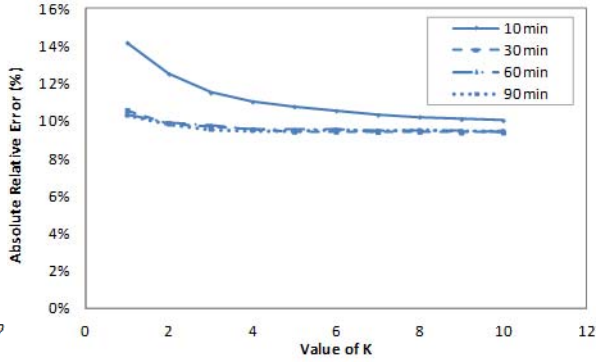
Figure 5.7 doesn't show results of using 10 minutes time window, because using a time series that is shorter than the prediction horizon (15 minutes) is not practical. The prediction results shown in Figure 5.7 from different lengths of time window are not notably different except when K=1. However, the prediction results of K=1 is not practical, as the KNN method aims to find a sub-set of the entire set of historical data in which the K>1, and the K is an optimal number that minimize the prediction error.

Based on the above results, we recommend using a 60 minute time window and K>1. The results show that as K increases, the results are improved, but these improvements increase at a decrease rate. The optimal value of K should be determined considering the size of database as well.

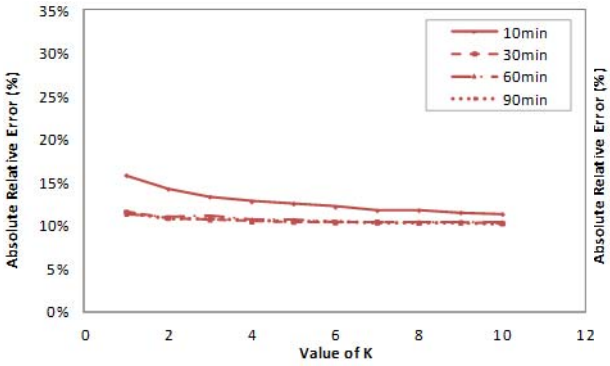
(a) MARE for all intervals



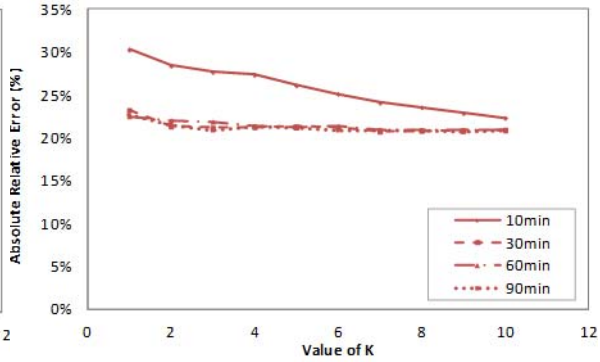
(b) MARE for congested intervals



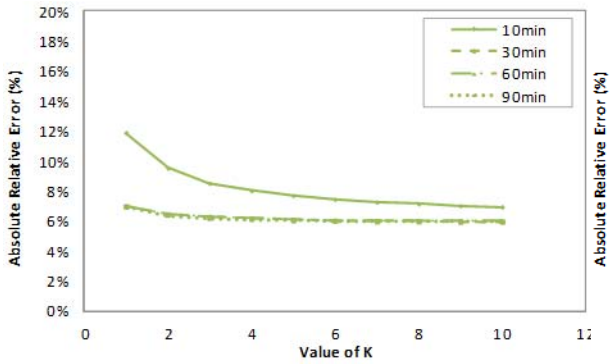
(c) 90th percentile error for all intervals



(d) 90th percentile error for congested intervals



(e) Standard deviation of prediction error for all intervals



(f) Standard deviation of prediction error for congested intervals

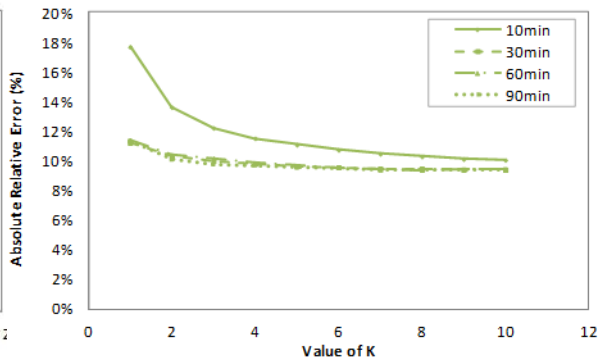


Figure 5.6: Travel time prediction accuracy as a function of K and length of time window (5-min prediction)

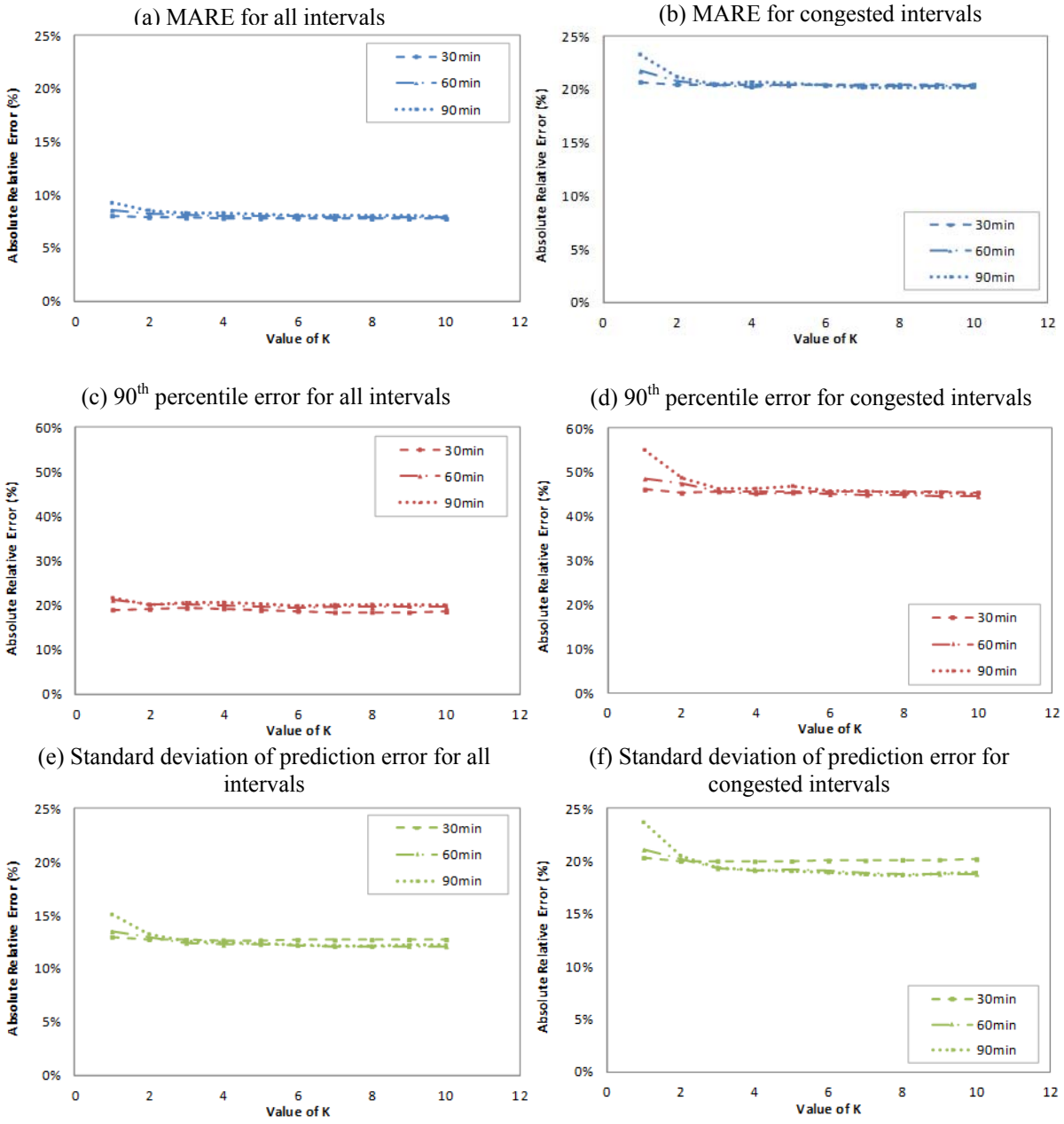


Figure 5.7: Travel time prediction accuracy as a function of K and length of time window (15-min prediction)

Figure 5.8 and 5.9 show the results of prediction in future 5 minutes and 15 minutes based on 60 minutes time window, with different sizes of database and different values of K. It can be observed from the results that the size of the historical database does not appear to have a strong influence on

prediction accuracy. That is to say, using a database containing historical data from previous 30 days (K>1) is able to provide similar prediction accuracy to that using a larger size of database (i.e. containing historical data from 90-210 days). Similar to the previous results, with increase of K the prediction results are improved but the improvements increase at a decrease rate. Therefore, we recommend using a size of database ≥ 1 month and K=4 in practical applications.

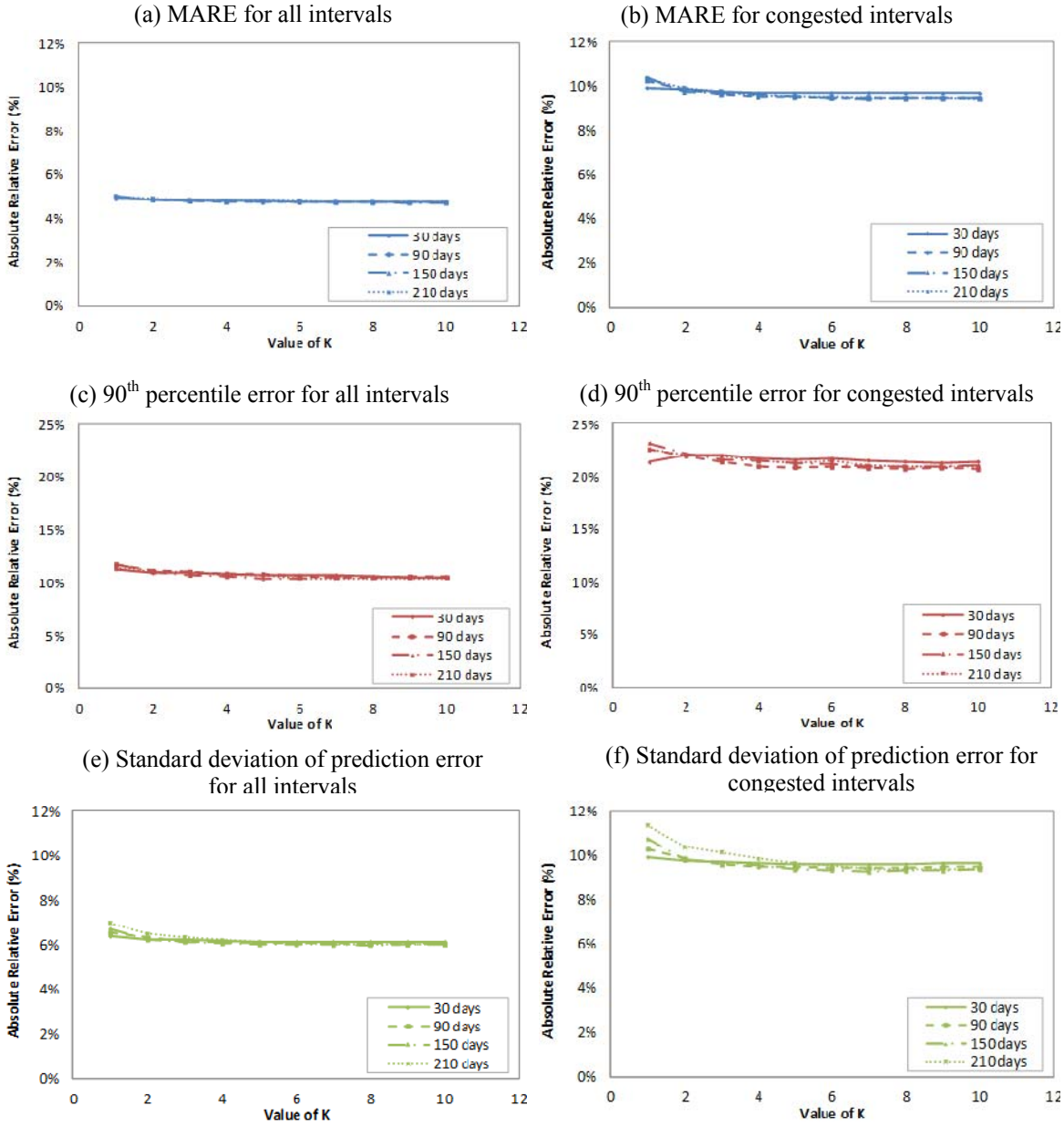
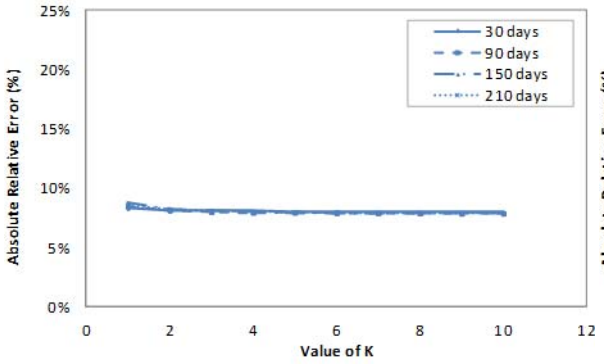
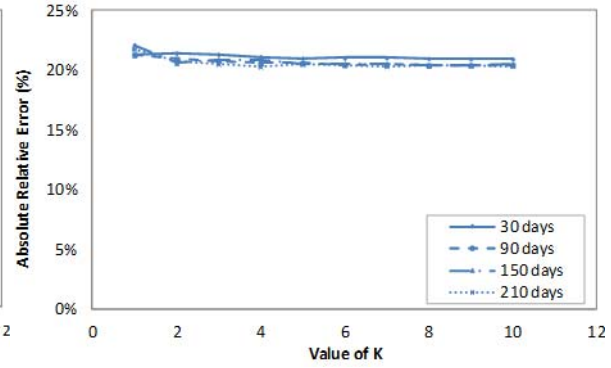


Figure 5.8: Travel time prediction accuracy as a function of K and size of the database (5-min prediction)

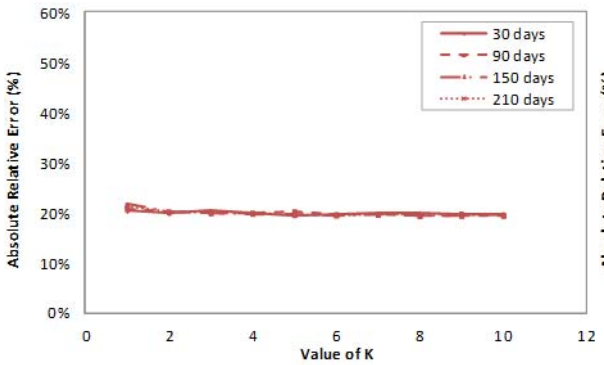
(a) MARE for all intervals



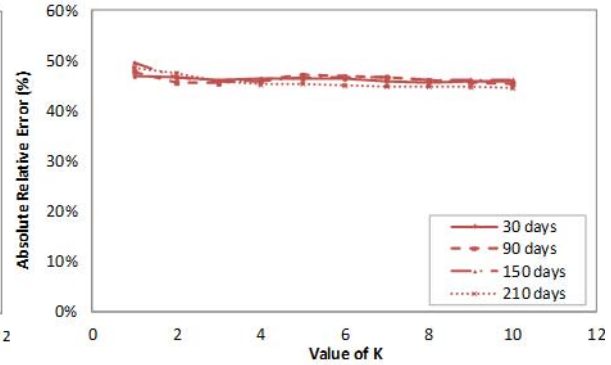
(b) MARE for congested intervals



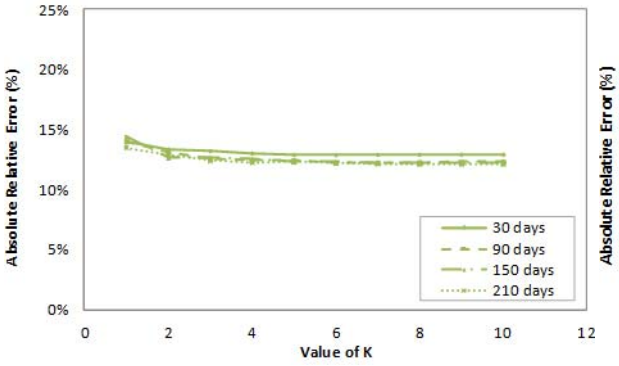
(c) 90th percentile error for all intervals



(d) 90th percentile error for congested intervals



(e) Standard deviation of prediction error for all intervals



(f) Standard deviation of prediction error for congested intervals

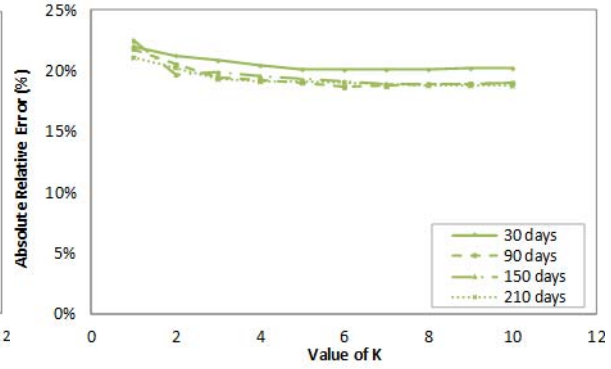


Figure 5.9: Travel time prediction accuracy as a function of K and size of the database (15-min prediction)

5.4 Summary

In this chapter, two methods (SA and KNN) of selecting historical data for short-term travel time prediction are compared, and the results show that the historical data selected by KNN provide significantly better estimation of travel time pattern (trend) than historical data selected by SA. Consequently, it is decided to use KNN method to select historical data for travel time prediction on the basis of the proposed prediction framework (discussed in Chapter 6).

Parameters associated with KNN method (e.g. the size of the historical database $\geq$ 1 month, length of the time window = 60 minutes, and number of data selected for aggregation i.e. K = 4) are calibrated using field data collected by Bluetooth detectors, and these values of the parameters are used to test the performance of the proposed travel time prediction method in Chapter 6.

Chapter 6
Short-term Travel Time Prediction - Kalman Filter Based Prediction Model

A considerable amount of work has been done on travel time prediction in the past few decades. The most common prediction methods are described in Section 2.2. These methods have their respective advantages and accuracy under different traffic conditions. Kalman filter-based model is attractive in short-term travel time prediction because the real-time estimations are updated continuously whenever new measurements are available, which enables the predictor to quickly respond to traffic fluctuations.

Most of the studies described in Section 2.2 have been developed to use data collected from conventional detectors (i.e. Loop detectors), but the focus of this study is on data collected from Bluetooth detectors. Although the core of the proposed prediction can be applied to any data source, the details of the prediction methods would likely change for different types of data.

The data collected from Bluetooth detectors are similar to data collected from Automatic Vehicle Identification (AVI) systems using dedicated transponders (e.g. such as electronic toll tags), and therefore using these data for travel time prediction faces some of the same challenges as using AVI measurements, namely: (1) dynamic outlier detection and travel time estimation must be able to respond quickly to rapid travel time changes; and (2) a time lag exists between the time when vehicles enter the segment and the time that their travel time can be measured (i.e. when the vehicle exits the monitored segment).

This chapter describes a proposed model for predicting near future freeway travel times using Bluetooth data with special attention to the above two challenges. The model combines a dynamic outlier filtering algorithm (described in chapter 4) with Kalman filtering and uses historical data to make up for the limitation of real-time Bluetooth measurements.

6.1 An Alternative Approach of Estimating Mean Travel Time

In many previous studies (Haghani et al. 2010; Dion and Rakh 2006; SwRI 1998; Mouskos et al. 1998), the mean travel time is computed on the basis of valid observations using Equation 6.1, in which the time instant that a vehicle is observed at downstream detector (i.e. end time) is used as the time label for carrying out the travel time aggregation.

$$ATT_{i,k} = \frac{\sum_{j=1}^{n_k}\left(t_{i,j} - t_{i-1,j}\right) \cdot \tau_{j,k}}{\sum_{j=1}^{n_k} \tau_{j,k}} \quad t_{i,j} \in k \tag{6.1}$$

Where: $t_{i,j}$ and $t_{i-1,j}$ are the time instants when vehicle j was detected at the downstream Bluetooth detectors i and the upstream detector $i-1$ respectively; $\tau_{j,k} = 0$ if the measured travel time $t_{i,j} - t_{i-1,j}$ is identified as an outlier, otherwise $\tau_{j,k} = 1$; n_k is the total number of measured travel times in period k.

An alternative approach is defined in Equation 6.2. The only difference between Equation 6.1 and Equation 6.2 is that in Equation 6.1 the time instant of a vehicle observed at downstream detector ($t_{i,j}$) is attributed to interval k however in Equation 6.2 the time instant of a vehicle observed at upstream detector ($t_{i-1,j}$) is attributed to interval k.

$$DTT_{i,k} = \frac{\sum_{j=1}^{n_k}\left(t_{i,j} - t_{i-1,j}\right) \cdot \tau_{j,k}}{\sum_{j=1}^{n_k} \tau_{j,k}} \quad t_{i-1,j} \in k \tag{6.2}$$

Equation 6.1 (ATT, arrival travel time) and Equation 6.2 (DTT, departure travel time) provide different estimates of the mean travel time. For real-time applications, such as posting travel time on variable message signs (VMS), the travel time of interest is the travel time that vehicles entering the segment during the given time interval will experience. Thus for these applications, DTT should be considered as the true travel time.

As discussed in Section 3.1.2, ATT estimated using Equation 6.1 lags behind the true travel time (DTT) which is the time that vehicles will experience before/when they enter the roadway section. The results from Section 3.1.2 have indicated that the errors caused by using ATT directly as an estimate of DTT will be unacceptably large in practical applications when traffic conditions vary.

In practice, it is easier to estimate ATT than DTT. However, many previous studies have ignored the distinction between ATT and DTT and in some cases have specifically considered ATT as "true" travel times when the prediction results are evaluated (Barcelo et al. 2010). This error is likely to result in reported performance that is artificially inflated and better than would actually be achieved.

It is possible to use Equation 6.2 to compute DTT in real-time, however if the time when prediction is made is the end time of interval k (t_k), then the measured travel time ($t_{i,j} - t_{i-1,j}$) is only available for those vehicles which entered the section during a time period satisfying Equation 6.3.

$$t_k - \Delta k \le t_{(i-1),j} \le t_k - \left(t_{i,j} - t_{(i-1),j}\right) \tag{6.3}$$

When the freeway section between detectors i and $i-1$ is relatively long, the section is congested or Δk is relatively short, then no vehicles satisfy Equation 6.3 and no travel times are available at time t_k to compute $DTT_{i,k}$.

In summary, the prediction of mean travel time on the basis of AVI travel time data requires: (1) a reliable real-time outlier detection algorithm; (2) a method for addressing data gaps caused by the time lag inherent in the travel time measurements. The next section describes a proposed approach for addressing these two needs.

6.2 Proposed Travel Time Prediction Method

The proposed model is based on Kalman filter theory and therefore involves both prediction and estimation. The proposed model consists of 4 steps: (1) Prior estimation; (2) Outlier detection; (3) Posterior estimation; (4) Traffic pattern recognition. These steps are described in the following sections.

6.2.1 Step 1: Prior Estimation (Prediction)

The variation in travel time over time is a function of the traffic state (i.e. uncongested or congested) and therefore different state functions are defined for these two traffic states. To meet the requirement of Kalman filter, the system is modeled with linear relationships between two consecutive states.

$$\hat{\bar{t}}t_{i,k} = \begin{cases} \hat{t}t_{i,k-1} + \omega_{i,k-1} & \hat{t}t_{i,k-1} < \theta_{tt} \\ \hat{t}t_{i,k-1} + \alpha \cdot \Delta\hat{t}t_{i,k-1} + (1-\alpha) \cdot \Delta\bar{t}t_{i,k} + \omega_{i,k-1} & \hat{t}t_{i,k-1} \ge \theta_{tt} \end{cases} \tag{6.4}$$

Where,

$\hat{\bar{t}}t_{i,k}$: The mean travel time between detectors i and $i-1$ at time interval k that is to be predicted

$\hat{t}t_{i,k-1}$: The mean travel time between detectors i and $i-1$ at time interval $k-1$ that is estimated

$\Delta\hat{t}t_{i,k-1}$: The real-time change in mean travel time (trend) between detectors i and $i-1$, $\Delta\hat{t}t_{k-1} = (\hat{t}t_{k-1} - \hat{t}t_{k-2})$

$\Delta\bar{t}t_{i,k}$: Change in mean travel time (trend) between detectors i and $i-1$ in a historical dataset with similar traffic pattern to current traffic state; this historical dataset is

selected using the nearest neighbor (KNN) method.

θ_{tt} : A threshold of travel time used to distinguish congestion traffic state from free flow traffic state

α : A weight factor to determine the level of confidence that should be placed on the real-time trend $\Delta\hat{t}t_{i,k-1}$ and the historical trend $\Delta\bar{t}t_{i,k}$

$\omega_{i,k-1}$: Noise term that has a normal distribution with zero mean and a variance of Q_{k-1}. A covariance matching method (Myers and Tapley 1976) is used to adaptively estimate the unknown noise for state process.

The predicted mean travel time at interval k is a prior estimation based on the state function (i.e. Equation 6.4) without considering the noise term. When traffic is operating in a free flow state (i.e. $\hat{t}t_{i,k-1} < \theta_{tt}$), then the prior estimation (i.e. predicted travel time) at interval k is equal to the estimated travel time from the previous interval. When traffic is congested (i.e. $\hat{t}t_{i,k-1} \geq \theta_{tt}$), the prior estimation at interval k is a linear combination of the estimated travel time from the previous interval and the historical and real-time travel time trend. The trend of travel time is represented by the difference rather than the ratio of the travel times between two consecutive intervals, because the results obtained at Chapter 5 indicated that using the ratio of travel times instead of the difference of travel times does not significantly improve the prediction results when data from multiple days are aggregated, and the trend represented by difference of travel times is easier to implement in the Kalman filter model. θ_{tt} is a user-defined threshold parameter that can be determined based on previous experience and/or engineering judgment and is specific to a given freeway segment. α is a dynamically adjusted weight factor that is determined using Equation 6.5, which is a modified equation based on Equation 2.13 developed by Dion and Rakha.

$$\alpha = 0.5 * (1 - (1 - \beta)^{n_{v,k-1}}) \quad (6.5)$$

The value of α depends on the number of valid travel time observations in the previous interval ($n_{v,k-1}$), β is a sensitivity parameter that determines how quickly α responds to the number of valid observations. Figure 6.1 shows the variation of α over the number of valid observations and the value of parameter β.

Equation 2.13 in Dion and Rakha's model is used to determine the level of confidence that should be placed on the observed data in the previous interval and the smoothed estimation of the previous interval when the expected smoothed average travel time of the current interval is estimated. However,

the weight factor α in Equation 6.5 is used to determine the level of confidence that should be placed on current trend of travel time and historical trend of travel time when the travel time is predicted in the proposed model. In order to use the historical data to make up the limitation of real-time data, A limit value of α (i.e. 0.5) is set to make sure that at least 50% of the weight will be placed on the historical aggregated data. Sensitivity analysis of this factor is conducted when the method is validated.

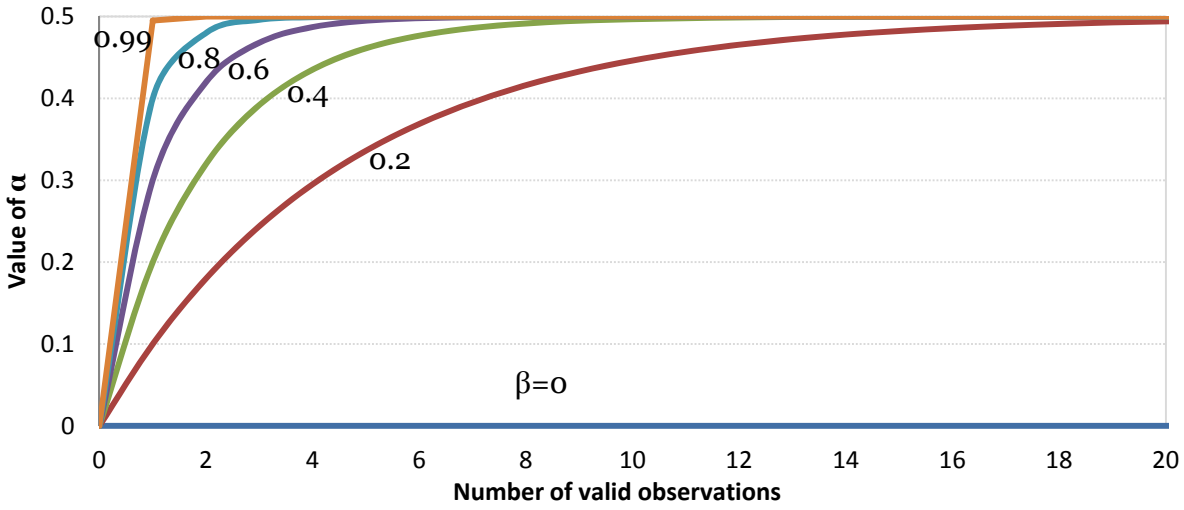


Figure 6.1: Variation of α over the number of valid observations and parameter β

The state function (Equation 6.4) can be transformed into the general form of the Kalman filter as follows. Define $x_k^1 = tt_k$, $x_k^2 = tt_k - tt_{k-1}$ and $u_k = \Delta\overline{tt}_k$, in which the notation i that represents a specific road section is omitted here and in the following description, then x_k^1 and x_k^2 can be expressed as functions of x_{k-1}^1, x_{k-1}^2,u_k and noise term ω_{k-1}:

$$x_k^1 = x_{k-1}^1 + \alpha \cdot x_{k-1}^2 + (1-\alpha) \cdot u_k + \omega_{k-1} \quad (6.6)$$

$$x_k^2 = x_k^1 - x_{k-1}^1 = \alpha \cdot x_{k-1}^2 + (1-\alpha) \cdot u_k + \omega_{k-1} \quad (6.7)$$

The state vector x_k can be expressed by Equation 6.8:

$$x_k = \begin{bmatrix} x_k^1 \\ x_k^2 \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ 0 & \alpha \end{bmatrix} \cdot \begin{bmatrix} x_{k-1}^1 \\ x_{k-1}^2 \end{bmatrix} + \begin{bmatrix} 1-\alpha \\ 1-\alpha \end{bmatrix} \cdot u_k + \omega_{k-1} \quad (6.8)$$

Where: $\alpha = 0$ and $u_k = 0$ if $x_{k-1}^1 < \theta_{tt}$; otherwise α is determined using Equation 6.2, and u_k is determined using historical data selected by nearest neighbor method.

6.2.2 Step 2: Outlier Detection and Travel Time Measurement

The set of valid travel times ($S_{tt_{i,k}}$) that are observed between two detectors i and $i-1$ during time interval k is defined in Equation 6.9, in which the valid observations during interval k are the travel times experienced by vehicles that pass through the upstream boundary of the road section (i.e. detector $i-1$) during interval k. Equation 6.10 defines the method used to calculate the average travel time during time interval k based on the valid observations identified by Equation 6.9.

$$S_{tt_{i,k}} = \{t_{i,j} - t_{i-1,j} | t_k - t_{k-1} < t_{i-1,j} \leq t_k \; and \; tt_{i,lower,k} \leq t_{i,j} - t_{i-1,j} \leq tt_{i,upper,k}\} \quad (6.9)$$

$$\tilde{tt}_{i,k} = \begin{cases} \hat{tt}_{i,k}^{-} & if \; n_{v,k} < \theta_{n_v} \\ \dfrac{\sum_{j=1}^{n_{v,k}} (t_{i,j} - t_{i-1,j})}{n_{v,k}} & if \; n_{v,k} \geq \theta_{n_v} \end{cases} \quad (6.10)$$

In the above equations, $t_{i,j}$ and $t_{i-1,j}$ are the time at which vehicle j was detected at AVI detectors i and $i-1$ respectively; t_k is the end time of time interval k; $tt_{i,lower,k}$ and $tt_{i,upper,k}$ are the lower bound and upper bound of the validity window at time interval k respectively; $\tilde{tt}_{i,k}$ is the measured average travel time during interval k; $n_{v,k}$ is the number of valid observations identified at interval k; θ_{n_v} is a threshold to determine the minimum number of valid observations required to update the priori estimations using measurements. The threshold θ_{n_v} is set equal to 1% of the total traffic volume.

An initial validity window is determined using Equations (6.11) and (6.12) following the method developed by Dion and Rakha (*3*):

$$tt_{i,lower,k} = e^{[ln(\hat{tt}_{i,k}^{-}) - n_{std} \cdot \sigma_{i,k-1}]} \quad (6.11)$$

$$tt_{i,upper,k} = e^{[ln(\hat{tt}_{i,k}^{-}) + n_{std} \cdot \sigma_{i,k-1}]} \quad (6.12)$$

Where, $\hat{tt}_{i,k}^{-}$ is the predicted travel time at time interval k; $\sigma_{i,k-1}$ is the sample standard deviation calculated on the basis of all valid observations in the previous interval $k-1$; n_{std} is a parameter representing the number of standard deviations. The sample variance in a specific interval k is calculated using Equation 6.13.

$$\sigma_{i,k}{}^2 = \begin{cases} \sigma_{i,k-1}{}^2 & if\ n_{v,k} < \theta_{n_v}\ or\ n_{v,k} \le 1 \\ \dfrac{\sum_{j=1}^{n_{v,k}} \left[ln\left(t_{i,j} - t_{i-1,j}\right)_k - ln\left(\hat{tt}_{i,k}{}^{-}\right) \right]^2}{n_{v,k} - 1} & if\ n_{v,k} \ge \theta_{n_v}\ and\ n_{v,k} > 1 \end{cases} \quad (6.13)$$

If the number of valid observations is less than the threshold θ_{n_v} or not greater than 1, the sample variance is equal to the sample variance computed for the previous interval, otherwise the sample variance is calculated using the variance equation. Following the method used by Dion and Rakha (2006), the variance is computed using the predicted average travel time ($\hat{tt}_{i,k}{}^{-}$) rather than the current average travel time ($\tilde{tt}_{i,k}$).

The initial validity window is determined using a method similar to that used by Dion and Rakha (2006), in which the travel times of individual vehicles are assumed to follow a lognormal distribution and the validity window is estimated using a number of standard deviations above and below the predicted mean travel time. The parameter n_{std} is set to 3, implying that 99.7% of the data lie within the validity window.

The initial validity window is only effective when the traffic state is relative stable, so an extension of the initial validity window based on traffic flow theory was developed in order to increase the responsiveness of the filtering algorithm to rapid changes in the underlying travel time. Details about the extension can be found in Chapter 4.

6.2.3 Step 3: Posterior Estimation (Correction)

The measurement function is defined with Equation 6.14 to relate the state to the measurement.

$$\tilde{tt}_{i,k} = \hat{tt}_{i,k}{}^{-} + \delta_{i\,k} \quad (6.14)$$

Where,

$\tilde{tt}_{i,k}$: The mean travel time at time interval k

$\delta_{i,k}$: Measurement error that has a normal distribution with zero mean and a variance of R_k

The measurement function (Equation 6.14) can be transformed into the general form of Kalman filter (notation i is omitted in the following description) by defining $z_k = \begin{bmatrix} \tilde{tt}_k \\ \tilde{tt}_k - \tilde{tt}_{k-1} \end{bmatrix}$, and then the measurement vector z_k can be expressed by Equation 6.15:

$$z_k = \begin{bmatrix} x_k^1 \\ x_k^2 \end{bmatrix} + \delta_k \quad (6.15)$$

The measurement in this study is a mean travel time calculated on the basis of the valid observations within each interval. This is a sample mean that is used to estimate the population mean, and therefore according to the central limit theorem, the measurement can be assumed to be approximately normally distributed. Figure 6.2 shows the mean travel time distribution based on the field travel time observations collected by Bluetooth detectors (the same dataset used in Section 4.1). Mean travel times (5-min interval) from over 200 days are investigated. From the results shown in Figure 6.2, we can see that the mean travel time can be assumed approximately normally distributed.

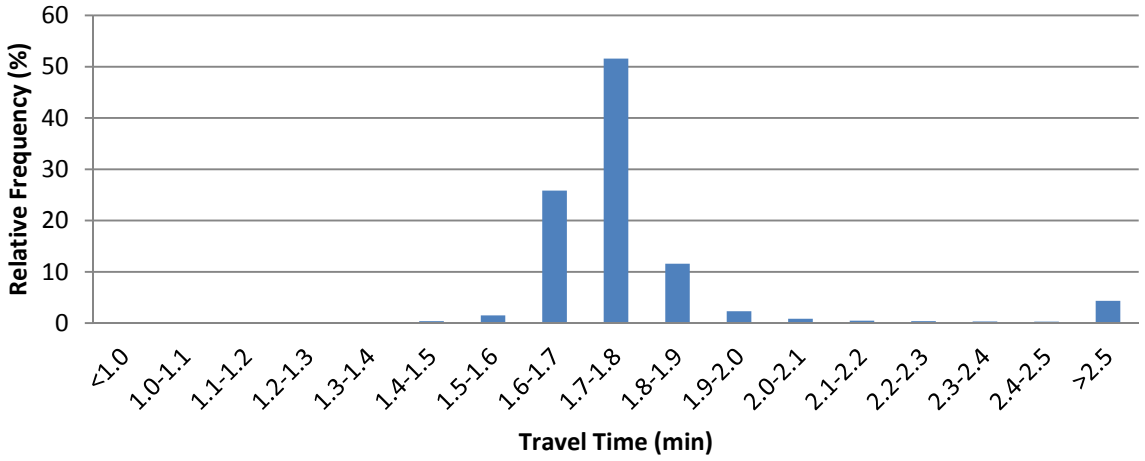


Figure 6.2: Distribution of the mean travel time based on field travel time observations

The difference between the sample mean and the population mean is considered as the measurement noise in this study, and this measurement noise is assumed to be normally distributed with zero mean and a variance of R_k. The variance R_k is quantified based on the sample mean variance using Equation 6.16.

$$R_k = \begin{bmatrix} {\sigma_{\hat{t}\hat{t}_{i,k}}}^2 & \sigma_{\hat{t}\hat{t}_{i,k}}\left(\sigma_{\hat{t}\hat{t}_{i,k}} + \sigma_{\hat{t}\hat{t}_{i,k-1}}\right) \\ \sigma_{\hat{t}\hat{t}_{i,k}}\left(\sigma_{\hat{t}\hat{t}_{i,k}} + \sigma_{\hat{t}\hat{t}_{i,k-1}}\right) & \left(\sigma_{\hat{t}\hat{t}_{i,k}} + \sigma_{\hat{t}\hat{t}_{i,k-1}}\right)^2 \end{bmatrix} \tag{6.16}$$

Where, $\sigma_{\hat{t}\hat{t}_{i,k}}$ is the sample mean variance of valid travel time observed at time interval k, ${\sigma_{\hat{t}\hat{t}_{i,k}}}^2 = {\sigma_{i,k}}^2/n_{v,k}$.

Once new measurements are available, previous estimations are updated based on the Kalman gain K_k:

$$\hat{x}_k = \hat{x}_k^{-} + K_k(z_k - \hat{x}_k^{-}) \tag{6.17}$$

The posterior estimation of travel time can be directly used as travel time ground truth for off-line analysis (e.g. evaluation of the prediction results), because individual travel times are aggregated/estimated as average departure travel time (DTT) which is the "true" travel time corresponding to the predicted travel time.

6.2.4 Step 4: Traffic Pattern Recognition

The K nearest neighbors (KNN) technique is used in the proposed method to find a sub-set of historical travel times, which contains only those data that are most similar to the conditions observed so far on the current day from the entire set of historical data. The similarity (normally termed the "distance") between the traffic conditions of the present day and conditions in the past days is measured by computing the Euclidean distance between two time series (Equation 6.18). Where, $a = \{\hat{tt}_{k-N_k+1}, \dots, \hat{tt}_{k-1}, \hat{tt}_k\}$ is a discrete time series of travel times estimated today (notation i is omitted here and in the following description); each estimation of the travel time in the time series is obtained at a constant time interval - say every 5 minutes; N_k is the number of the data points contained in the time series; the same discrete time series exists in each of the historical days $h_d = \{(\hat{tt}_{k-N_k+1})_d, \dots, (\hat{tt}_{k-1})_d, (\hat{tt}_k)_d\}$, where $d = \{1,2, \dots, D\}$ and D is the total number of days in the historical data.

$$Dist_d(a,\ h_d) = \sqrt{\sum_{n=1}^{N_k} [\hat{tt}_{k-n+1} - (\hat{tt}_{k-n+1})_d]^2} \tag{6.18}$$

Consequently, a sub-set which contains only those data that are most similar to the conditions observed so far on the current day can be identified. Once the sub-set of the historical data is identified, a weighted average of the change in travel time (trend) in the historical data is computed using Equation 6.19.

$$\Delta\bar{tt}_k = \frac{\sum_{d=1}^{D_s} (\hat{tt}_k - \hat{tt}_{k-1})_d / Dist_d}{\sum_{d=1}^{D_s} 1/Dist_d} \tag{6.19}$$

Where, $(\hat{tt}_k - \hat{tt}_{k-1})_d$ is the trend of travel time in historical day d; $Dist_d$ represents the dissimilarity error (i.e. "distance") between current time series and the time series in historical day d; D_s is the total number of historical days selected in the sub-set.

The operation process of the proposed model is illustrated in Figure 6.2, in which we can see that the proposed 4 steps are recursively applied.

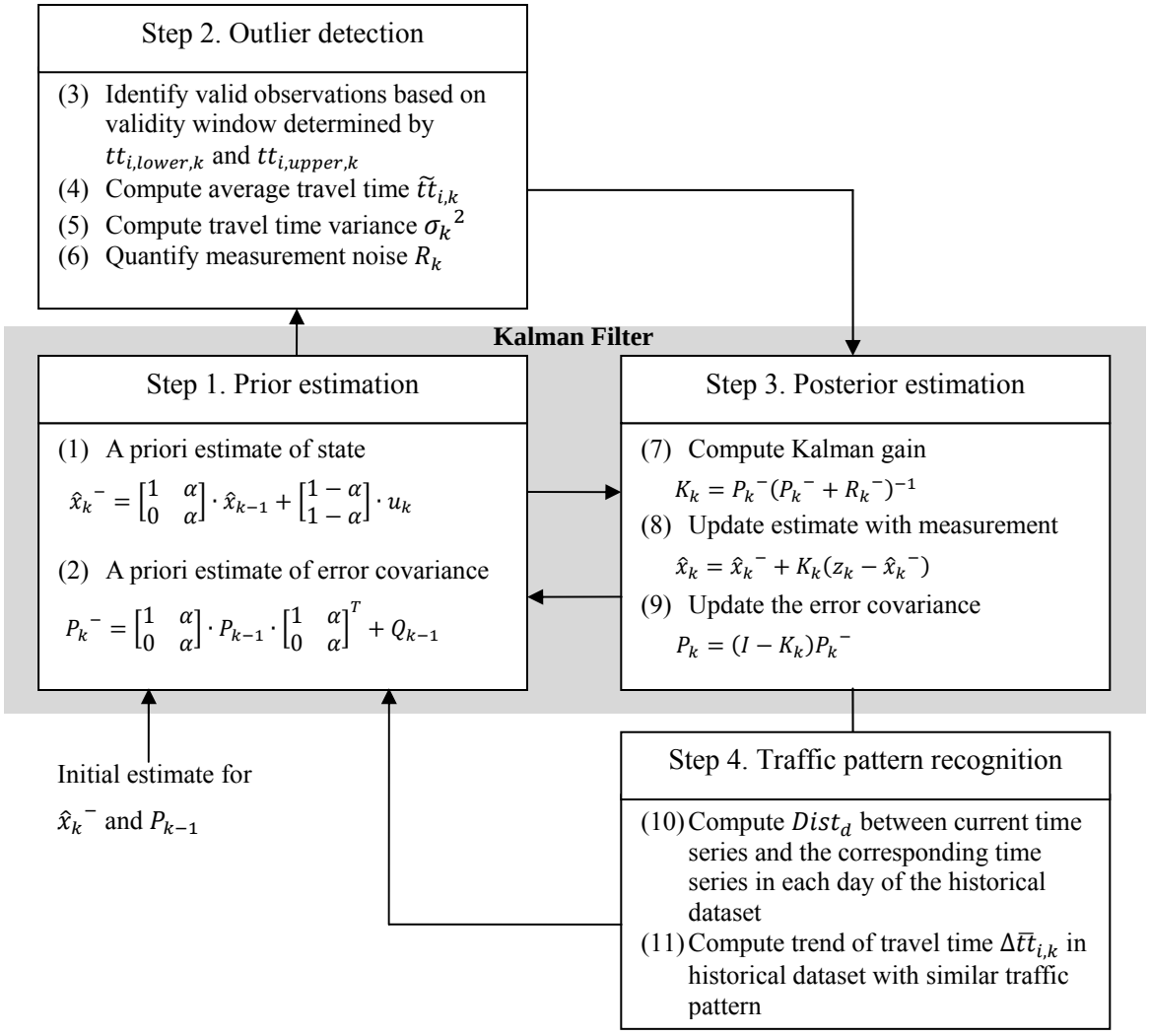


Figure 6.3: Operation process of the proposed model

The proposed model is designed to be able to predict travel time n_s steps into the future ($n_s = 1,2,\dots n$). For example, the data are aggregated in a pre-determined length of time interval (Δk). If $n_s = 1$, then the prediction horizon is Δk; if $n_s = n$, then the prediction horizon is $n\Delta k$. The prediction step (defined as a time interval at which the prediction is updated) is equal to the length of data aggregation interval (Δk), that is to say the prediction is updated every 5 minutes if the pre-determined length of time interval is 5 minutes.

Every time the prediction is updated, previous estimations have to be updated first. This is the feedback control in the proposed model. The proposed model aggregates travel times as DTT (Equation 6.10). If measured travel times are not available at the present time as a result of the lag effect, then the prediction is made based on the state function (Equation 6.4) with more weight (α) being given to historical data trend. The measurements are set equal to predictions when there are too few (less than θ_{n_v}) valid observations in specific time intervals (Equation 6.10), and consequently the posterior estimations are equal to predictions in these intervals. As time goes on, these priori estimations will be updated when new measurements are available.

6.3 Validation and Calibration

The proposed model was applied to the dataset described in Chapter 4. A set of data from the eastbound direction was used to test the proposed model. Data from 5 days were selected as "current" days, and the rest of the data are considered as historical data in the following tests. The selected 5 days (Jun. 15th 2012, Sep. 20th 2012, Oct. 5th 2012, Oct. 19th 2012 and Nov. 5th 2012) experienced varying levels of traffic congestion, but most congestion on this segment is non-recurrent. Sample of travel time observations collected from these 5 days can be found in Appendix A, in which we can see that a number of outliers must be removed, and the sample variance becomes large when traffic is congested.

Travel times are aggregated in 5 minutes intervals, and mean travel times are predicted for 5, 10, and 15 minutes into the future. A time series of travel time with length of 1 hour is used in the process of traffic pattern recognition, and the most similar 4 days are selected from the entire historical dataset. These parameters were selected through off-line calibration (Chapter 5). On the basis of engineering judgment, θ_{tt} is set equal to the travel time corresponding to an average travel speed of 80km/h. θ_{n_v} is set equal to 5 observations in a 5 minute interval, which is approximately equal to 1% of the total traffic volume.

The value of the use of the trend terms in the proposed model was determined by comparing the performance of the proposed model (Model_P) to a benchmark model (noted as Model_B) which uses the same state function for predictions in both free flow and congestion states (i.e. Equation 6.20). The sensitivity of the performance of the proposed model to parameter β was determined for the values β= {0, 0.2, 0.4, 0.6, 0.8, 0.99}.

$$\bar{\hat{tt}}_{i,k} = \hat{tt}_{i,k-1} + \omega_{i,k-1} \tag{6.20}$$

The prediction models (Model_B and Model_P) are evaluated using three measures of performance (MOPs): (1) mean absolute relative error (MARE); (2) 90th percentile of the absolute relative error (90th P ARE); and (3) standard deviation of the absolute relative error (Std. ARE). The travel time ground truth is the posterior estimation of travel time from the proposed Kalman filter-based model, because valid individual travel times are aggregated/estimated as average departure travel time (DTT), which is the "true" travel time in real-time applications. Each MOP is calculated in three different ways:

- Based only on the intervals for which traffic is in a free flow state ("free flow" defined as the case when the average travel speed is greater than 80 km/h).
- Based only on the intervals for which traffic is in a transition state ("transition" defined as the case when the average travel speed is equal to or less than 80 km/h and congestion is forming or dissipating; the period when congestion is forming or dissipating is determined through observation on the basis of the true travel times, see Appendix B).
- Based only on the intervals for which traffic is in a congested state ("congested" defined as the case when the average travel speed is less than 80 km/h and the level of congestion is remaining relatively stable; the period when traffic is in a congested state is determined through observation on the basis of the true travel times, see Appendix B).

The statistical significance of differences in the ARE between different models is tested using the paired-z test. The Model_B is compared to 6 tested models (noted as Model-P*) with different values of β. Percentage of improvement (i.e. $(\text{Model\_B} - \text{Model\_P}^*)/\text{Model\_B}$) is computed for MARE, 90th P ARE and Std. ARE and noted as ΔM, ΔP and ΔS respectively in the following descriptions. Positive values for ΔM, ΔP or ΔS imply that the tested model provides better performance than the benchmark model. The test results for 5 minutes, 10 minutes and 15 minutes predictions are shown in Table 6.1, 6.2 and 6.3 respectively, and the percentage of improvements on MARE are shaded if the difference is statistically significant at the 95% level of confidence.

From the results in Table 6.1, 6.2 and 6.3, the following observations can be made:

1. For 5 minutes prediction, there is no statistically significant difference between the prediction results of the proposed model and the prediction results of the benchmark model for all the tested traffic states (free flow, transition and congestion).
2. For 10 and 15 minutes predictions, the differences of the mean errors for predictions at free flow and congestion states between the proposed model and the benchmark model are not statistically

significant. However, the proposed model significantly improves the prediction accuracy when traffic is in a transition state for both 10 and 15 minutes prediction.

3. The proposed model improves the 90th percentile error and the standard deviation of the prediction errors at transition state as well, and the improvements are more obvious as the prediction horizon increases.
4. The prediction errors of the proposed models (for all the tested prediction horizons) at transition state decrease with the value of parameter β increases, and the improvements of the proposed model over benchmark model on MAREs for prediction in future 10 minutes are statistically significant only when $\beta > 0$.

The above observations from 1 to 3 validate that the trend terms added in the proposed model effectively improve the accuracy of prediction results when traffic is in a transition state. The 4th observation indicates that the trend term from real-time data combined with that from historical data is able to improve the accuracy of prediction, and the value of β should be greater than 0. Variation of the percentage of the improvements on MAREs (i.e. ΔM) with different values of parameter β is shown in Figure 6.3, and the results show that the improvements on MAREs (i.e. ΔM) based on different values of parameter β do not change much as long as $\beta > 0$, therefore the value of parameter β is selected to be 0.2 when the proposed model is applied in Chapter 7.

Table 6.1: Results of model validation (5-min)

5-min		Model-B	Model-P					
			β=0	β=0.2	β=0.4	β=0.6	β=0.8	β=0.99
Free Flow	MARE	0.025	0.026	0.026	0.026	0.026	0.026	0.026
	90th P ARE	0.057	0.057	0.057	0.057	0.057	0.057	0.057
	Std. ARE	0.030	0.030	0.030	0.030	0.030	0.030	0.029
	ΔM		-0.009	-0.008	-0.008	-0.008	-0.008	-0.008
	ΔP		0.001	0.001	0.001	0.001	0.001	0.001
	ΔS		-0.007	0.001	0.001	0.001	0.001	0.002
Transition	MARE	0.187	0.175	0.164	0.164	0.164	0.164	0.162
	90th P ARE	0.415	0.371	0.316	0.316	0.316	0.316	0.316
	Std. ARE	0.189	0.170	0.160	0.159	0.159	0.159	0.158
	ΔM		0.067	0.122	0.124	0.124	0.125	0.133
	ΔP		0.106	0.238	0.239	0.239	0.239	0.238
	ΔS		0.099	0.150	0.155	0.157	0.157	0.163
Congestion	MARE	0.137	0.142	0.156	0.158	0.165	0.157	0.159
	90th P ARE	0.256	0.297	0.302	0.308	0.351	0.302	0.324
	Std. ARE	0.137	0.144	0.145	0.146	0.158	0.144	0.146
	ΔM		-0.040	-0.142	-0.157	-0.203	-0.149	-0.164
	ΔP		-0.159	-0.180	-0.205	-0.372	-0.180	-0.266
	ΔS		-0.045	-0.057	-0.063	-0.152	-0.051	-0.065

Table 6.2: Results of model validation (10-min)

10-min		Model-B	Model-P					
			β=0	β=0.2	β=0.4	β=0.6	β=0.8	β=0.99
Free Flow	MARE	0.024	0.024	0.024	0.024	0.024	0.024	0.024
	90th P ARE	0.048	0.047	0.047	0.047	0.047	0.047	0.047
	Std. ARE	0.034	0.034	0.034	0.034	0.034	0.034	0.034
	ΔM		0.001	0.001	0.001	0.001	0.001	0.001
	ΔP		0.011	0.011	0.011	0.011	0.011	0.011
	ΔS		0.003	0.003	0.003	0.003	0.003	0.004
Transition	MARE	0.275	0.246	0.238	0.237	0.237	0.237	0.236
	90th P ARE	0.592	0.547	0.537	0.533	0.533	0.533	0.525
	Std. ARE	0.250	0.227	0.213	0.211	0.210	0.209	0.208
	ΔM		0.106	0.135	0.138	0.140	0.140	0.144
	ΔP		0.076	0.092	0.100	0.100	0.100	0.114
	ΔS		0.092	0.147	0.157	0.159	0.162	0.167
Congestion	MARE	0.129	0.133	0.140	0.142	0.144	0.142	0.143
	90th P ARE	0.251	0.271	0.278	0.280	0.283	0.280	0.276
	Std. ARE	0.135	0.139	0.137	0.138	0.144	0.138	0.138
	ΔM		-0.026	-0.085	-0.094	-0.117	-0.097	-0.106
	ΔP		-0.080	-0.110	-0.117	-0.130	-0.117	-0.098
	ΔS		-0.026	-0.014	-0.019	-0.067	-0.020	-0.021

Table 6.3: Results of model validation (15-min)

15-min		Model-B	Model-P					
			β=0	β=0.2	β=0.4	β=0.6	β=0.8	β=0.99
Free Flow	MARE	0.023	0.023	0.023	0.023	0.023	0.023	0.023
	90th P ARE	0.046	0.045	0.045	0.045	0.045	0.045	0.045
	Std. ARE	0.034	0.034	0.034	0.034	0.034	0.034	0.034
	ΔM		-0.001	0.002	0.002	0.002	0.002	0.002
	ΔP		0.010	0.015	0.015	0.015	0.015	0.015
	ΔS		0.000	0.001	0.001	0.001	0.001	0.002
Transition	MARE	0.368	0.326	0.309	0.308	0.307	0.307	0.306
	90th P ARE	0.719	0.621	0.603	0.603	0.603	0.603	0.605
	Std. ARE	0.290	0.271	0.266	0.263	0.261	0.260	0.259
	ΔM		0.116	0.161	0.164	0.166	0.167	0.169
	ΔP		0.136	0.161	0.162	0.162	0.162	0.159
	ΔS		0.064	0.082	0.093	0.099	0.101	0.104
Congestion	MARE	0.132	0.134	0.143	0.144	0.145	0.144	0.145
	90th P ARE	0.303	0.297	0.301	0.301	0.301	0.301	0.301
	Std. ARE	0.140	0.141	0.139	0.139	0.141	0.139	0.139
	ΔM		-0.020	-0.085	-0.091	-0.099	-0.092	-0.101
	ΔP		0.021	0.006	0.006	0.005	0.006	0.005
	ΔS		-0.008	0.003	0.002	-0.011	0.002	0.004

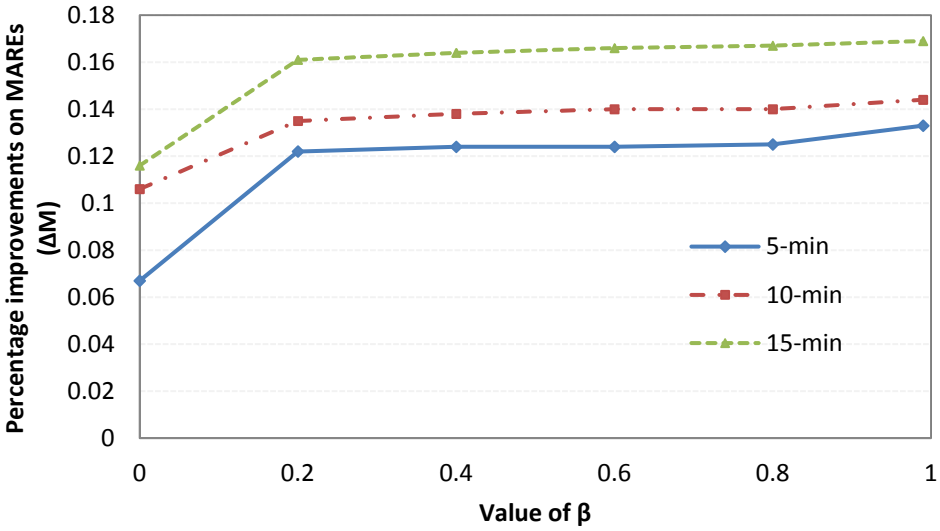


Figure 6.4: Variation of ΔM with different values of parameter β

6.4 Summary

This chapter proposes a short-term prediction model for predicting freeway travel times using data collected by Bluetooth detectors. The proposed model, which is a combination of dynamic outlier filtering algorithm and Kalman filter, focuses on two challenges of using AVI measurements for travel time prediction: (1) dynamic outlier detection must be able to respond to rapid traffic fluctuations, and (2) the time lag that exists in AVI measurements.

The proposed model is calibrated and validated using a dataset of freeway travel times collected by Bluetooth detectors. On the basis of the calibration and validation results, the following conclusions are made:

1. The use of the travel time trend terms (from real-time data and historical data) has a significant effect on improving the prediction accuracy of the proposed model for predictions 10 or more minutes into the future when traffic is in a transition state. For other conditions, the performance improvement resulting from the inclusion of these terms was not statistically significant.

2. The performance of the proposed model is relatively insensitive to the value of the parameter β as long as $\beta > 0$. On the basis of the sensitivity analysis, it is recommended to use a value of β=0.2.

Chapter 7
Model Application and Evaluation

The objective of this chapter is to evaluate the performance of the proposed prediction model by comparing the prediction results of the proposed model to the prediction results from two benchmark models - TransGuide and D&R (described in Section 2.1.3). To achieve this objective the proposed prediction model and the two benchmark models were applied to data associated with 2 field datasets, namely: (1) 401-H8/H24 (eastbound); and (2) 401-H24/H8 (westbound). Description of the study area can be found in Section 4.1. Detailed flow charts for implementation of the proposed method can be found in Appendix D.

The three prediction models were applied to the two datasets for travel time predictions in future 5, 10 and 15 minutes. For each dataset, data from 5 days were selected as "current" days (see Appendix A), and the rest of the data are considered as historical data. The selected 5 days from each dataset experienced varying levels of traffic congestion, but most congestion on this freeway segment (for both directions) is non-recurrent.

Parameters associated with the two benchmark models are selected using the same values (i.e. δ= 50% for TransGuide model, $\lambda = 3, \beta = 0.2, \beta_\sigma = 0.2$ for D&R model) that are used in Chapter 4.

The prediction models were evaluated using the same measures of performance (MOPs) that are used in Section 6.3, i.e. (1) mean absolute relative error (MARE); (2) 90th percentile of the absolute relative error (90th P ARE); and (3) standard deviation of the absolute relative error (Std. ARE). Each MOP is calculated in two different ways:

- Based only on the intervals for which traffic is in a free flow state ("free flow" defined as the case when the average travel speed is greater than 80 km/h).
- Based on only on the intervals for which traffic is in non-free flow state ("non-free flow" defined as the case when the average travel speed is equal to or less than 80 km/h).

The average travel speed is equal to the distance divided by the estimated mean travel time.

7.1 Application to Data from 401-H8/H24 (eastbound)

The dataset of 401-H8/H24 (eastbound) is the same dataset that was used to calibrate and validate the proposed model in Section 6.3. The following comparisons between the proposed model and the benchmark models are based on the prediction results of the proposed model with parameter $\beta = 0.2$. TransGuide model and D&R model were also applied to this dataset for travel time predictions in future 5, 10 and 15 minutes. Comparisons between the proposed model and the benchmark models (TransGuide and D&R) are conducted on the basis of the MOPs described previously. ΔM, ΔP and ΔS are computed for each comparison between the proposed model and one of the benchmark models, and the differences of the absolute relative error (ARE) between two compared models are tested using paired-z test. The results are shown in Table 7.1 and the percentage of improvements on MARE are shaded if the difference is statistically significant at the 95% level of confidence.

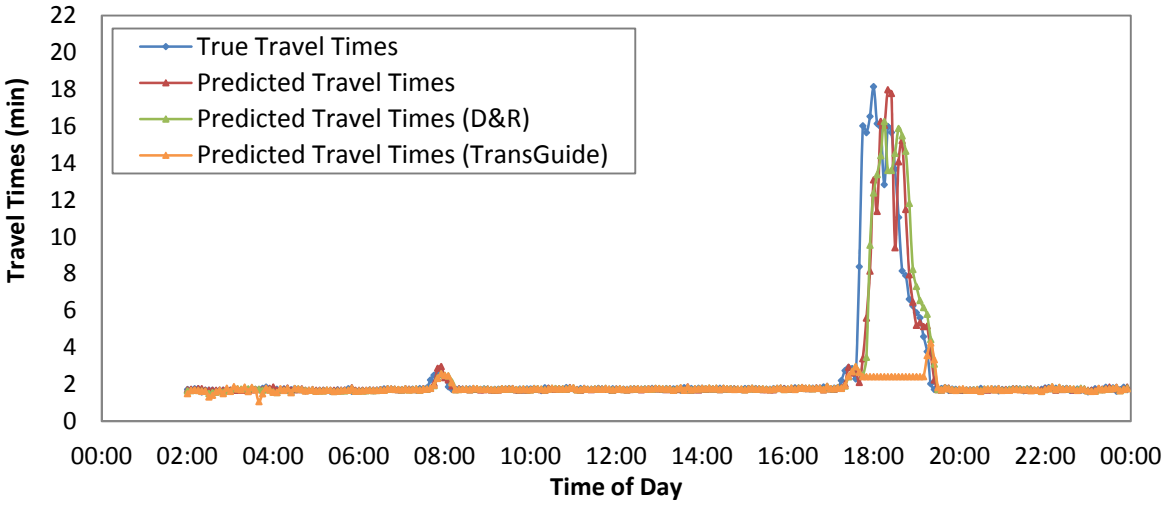
Table 7.1: Results of comparisons between proposed model and benchmark models (401 East)

5-min	Free Flow			Non-Free Flow		
	Proposed	TransGuide	D&R	Proposed	TransGuide	D&R
MARE	0.026	0.031	0.028	0.161	0.193	0.180
90th P ARE	0.057	0.065	0.060	0.340	0.510	0.399
Std. ARE	0.030	0.035	0.030	0.166	0.219	0.195
ΔM		0.177	0.088		0.167	0.104
ΔP		0.116	0.046		0.333	0.149
ΔS		0.161	0.000		0.242	0.148
10-min	Free Flow			Non-Free Flow		
	Proposed	TransGuide	D&R	Proposed	TransGuide	D&R
MARE	0.024	0.028	0.026	0.187	0.231	0.226
90th P ARE	0.047	0.058	0.053	0.418	0.603	0.549
Std. ARE	0.034	0.037	0.034	0.185	0.244	0.249
ΔM		0.154	0.092		0.190	0.174
ΔP		0.191	0.104		0.307	0.238
ΔS		0.088	0.000		0.241	0.257
15-min	Free Flow			Non-Free Flow		
	Proposed	TransGuide	D&R	Proposed	TransGuide	D&R
MARE	0.023	0.028	0.026	0.220	0.271	0.276
90th P ARE	0.045	0.056	0.051	0.493	0.674	0.686
Std. ARE	0.034	0.038	0.036	0.228	0.272	0.298
ΔM		0.167	0.118		0.186	0.202
ΔP		0.199	0.121		0.268	0.281
ΔS		0.095	0.055		0.160	0.235

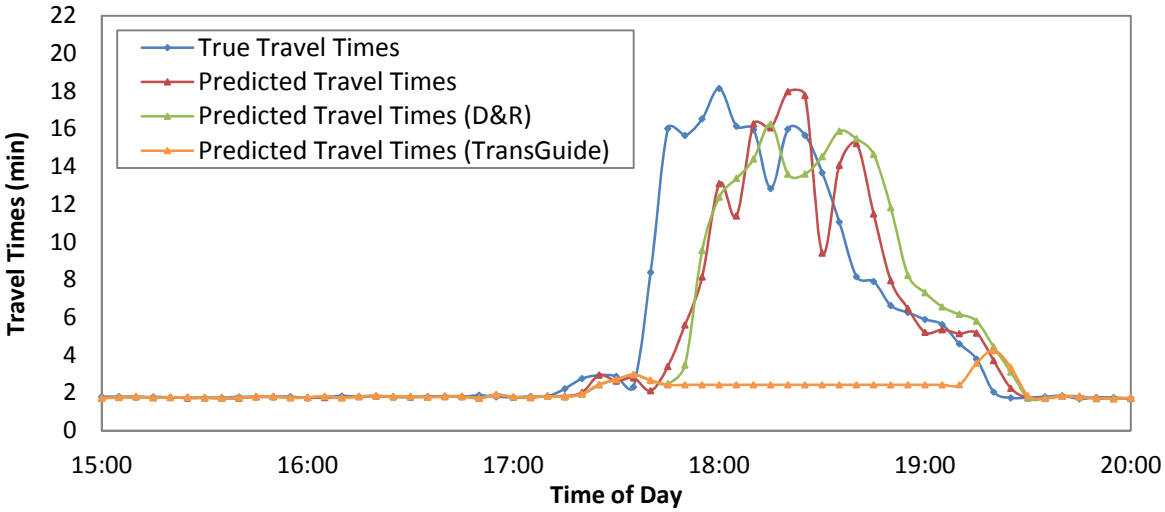
The results in Table 7.1 show that comparing to the benchmark models, the proposed model significantly improves prediction accuracy for both free flow and non-free flow states through all the tested prediction horizons (5, 10 and 15 minutes). The percentage of improvements on MARE varies from 8.8% to 17.7% under free flow conditions, and from 10.4% to 20.2% under non-free flow

conditions. Furthermore, the 90th percentile errors and standard deviations of errors also show improvements.

An example of the prediction results (prediction in future 5 minutes) for data collected from one day (Nov. 5th, 2012) are illustrated in Figure 7.1. Predictions were performed for a period from 2:00 am – 12 midnight for each day (as shown in Figure 7.1 (a)). The two hours of data prior to the current period were used to identify the similar traffic pattern in historical dataset.



(a) Overall illustration



(b) Detailed illustrations

Figure 7.1: Illustrations of the prediction results (for data collected at Nov. 5th, 2012)

The results in Figure 7.1 (a) show that in some cases the TransGuide model (i.e. results represented by orange line) was not able to track the sudden changes in travel times (e.g. predictions in time

period from 17:50 pm-19:10 pm), resulting in large prediction errors. This problem was solved when both D&R model and the proposed model were applied.

Figure 7.1 (b) shows the difference between the prediction results from the proposed model (red line) and from the D&R model (green line) for the time of day when severe non-recurrent congestion occurred. From the results of Figure 7.1 (b) we can observe that the proposed model responds to change of travel time more quickly than the D&R model, and therefore provides more accurate prediction results when traffic congestion is forming or dissipating. In particular, the proposed model performs better than the D&R model when congestion is dissipating.

7.2 Application to Data from 401-H24/H8 (westbound)

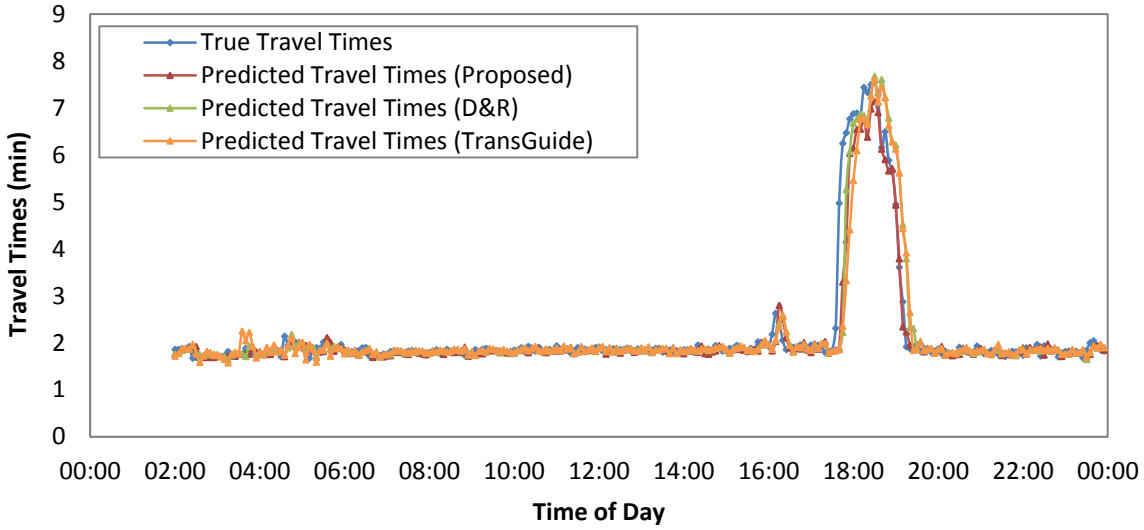
The data of 401 from H24 to H8 are collected from westbound of the study freeway segment. Similar to the eastbound dataset, data from 5 days were selected as "current" days (see Appendix A for days Jul. 27th 2012, Jan. 6th 2013, Jan. 28th 2013, Feb. 19th 2013 and Feb. 24th 2013), and the rest of the data are considered as historical data. Three prediction models (i.e. proposed model, TransGuide model and D&R model) were applied to this dataset for travel time predictions in future 5, 10 and 15 minutes. Comparisons between the proposed model and the benchmark models (TransGuide and D&R) are conducted using the same way as that was used in section 7.1 (shown in Table 7.2).

The results in Table 7.2 show that comparing to the benchmark models, the proposed model improves prediction accuracy for both free flow and non-free flow states through all the tested prediction horizons (5, 10 and 15 minutes). These improvements are statistically significant for all comparisons except one. The overall results show that the percentage of improvements of MARE varies from 14.3% to 30.6% under free flow conditions, and from 7.5% to 49.9% under non-free flow conditions. Furthermore, the 90th percentile errors and standard deviations of errors show improvements as well.

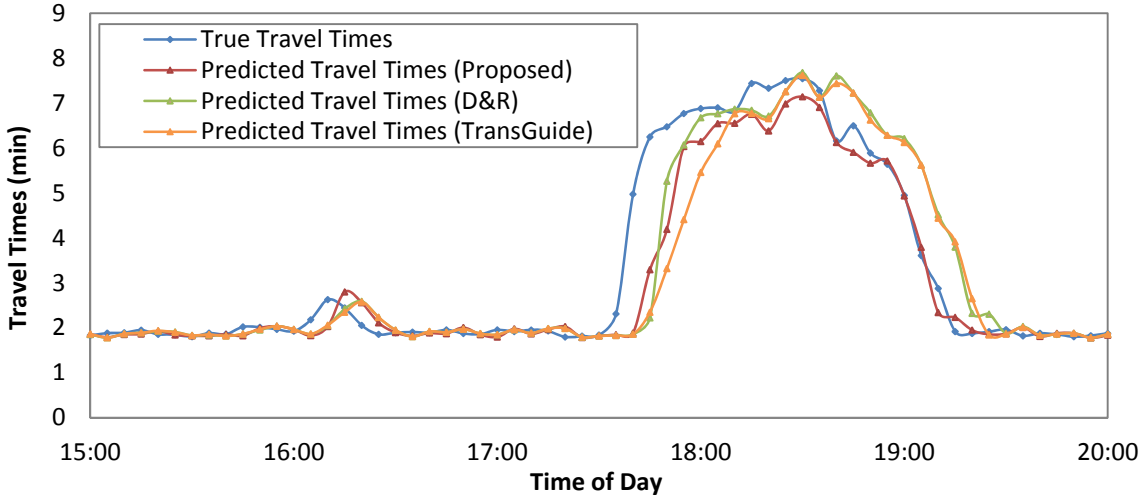
Table 7.2: Results of comparisons between proposed model and benchmark models (401 West)

5-min	Free Flow			Non-Free Flow		
	Proposed	TransGuide	D&R	Proposed	TransGuide	D&R
MARE	0.036	0.045	0.043	0.180	0.219	0.193
90th P ARE	0.087	0.093	0.086	0.405	0.571	0.501
Std. ARE	0.036	0.045	0.042	0.168	0.278	0.269
ΔM		0.274	0.199		0.217	0.075
ΔP		0.073	-0.010		0.410	0.239
ΔS		0.265	0.174		0.653	0.597
10-min	Free Flow			Non-Free Flow		
	Proposed	TransGuide	D&R	Proposed	TransGuide	D&R
MARE	0.032	0.042	0.038	0.190	0.284	0.226
90th P ARE	0.070	0.090	0.077	0.466	0.715	0.560
Std. ARE	0.031	0.044	0.041	0.201	0.369	0.348
ΔM		0.306	0.164		0.499	0.195
ΔP		0.301	0.115		0.535	0.201
ΔS		0.421	0.305		0.839	0.734
15-min	Free Flow			Non-Free Flow		
	Proposed	TransGuide	D&R	Proposed	TransGuide	D&R
MARE	0.032	0.038	0.036	0.215	0.271	0.262
90th P ARE	0.072	0.081	0.076	0.529	0.687	0.580
Std. ARE	0.034	0.042	0.042	0.249	0.420	0.420
ΔM		0.188	0.143		0.257	0.217
ΔP		0.131	0.060		0.299	0.097
ΔS		0.237	0.212		0.689	0.687

An example of the prediction results (prediction in future 5 minutes) for data collected from one day (July. 27th, 2012) are illustrated in Figure 7.2.



(a) Overall illustration



(b) Detailed illustrations

Figure 7.2: Illustrations of the prediction results (for data collected at July. 27th, 2012)

The results in Figure 7.2 (a) show that the sudden changes in travel times during time period from 17:35 pm-19:15: pm was detected by all the three tested prediction models, however the predicted travel times lag behind the true travel times within this time period. This phenomenon can be seen more clearly from Figure 7.2 (b), and the results also show that the proposed model responds to change of travel time more quickly than the TransGuide model and D&R model, especially for time period when congestion is dissipating. The performance of the proposed model when congestion is forming is notably better than TransGuide model, but only marginally better than the D&R model.

It should be noted that the proposed model performs well for data collected from the westbound direction even through the model was calibrated using eastbound data. This suggests that the model is robust to different traffic conditions.

7.3 Results Discussion

Combining the results from eastbound and westbound, the variations of the prediction errors of the proposed model with true travel times are shown in Figure 7.3.

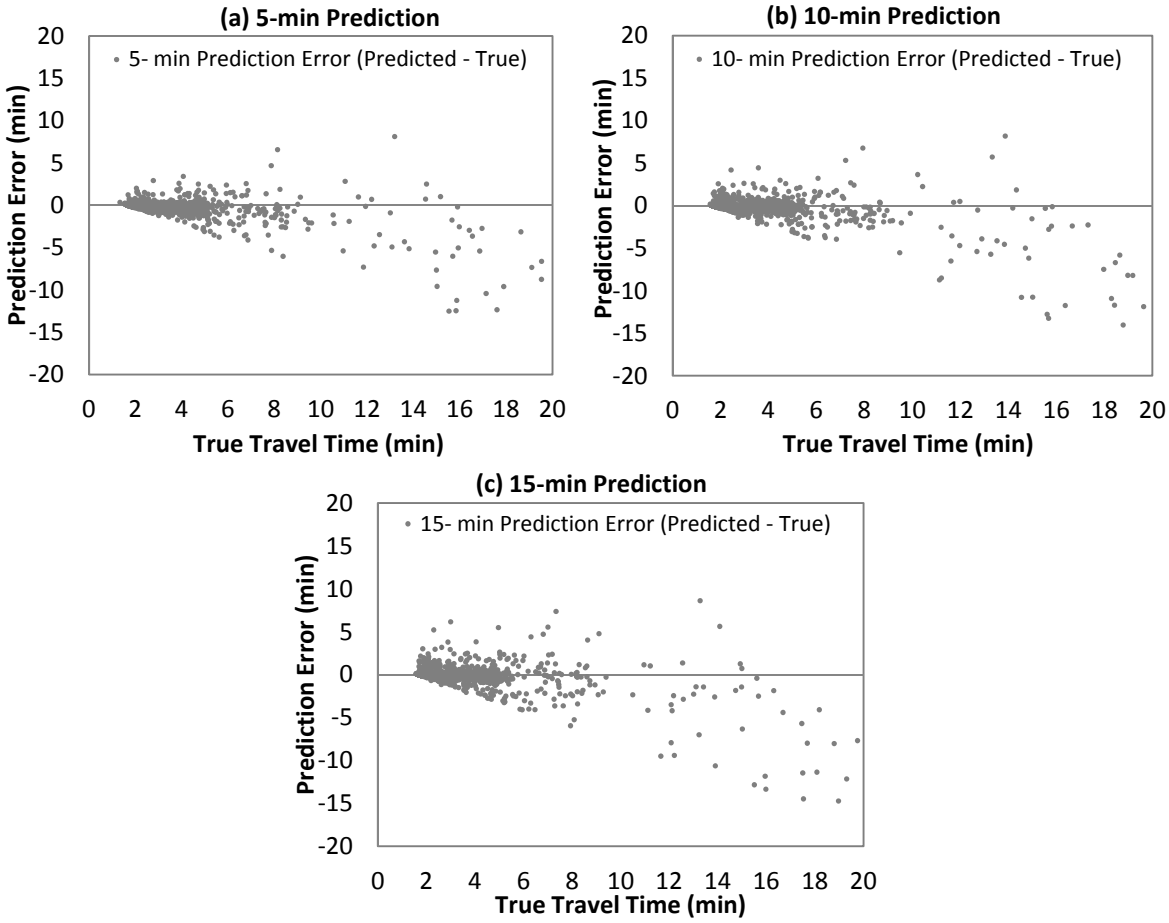


Figure 7.3: Variation of the prediction errors with true travel time

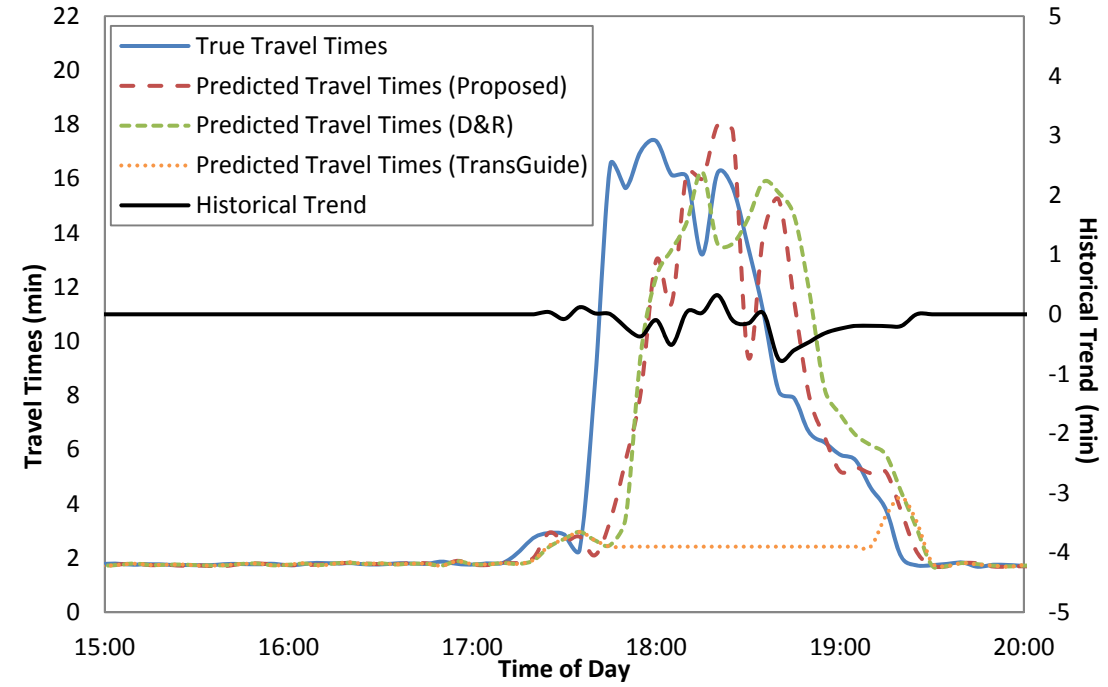
The results in Figure 7.3 indicate that when the true travel time (for either 5, 10 or 15 minutes into future) is less than 5 minutes (which is approximately three times the free speed travel time), then the prediction errors tend to relatively consistent. However, as the true travel time increases, the prediction error also tends to increase, and there is a bias to under-estimating the true travel time. The prediction results (5-min prediction horizon) of the proposed model to the 10 tested days are shown in

Appendix C. These results indicate that despite the improved performance of the proposed model relative to the benchmark methods, the proposed model still has difficulties tracking rapid increases in travel time associated with non-recurrent congestion.

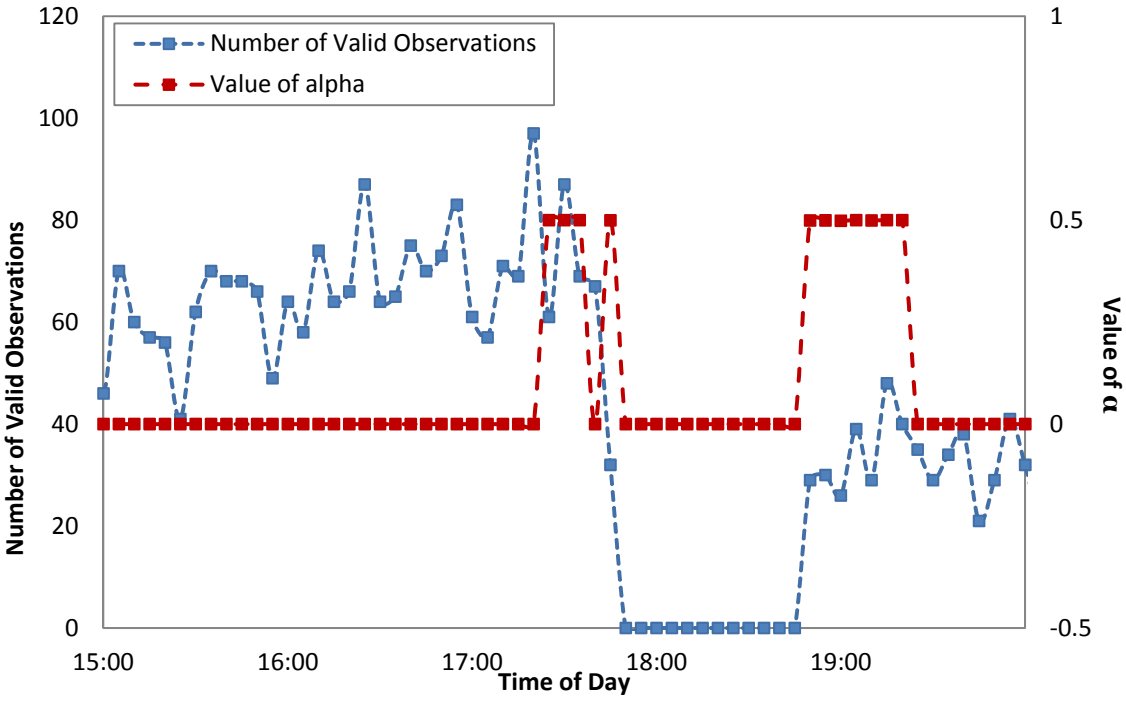
Detailed illustrations of the proposed algorithm being applied to datasets from two representative days (Nov. 5th, 2012 and July. 27th, 2012) for the time of day when severe non-recurrent congestion occurred are shown in Figure 7.5 and 7.6. Figures 7.5 (a) and 7.6 (a) illustrate the prediction results and the associated historical trend value, and Figures 7.5 (b) and 7.6 (b) illustrate the number of valid observations in real-time and the value of α determined using Equation 6.5.

Based on the results shown in Figure 7.5 and 7.6, the following observations are made:

1. Use of historical data does help for improving the prediction accuracy when congestion is dissipating, but it does very little to improve the prediction accuracy for congestion forming as a result of an incident.
2. The average of historical trend data results in only very small changes in any given time period, consequently the use of historical data cannot significantly improve the accuracy of prediction for non-recurrent congestion, especially for congestion forming. It's not clear yet about how much it helps for recurrent congestion.
3. When the real-time data are not available for a time period that serious traffic congestion occurs (e.g. 17:55pm – 18:45pm in Figure 7.5 (b)), the α behave as expected ($\alpha = 0$), i.e. the predicted travel time is a combination of the estimated travel time in previous interval and historical trend value. However, the value of α determined based on the number of valid observations does not change smoothly as expected – it is almost binary, which is mainly because the α is very sensitive to the change of the number of valid observations, and the change of the number of valid observations is not smooth.

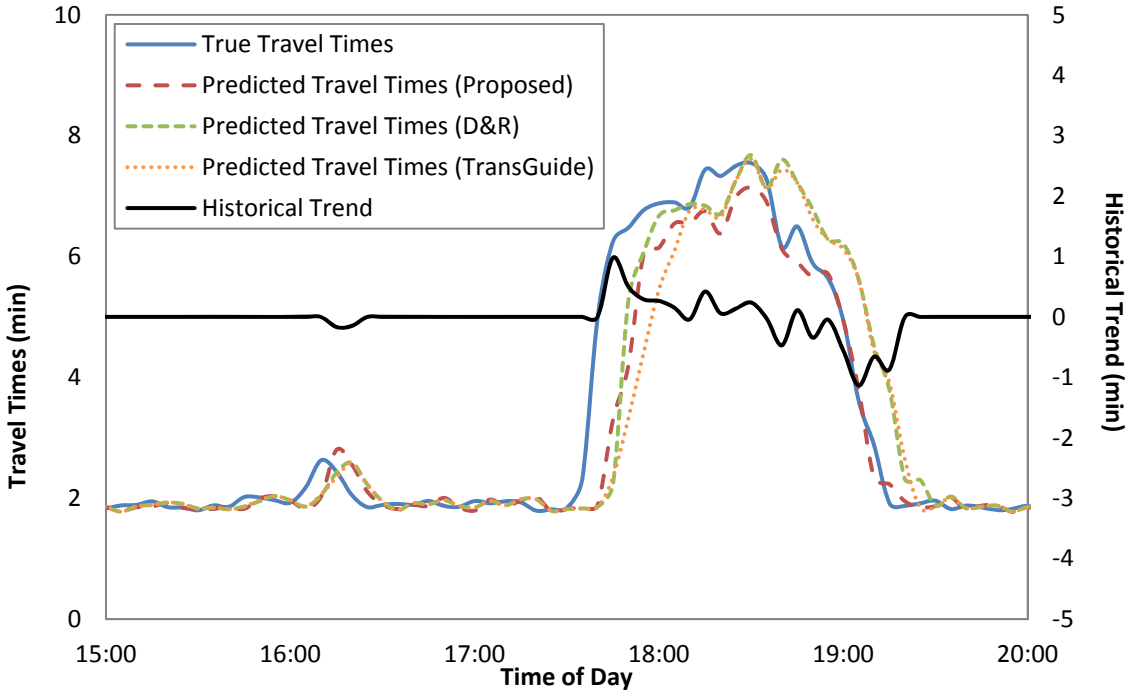


(a)

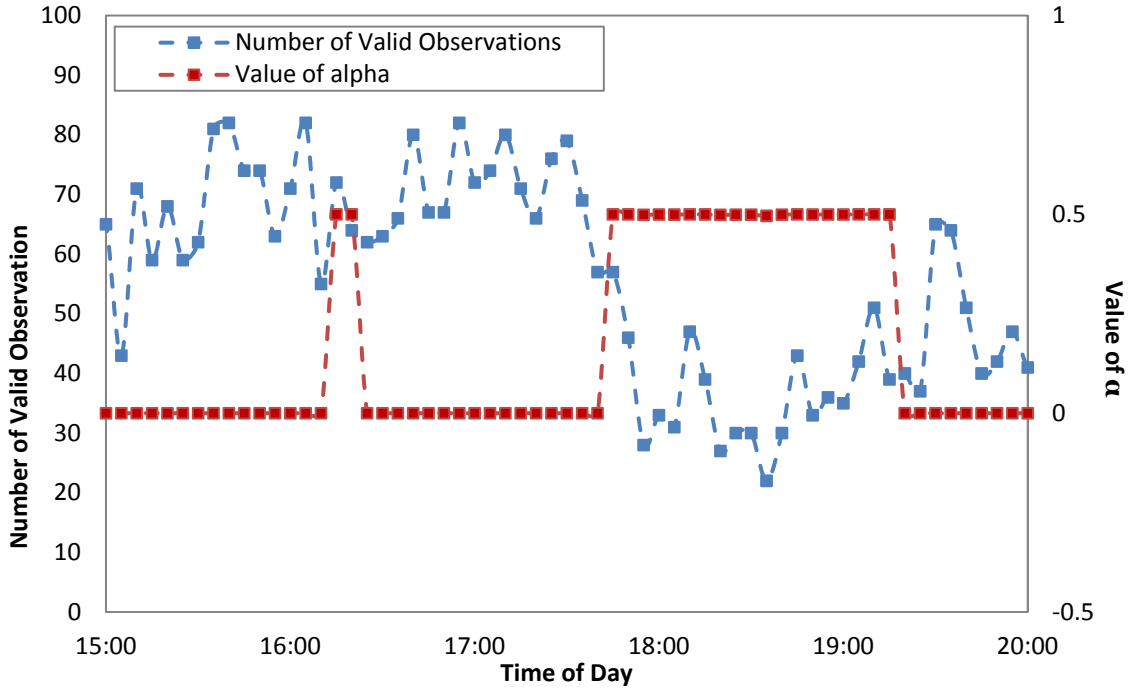


(b)

Figure 7.4: Illustration of the algorithm details (for data collected from eastbound at Nov. 5th, 2012)



(a)



(b)

Figure 7.5: Illustration of the algorithm details (for data collected from westbound at July. 27th, 2012)

To sum up, the large prediction error of the proposed model is mainly caused by the lag effects of AVI measurements. The use of the historical trend data in the proposed model is intended to address this lag effect; however it is hypothesized that this approach is most effective for recurrent congestion rather than non-recurrent congestion. Unfortunately, the existing data set does not contain significant recurrent congestion and therefore the extent to which performance improvements under recurrent conditions can be attributed to the use of historical trend data could not be quantified. Further investigation on improving the prediction accuracy of non-recurrent congestion should focus on using information that does not suffer from the lag effect (e.g. the variation of the number of detections).

7.4 Summary

In this chapter, the performance of the proposed model is evaluated. The prediction results from the proposed model were compared to the prediction results from two benchmark models (TransGuide and D&R). In addition, an investigation on the proposed algorithm details was provided. On the basis of these evaluations the following conclusions are made:

1. The proposed model is superior to two benchmark models (TransGuide and D&R) for both congested and uncongested traffic conditions. Performance improvements (in terms of the reduction in the MARE) vary from 8.8% to 30.6% under free flow conditions, and from 7.5% to 49.9% under non-free flow conditions.
2. The proposed model also shows improvements (reductions) in the 90th percentile errors and standard deviations of errors for all traffic conditions.
3. The proposed model still has difficulties tracking rapid increases in travel time associated with non-recurrent congestion, which is mainly caused by the lag effects of AVI measurements.

The evaluations described in this chapter are for a suburban freeway section which does not contain any intermediate junctions, and does not experience significant recurrent congestion. The method is applicable to sections with intermediate junctions, so method evaluations for freeway section which contain intermediate junctions are recommended. Also, it is recommended that the proposed model be evaluated on another freeway section which does experience substantive recurrent congestion.

Further efforts to improve the accuracy of travel time predictions made on the basis of Bluetooth data should focus on improving performance for non-recurrent events when congestion is forming. Potential approaches could include: (1) using other attributes of data that can be collected by Bluetooth detectors, such as the number of detections; and/or (2) considering the spatial interactions

along a roadway such that the effect of queues spilling back into the road segment of interest can be captured.

Chapter 8
Conclusions and Recommendations

Pre-active traffic management used to maximize the efficiency of the existing transportation systems, as one of the solutions to alleviate the growing problem of traffic congestion, become more and more attractive to traffic managers. The prerequisite of implementing pre-active traffic management is dynamically estimating/predicting roadway conditions using traffic data collected in-real-time.

Bluetooth traffic monitoring technologies provide the opportunity to collect wide area real-time travel time data with low cost, something that is not feasible with the traditional traffic monitoring technologies (e.g. Loop detectors). However, as with other AVI technologies (e.g. electronic toll tags, license plate recognition, etc.), using the Bluetooth data (i.e. travel time measurements) for travel time prediction requires: (1) determining the optimal spacing between detectors; (2) a reliable real-time outlier detection algorithm; and (3) a method for addressing data gaps caused by the time lag inherent in the travel time measurements.

In this dissertation, we developed methods to address the above problems for providing reliable travel time prediction in real-time using Bluetooth data. This chapter highlights the main contributions of this thesis research and presents directions for future work.

8.1 Major Contributions

The major contribution of this research concerns the practical solutions to the critical problems of reliable travel time prediction using Bluetooth data.

The specific contributions made in this dissertation are as follows:

1. Quantified the difference between Bluetooth measured travel time (ATT) and true travel time (DTT), and demonstrated that the real-time estimation error caused by using ATT directly as an estimate of DTT is not negligibly small, especially when traffic is in congestion state. Moreover, evidence was provided to show that the temporal variation pattern of ATT is different from DTT which further degrades the accuracy of travel time prediction.
2. Quantified the impact that Bluetooth detector spacing has on the real-time estimation errors, and developed a generalized regression model that can be used to determine the optimal average

spacing of Bluetooth detector deployments on urban freeways as a function of the length of the route for which travel times are to be estimated.

3. Implemented and evaluated two existing real-time travel time outlier filtering algorithms (TransGuide and D&R), and identified that both of these two algorithms are not able to perform reliably when travel times change rapidly.

4. Developed and validated a traffic flow theory based travel time outlier filter enhancement, which can be used as an extension to existing data driven outlier detection algorithms. The proposed method improves the performance of existing data driven outlier detection algorithms for periods when travel times are changing rapidly, as when congestion is forming or dissipating. The proposed method doesn't require off-line calibration of parameters and therefore is simple and easy to implement.

5. Validated that the historical data selected by K nearest neighbor (KNN) method provide estimates of the travel time pattern (trend) that are statistically significantly better than historical data selected by simple aggregation (SA) method. A method for calibrating the KNN method using field Bluetooth data was developed and demonstrated.

6. Developed a model for predicting near future freeway travel times using Bluetooth data with special attention to the time lag that exists in the Bluetooth measurements. Calibrated and validated the proposed model, and showed with evidence that the use of the travel time trend terms (from real-time data and historical data) has a significant effect on improving the prediction accuracy of the proposed model for predictions 10 or more minutes into the future when traffic is in a transition state. On the basis of a sensitivity analysis, it was found that the performance of the proposed model is relatively insensitive to the value of the parameter β as long as $\beta > 0$, therefore it is recommended to use a value of β=0.2.

7. Demonstrated the performance of the proposed model by comparing the prediction results between the proposed model and two benchmark models. The models were applied to two datasets of freeway travel times collected by Bluetooth detectors. The comparison results indicate that the proposed model significantly improves the accuracy of travel time prediction for 5, 10 and 15 minutes prediction horizon under both free flow and non-free flow traffic states. The 90^{th} percentile errors and standard deviation of the prediction errors are also improved.

8.2 Future Research

For improving and complementing this research, the following problems are identified for future research:

1. In this research, the generalized model used to determine the optimal detector spacing was developed based on the simulation data because the field data were not available. It is recommended to test and validate the proposed model using field data if it is available.
2. The proposed traffic flow theory based travel time outlier filtering model was tested and validated using data from a 3.1km freeway segment with sampling rate of approximately 9%. It is recommended that the proposed model be tested on data from different road sections (e.g. road sections with different lengths and/or with different sampling rate) to verify its transferability.
3. The dataset used to test and evaluate the proposed short-term travel time prediction model does not contain periods of significant recurrent congestion, therefore it is recommended that the proposed model be tested and evaluated on freeway sections which experience substantive recurrent congestion, in order to verify the hypothesis that the problem of tracking rapid changes of travel time from free flow to congestion (i.e. congestion forming) can be improved by using the historical trend data.
4. Further efforts to improve performance for non-recurrent events when congestion is forming should focus on: (1) using other attributes of data that does not suffer from the lag effect, such as the number of detections; and/or (2) considering the spatial interactions along a roadway such that the effect of queues spilling back into the road segment of interest can be captured.

Bibliography

Aliari, Y. and A. Haghani. 2012. Bluetooth Sensor Data and Ground Truth Testing of Reported Travel Times. Transportation Research Record: Journal of the Transportation Research Board, Vol.2308(-1), pp.167-172.

Ban, X., L. Chu, R. Herring and J. Margulici. 2009. Optimal Sensor Placement for Both Traffic Control and Traveler Information Applications. Transportation Research Board 88th Annual Meeting, Washington D.C.

Barcelo, J., L. Montero, L. Marqués, P. Marinelli, and C. Carmona. 2010. Travel Time Forecasting and Dynamic OD Estimation in Freeways Based on Bluetooth Traffic Monitoring. Journal of the Transportation Research Board, Vol. 2175(-1), pp. 19-27.

Billings, D., and J. S. Yang. 2006. Application of the ARIMA models to urban roadway travel time prediction-a case study. Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference.

Blue, V., G.F. List, and M.J. Embrechts. 1994. Neural Network Freeway Travel Time Estimation. Proceedings of Intelligent Engineering Systems through Artificial Neural Networks American Society of Mechanical Engineers, New York, 1994, pp. 1135-1140.

BluSTATs Operations Manual, BluSTATs version 1.4e, August 6, 2009.

Bucur, L., A. Florea, B.S. Petrescu. 2010. An adaptive fuzzy neural network for traffic prediction. 18th Mediterranean Conference on Control and Automation, MED'10 - Conference Proceedings, pp. 1092- 1096.

Chang, EC-P. 1999. Traffic Estimation for Proactive Freeway Traffic Control. Transportation Research Record, no. 1679, pp. 81-86, 1999

Chen, M. and S.I.J. Chien. 2001. Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data: Link Based vs. Path Based. In Transportation Research Record: Journal of the Transportation Research Board, No. 1768, Transportation Research Board of the National Academies, Washington, D.C., pp. 157-161

Chien, S. I-J. and C.M. Kuchipudi. 2003. Dynamic travel time prediction with real-time and historic data. Dynamic Travel Time Prediction with Real-Time and Historic Data. Journal of Transportation Engineering © ASCE.

Chien, S.I-J. X. Liu, and K. Ozbay. 2003a. Predicting Travel Times for the South Jersey Real –Time Motorist Information System. Transportation Research Record: Journal of the Transportation Research Board, No. 1855, Transportation Research Board of the National Academies, Washington, D.C., pp. 32-40.

Chrobok, R., et al., Different methods of traffic forecast based on real data. European Journal of Operational Research, 2004. 155: p. 558-568.

City of Calgary Website: http://www.calgary.ca/Transportation/Roads/Pages/Traffic/Traffic-management/ Bluetooth-detection-system.aspx, Accessed at April 26th, 2013

Chu, L. 2005. Adaptive Kalman filter based freeway travel time estimation. In: Transportation Research Board 84th Annual Meeting, Washington DC, 2005.

Clark, S. D., S. Grant-Muller, and H. Chen. Cleaning of Matched License Plate Data. In Transportation Research Record: Journal of the Transportation Research Board, Vol. 1804, 15 2002, pp. 1-7.

Coifman, B.. Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors. Transportation Research: Part A, Vol 36, No 4., 2002, pp. 351-364.

Cortes, C.E., Lavanya, R., Oh, J.S., and Jayakrishnan, R. A General Purpose Methodology for Link Travel Time Estimation Using Multiple Point Detection of Traffic. Institute of Transportation Studies, U of California, 2001.

Davis, G. A., N. L. Nihan, M. M. Hamed, and L. N. Jacobson. 1990. Adaptive forecasting of freeway traffic congestion. Transportation Research Record: Journal of the Transportation Research (1287): 29-33.

Edara, P., J. Guo, B. L. Smith, and C. McChee. Optimal Placement of Point Detectors on Virginia's Freeways: Case Studies of Northern Virginia and Richmond. Final Contract Report, VTRC 08-CR3, Virginia Transportation Research Council. 2008.

Fei, X., Lu, C-C., Liu, Ke Liu, A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction, Transportation Research Part C 19,2011, 1306–1318

Fowkes, A. The use of Number Plate Matching for Vehicle Travel Time Estimation. In PTRC Proceedings of the 11th Annual Conference, University of Sussex, 1983, pp. 141-148.

Guin, A., Travel Time Prediction using a Seasonal Autoregressive Integrated Moving Average Time Series Model. 2006 IEEE Intelligent Transportation Systems Conference, Toronto, Canada, September 17-20, 2006, pp. 493-498

Haghani, A., M. Hamedi, K. F. Sadabadi, S. Young, and P. Tarnoff. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2160, 2010, pp. 60-68.

Haas, C. T., H. S. Mahmassani, T. Rioux, M. Haynes, J. Khoury, and H. Logman. Recommendations and Implementation of Automatic Vehicle Identification for Incident Detection and Advanced Traveler Information Systems. Project Summary Report 7-4957-S Project 7-4957: Evaluation of Automatic Identification System Implementation in San Antonio , 2001.

Hellinga, B. and G. Knapp. AVI Based Incident Detection. Transportation Research Record - Journal of the Transportation Research Board 1727. National Academy Press, Washington, D.C. 2000, pages 142 – 153.

Hellinga B. and Izadpanah P.. An Opportunity Assessment of Wireless Monitoring of Network-Wide Road Traffic Conditions-Final Report. Prepared for: Mr. Ataur Bacchus, Intelligent Transportation Systems Office, Ministry of Transportation of Ontario, 2007.

Dion, F., and H. Rakha. Estimating Dynamic Roadway Travel Times using Automatic Vehicle Identification Data for Low Sampling Rates. Transportation Research Part B: Methodological, 20 Vol. 40, No. 9, 2006, pp. 745-766.

Guin Angshuman, Travel Time Prediction using a Seasonal Autoregressive Integrated Moving Average Time Series Model. 2006 IEEE Intelligent Transportation Systems Conference, Toronto, Canada, September 17-20, 2006.

Houston TranStar Website: www.houstontranstar.org/ , Accessed at April 26th, 2013

Ishak, S., and H. Al-Deek. 2002. Performance Evaluation of Short-Term Time-Series Traffic Prediction Model. ASCE Journal of Transportation Engineering, Vol. 128, No. 6, pp. 490-498.

Izadpanah P.. Freeway Travel Time Prediction Using Data from Mobile Phone Probes. A proposal presented to the University of Waterloo in fulfillment of the thesis requirement for the degree of

Doctor of Philosophy in Civil Engineering Department of Civil and Environmental Engineering Waterloo, Ontario, Canada, 2007(a).

Izadpanah P. and Hellinga B.. Wide-area Wireless Traffic Conditions Monitoring: Reality or Wishful Thinking? Proceedings of the ITE Canadian District Annual Conference, Toronto, Canada, 2007(b).

Izadpanah, P., B. Hellinga, L. Fu. Real-Time Freeway Travel Time Prediction Using Vehicle Trajectory Data. Presented at 90th Annual Meeting of Transportation Research Board, Washington D.C, 2011

Jiang, X. and H. Adeli. 2005. Dynamic wavelet neural network model for traffic flow forecasting. Journal of transportation engineering. 131(10): p. 771-779.

Klein, et al. Traffic Detector Handbook published by the U.S. Federal Highway Administration, 2006.

Kurzhanskiy, Alex A; Varaiya, Pravin. 2010. Philosophical transactions. Active traffic management on road networks: a macroscopic approach. Series A, Mathematical, physical, and engineering sciences, 368(1928): 4607- 4626, October 13, 2010

Lee, S. and D.B. Fambro, Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transportation Research Record, 1999. 1678: p. 179-188

Lin W., Dahlgren J., and Huo H.. Enhancement of Vehicle Speed Estimation with Single Loop Detectors. In Transportation Research Record: Journal of the Transportation Research Board, No. 1870, TRB, National Research Council, Washington D.C., 2004, pp. 147-152.

Lu,X.Y., and A.Shabardonis. Freeway Traffic Shockwave Analysis: Exploring the NGSIM Trajectory Data. 86th Annual Meeting Transportation Research Board Washington, D.C., 2007.

Martchouk, M., F. Mannering and D. Bullock. Analysis of Freeway Travel Time Variability Using Bluetooth Detection. Journal of Transportation Engineering. 137(10), 2011, pp. 697–704.

Ministry of Transportation Ontario website, http://www.mto.gov.on.ca/english/transtek/roadtalk/rt17-3/index.shtml#a3, Accessed at Nov. 20th, 2013

May, A.D. Traffic Flow Fundamentals. Prentice-Hall, ISBN 0-13-926072-2, 1990.

Mouskos, K.C., Niver, E., Pignataro, L.J., Lee, S.. Transmit System Evaluation. Final Report, Institute for Transportation, New Jersey Institute of Technology, Newark, NJ, 1998.

Myers K.A. and B.D.Tapley. Adaptive Sequential Estimation with Unknown Noise Statistics. IEEE Transactions on Automatic Control, Vol. 21, No. 4, 1976, pp.520-523.

Myung, J. Kim, D. Kho, S. and Park, C. 2011. Travel Time Prediction Using the K-Nearest Neighbourhood Method with Combined VDS and ATC Data. Transportation Research Board, 90th Annual Meeting, Washington D.C.

Nikovski, D., et al. Univariate short-term prediction of road travel times. In Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems. 2005

OECD Territorial Reviews: Toronto, Canada. 2010.

Oh, J.S., Jayakrishnan, R. and W.Recker. Section Travel Time Estimation from Point Detection Data. Institute of Transportation Studies, U of California, Irvine/Submitted to 82nd TRB Conference, 2002.

Oh, S. and K. Choi. Optimal Detector Location for Estimating Link Travel Time Speed in Urban Arterial Roads. KSCE Journal of Civil Engineering May 2004, Volume 8, Issue 3, pp 327-3335.

Park, D. and Rilett L.R., Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks. Transportation Research Record, 1617, 1998, pp. 163-170.

Park, D, and L.R. Rilett. Forecasting Freeway Link Travel Times with A Multilayer Feedforward Neural Network. Computer Aided Civil and Infrastructure Engineering Vol. 14, 1999, pp.357-367.

Pekilis B.. Overview of IntelliDriveSM. Networked Vehicle Foundation Workshop, CU-ICAR, Greenville, South Carolina, 2009.

Petty, K. F. and J. J. Peter Bickel, Michael Ostland, John Rice, Ya'acov Ritov, Frederic Schoenberg. Accurate estimation of travel times from single-loop detectors. Transportation Research Part A: Policy and Practice, Volume 32, Number 1, 1998, pp. 1-17(17).

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

RITA Intelligent Transportation Systems website: http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm, Accessed at Nov. 20th, 2013

Robinson, S., and J. Polak. Overtaking Rule Method for the Cleaning of Matched License-Plate Data. Journal of Transportation Engineering, Vol. 132, 2006, pp. 609.

Salek Moghaddam S. and B. Hellinga. Quantifying Measurement Error in Arterial Travel Times Measured by Bluetooth Detectors. In Transportation Research Record: Journal of the Transportation Research Board, 2013

Sherali, H.D., J. Desai, and H. Rakha. A Discrete Optimization Approach for Locating Automatic Vehicle Identification Readers for the Provision of Roadway Travel Times. Transportation Research B, 2006. 40:857-871

Smith, B.L., B.M. Williams, and R.K. Oswald, Comparison of parametric and nonparametric models for traffic flow forecasting, Transportation Research Part C, 2002, 10: p. 303-321.SwRI, Automatic Vehicle Identification Model Deployment Initiative – System Design Document. Report prepared for TransGuide, Texas Department of Transportation, Southwest Research Institute, San Antonio, TX, 1998.

Tan, G., W. Yuan, and H. Ding. 2004. Traffic flow prediction based on generalized neural network. Proceedings in IEEE 2004 Intelligent Transportation Systems Conference.

Traffax Inc. BluSTATs Operations Manual. BluSTATs version 1.4e. 2009

University of Maryland, Bluetooth Traffic Monitoring Technology --- Concept of Operation &Deployment Guidelines. Center for Advanced Transportation Technology, 2008.

U.S. DOT, Federal Highway Administration. 2005. Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation. Report prepared by Cambridge Systematics, Inc.100 Cambridge Park Drive, Suite 400 Cambridge, Massachusetts 02140 with Texas Transportation Institute

U.S. DOT, Federal Highway Administration. 2003. Freeway Management and Operations Handbook.

van Hinsbergen, C. P. IJ., van Lint, J. W. C.. Bayesian Combination of Travel Time Prediction Models. Transportation Research Record: Journal of the Transportation Research Board, No. 2064, 2008, pp. 73–80.

van Hinsbergen, C.P., J.W.C. van Lint, and F. M. Sanders, Short Term Traffic Prediction Models. In Proc., 14th World Congress on ITS, Beijing, China. 2007

van Lint J.W.C., S. P. Hoogendoorn, and H. J. van Zuylen. Freeway Travel Time Prediction with State-Space Neural Networks-Modeling State-Space Dynamics with Recurrent Neural Networks. Transportation Research Record 1811, 2002, pp.30-39.

van Lint, J.W.C.. Reliable Travel Time Prediction for Freeways. PhD thesis, Faculty of Civil Engineering and Geosciences, Transportation and Planning Section, Delft University of Technology, 2004.

van Lint, J.W.C., Hoogendoorn, S.P., and H.J. Van Zuylen. Accurate Freeway Travel Time Prediction with State-Space Neural Networks under Missing Data. Transportation Research Part C: Emerging Technologies, Vol. 13, 2005, pp. 347-369.

Waterloo traffic website: http://transblue.uwaterloo.ca

Waller, S. T., Y.-C. Chiu, N. Ruiz-Juri, A. Unikrishnan, and B. Bustillos. Short Term Travel Time Prediction on Freeways in Conjunction with Detector Coverage Analysis. Technical Report FHWA/TX-08/0-5141-1. Center for Transportation Research, University of Texas at Austin for Texas Department of Transportation, Austin, 2007.

Wang, L. Shen, H. Liu. Adjustments based on wavelet transform ARIMA model for network traffic prediction. Computer Engineering and Technology (ICCET), 2010 2nd International Conference on 16-18 April 2010. V4-520 - V4-523.

Wang, Y., M. Papageorgiou,; A. 2007. Messmer. Investigation of the adaptive features of a real time nonlinear freeway traffic state estimator. Nonlinear Dynamics. Vol. 49, no. 4, pp. 511-524.

Wang, Y., M. Papageorgiou, A. Messmer. 2008. Real-time freeway traffic state estimation based on extended Kalman filter: adaptive capabilities and real data testing. Transportation Research, Part A (Policy and Practice), vol.42,no.10,pp.1340-1358.

Welch Greg and Bishop Gary. An Introduction to the Kalman Filter. Department of Computer Science University of North Carolina at Chapel Hill, July, 2006.

Wild, D., 1997. Short-term forecasting based on a transformation and classification of traffic volume time series. International Journal of Forecasting 13 (1), 63–72.

Work, D.B. Tossavainen, O.-P. Blandin, S. Bayen, A.M. Iwuchukwu, T. Tracton, K. An Ensemble Kalman Filtering Approach to Highway Traffic Estimation Using GPS Enabled Mobile Devices. Decision and Control, 47th IEEE Conference, 2008, pp. 5062-5068

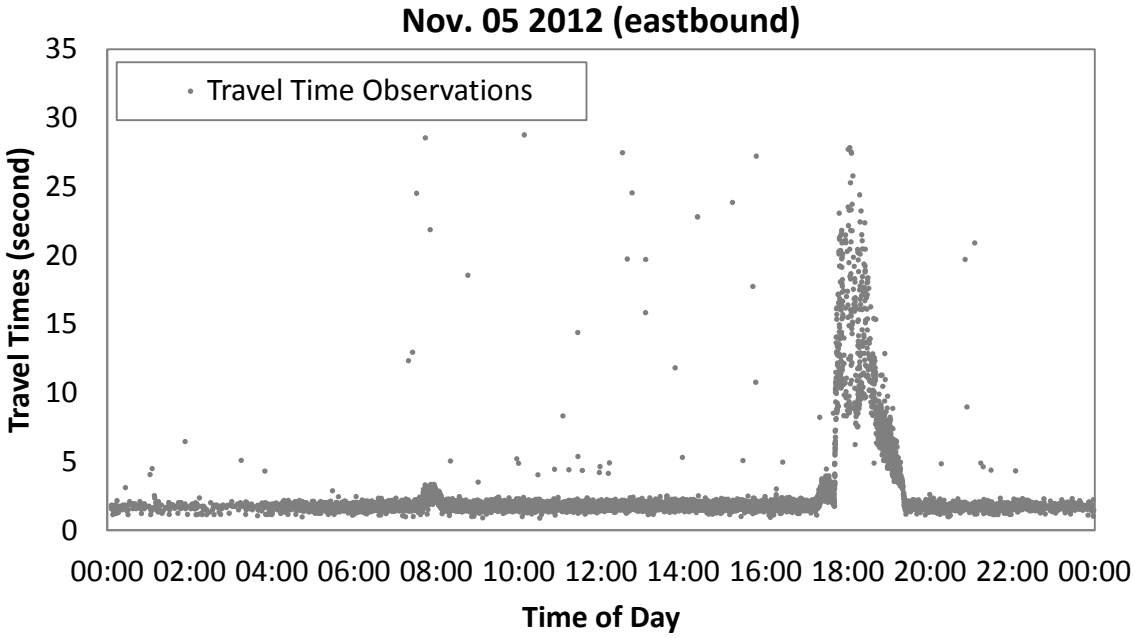
Wunderlich K., Burnier C., Larkin J., Shah V., and Vasudevan M.. Final Report : Vehicle-Infrastructure Integration (VII) Probe Data Characteristics: Analyses of Three Field Data Sets. Falls Church Virginia, Contract Sponsor: Federal Highway Administration, 2007.

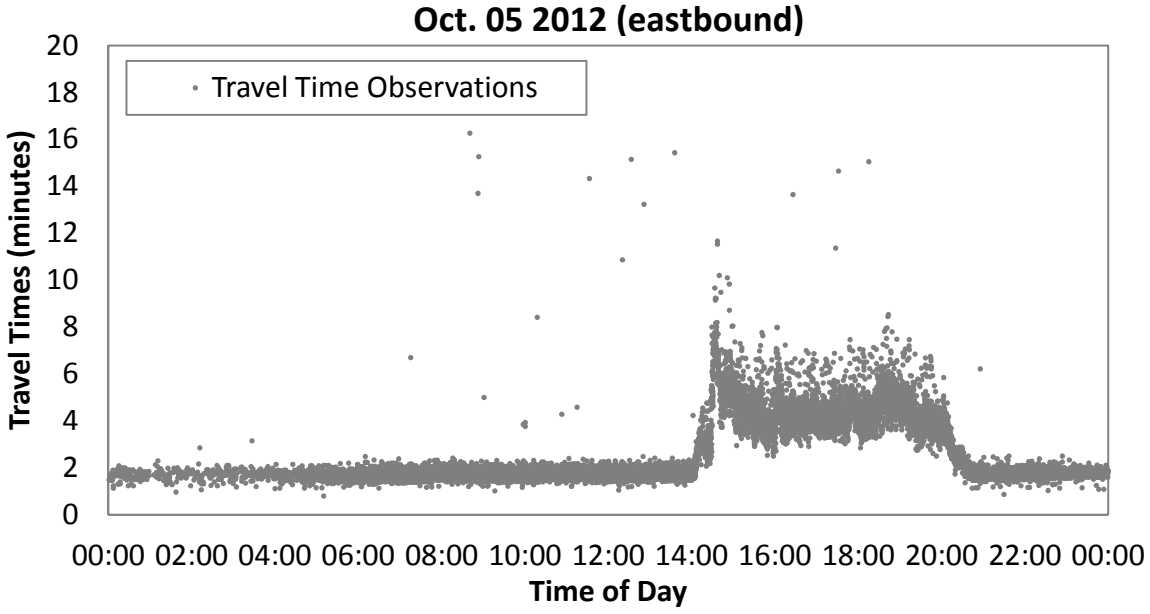
Yang J-S.. A Study of Travel Time Modeling via Time Series Analysis. 2005 IEEE Conference on Control Applications, Toronto, Canada, August 28-31, 2005.

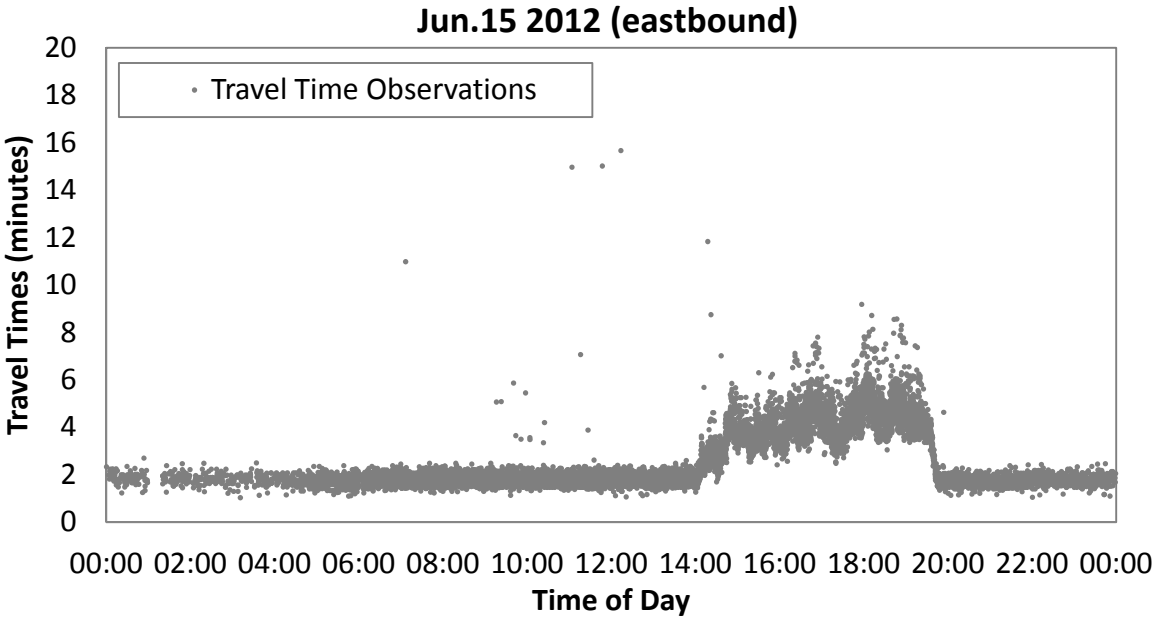
Zhang, X., Rice, John A..2003. Short-Term Travel Time Prediction Using A Time-Varying Coefficient Linear Model, Transportation Research Part C 11 187–210.

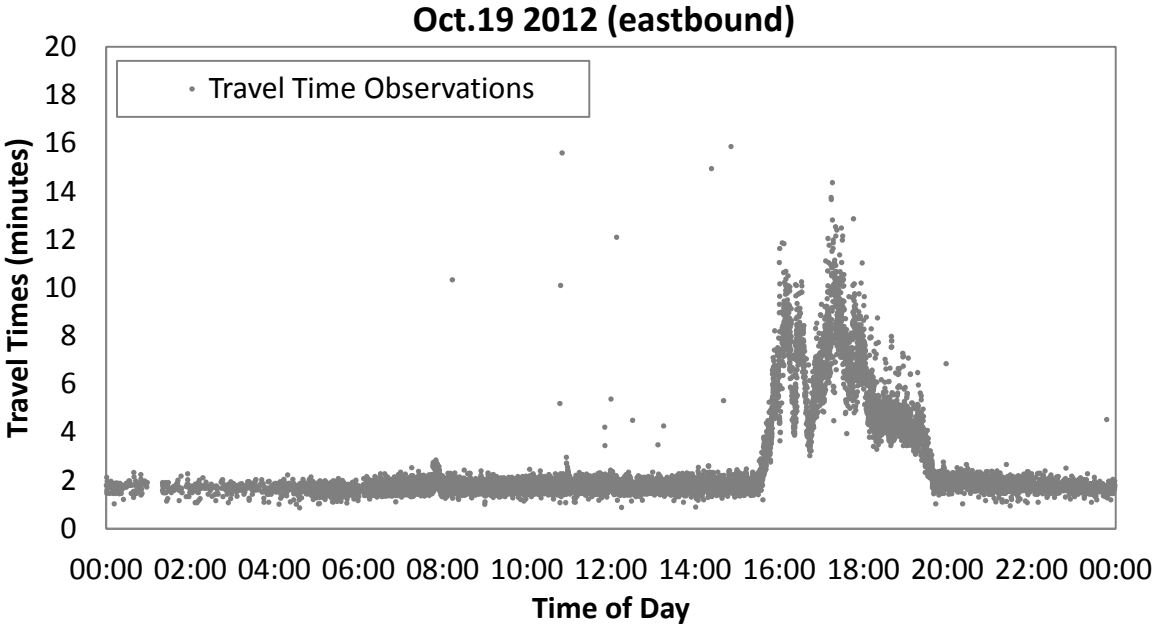
Appendix A

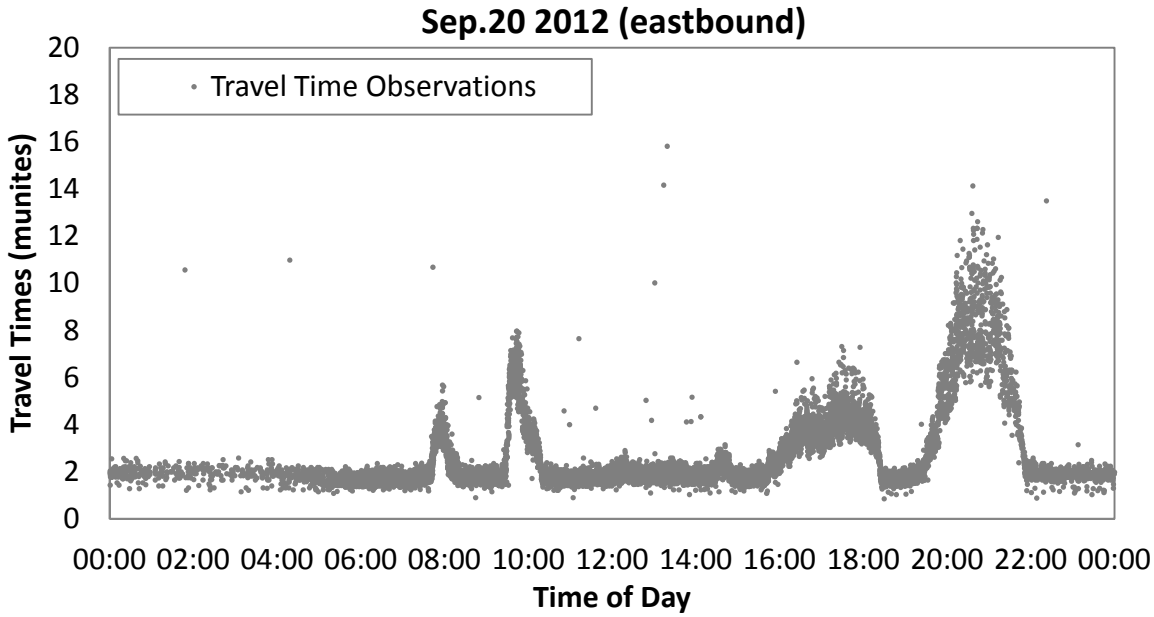
Travel time observations of the tested datasets

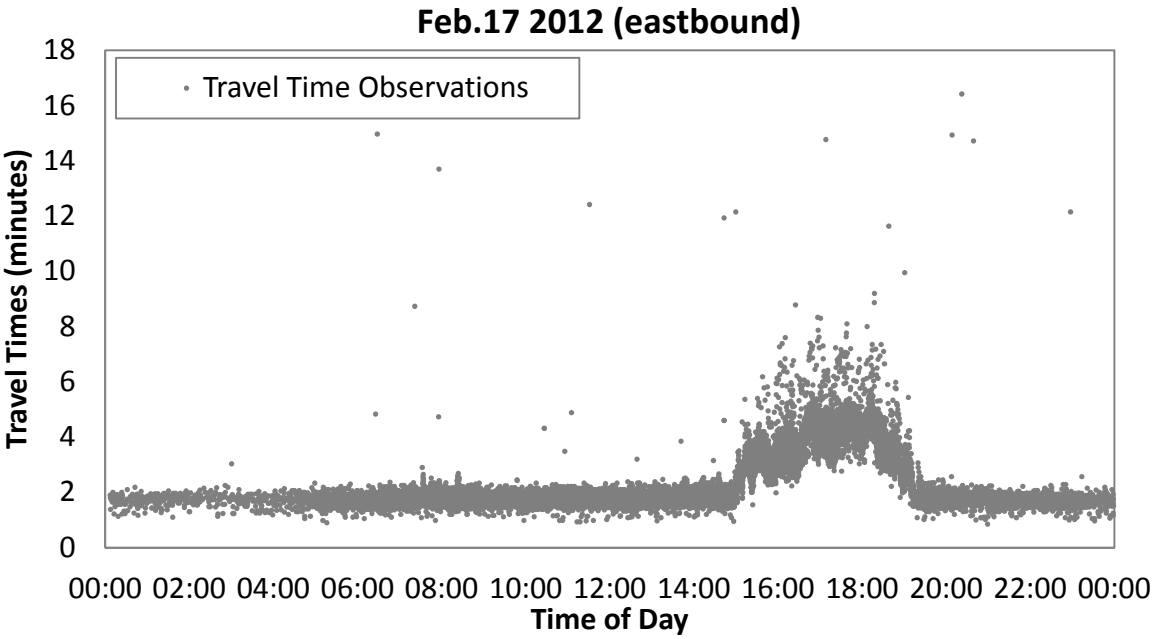


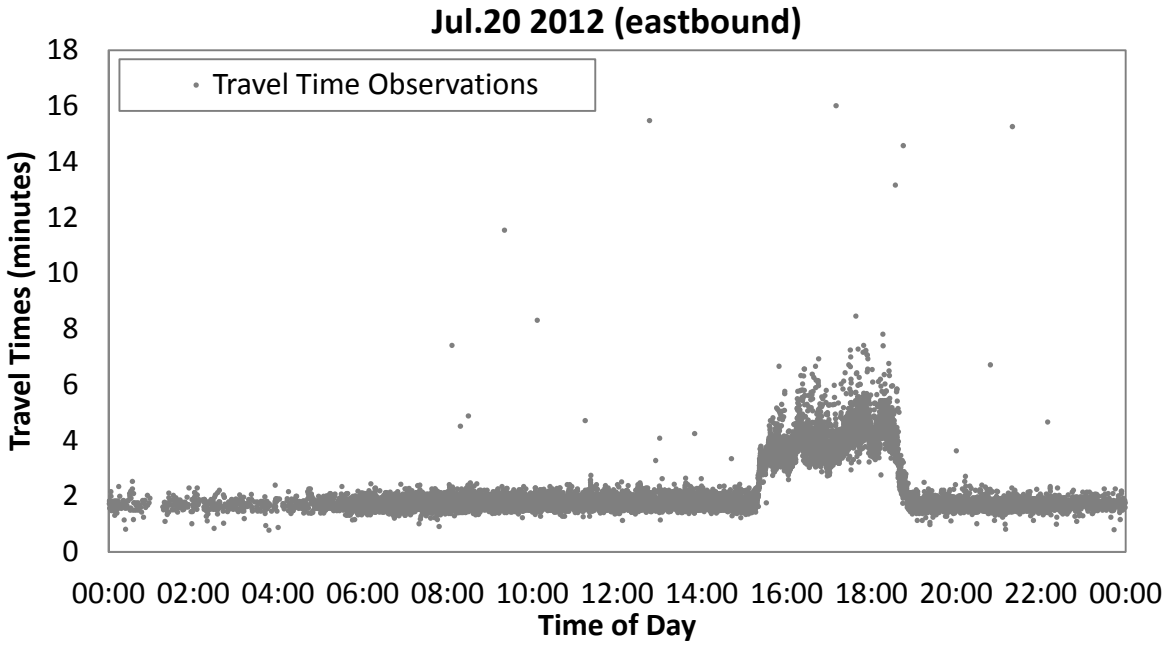


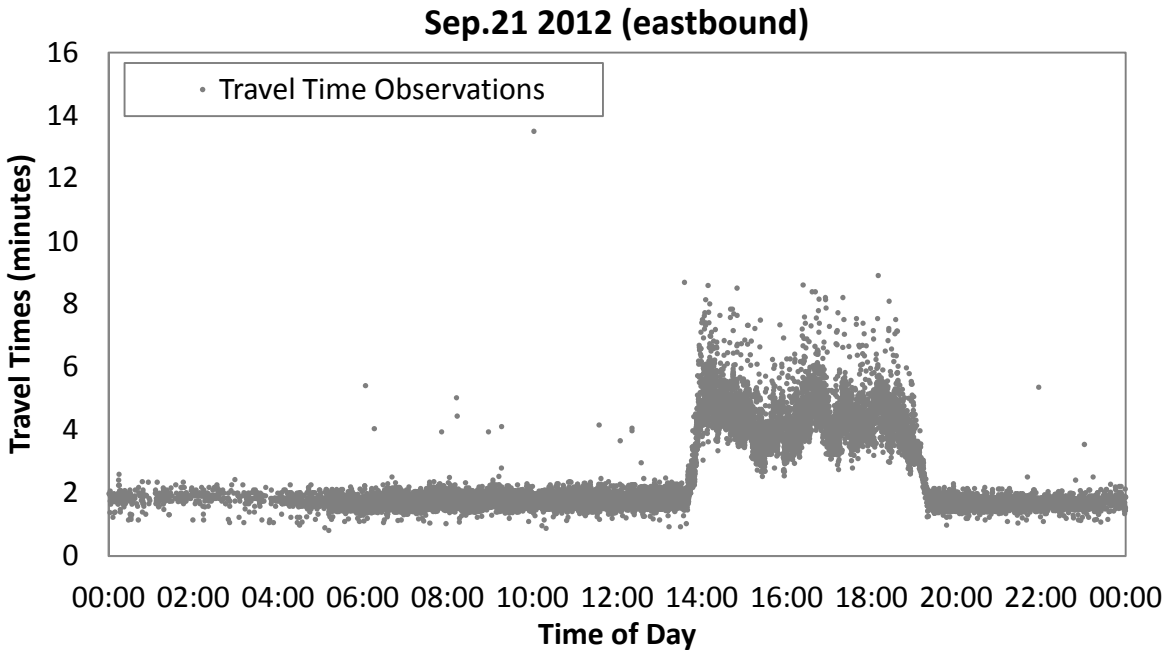

Jun.15 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
20
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

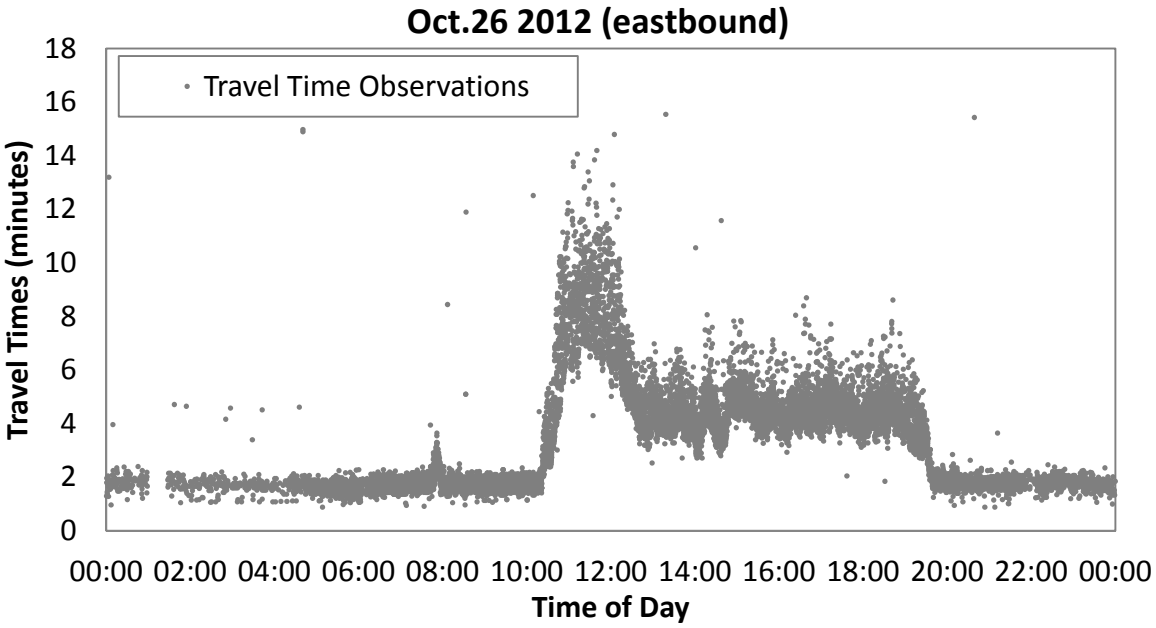

Oct.19 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
20
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

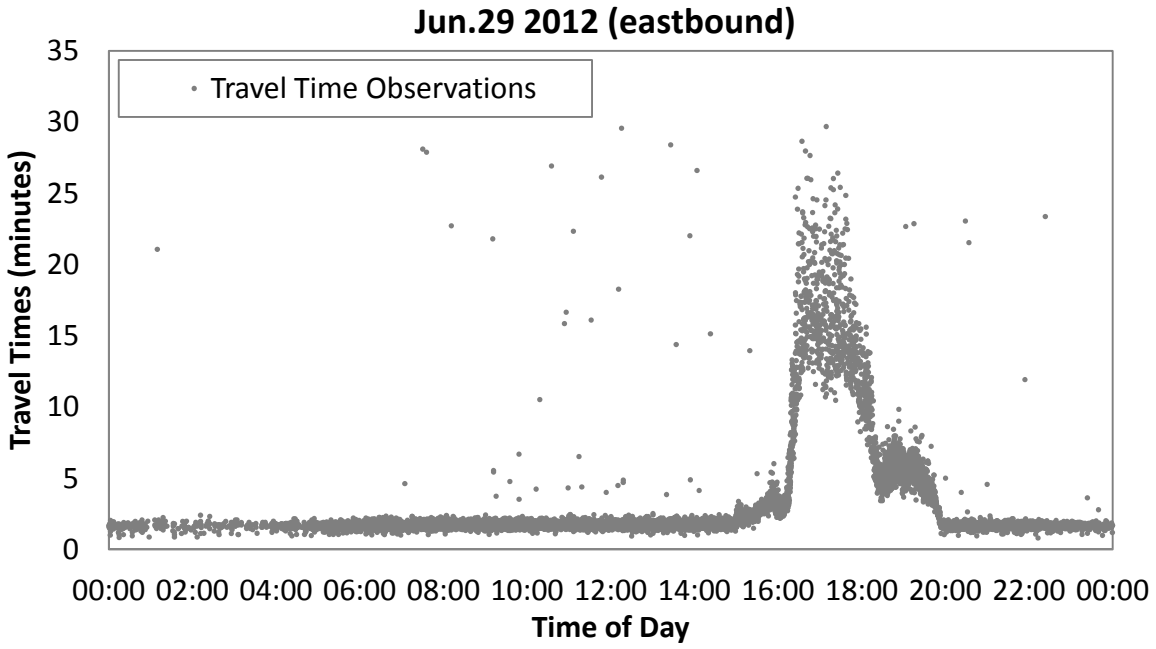

Sep.20 2012 (eastbound)
Travel Time Observations
Travel Times (munites)
20
18
16
14
12
10
8
6
4
2
0
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

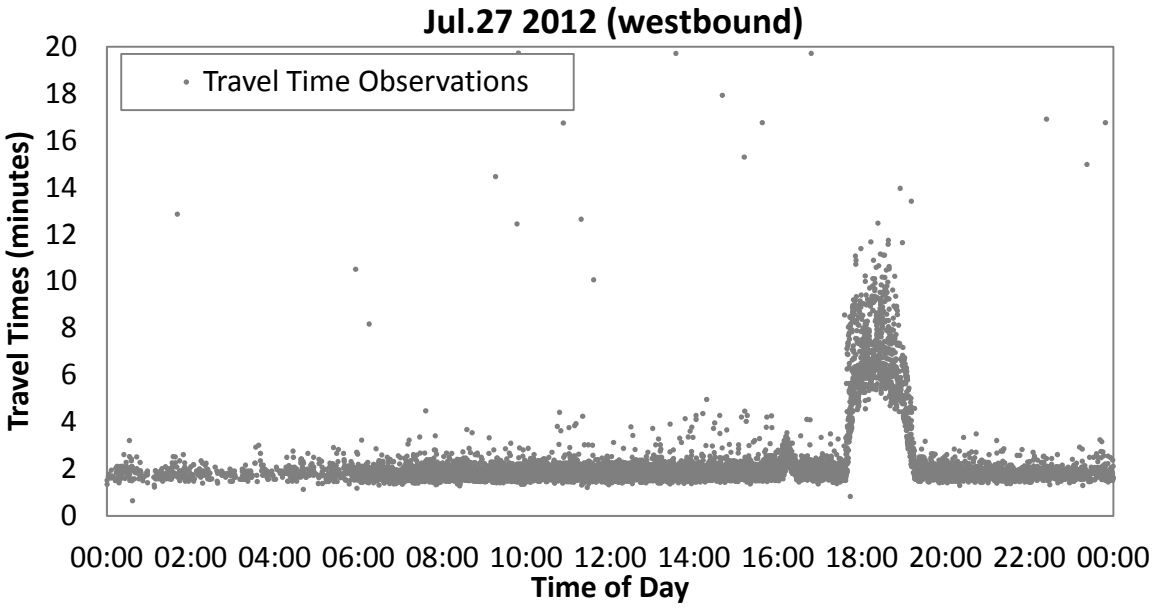

Feb.17 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
18
16
14
12
10
8
6
4
2
0
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

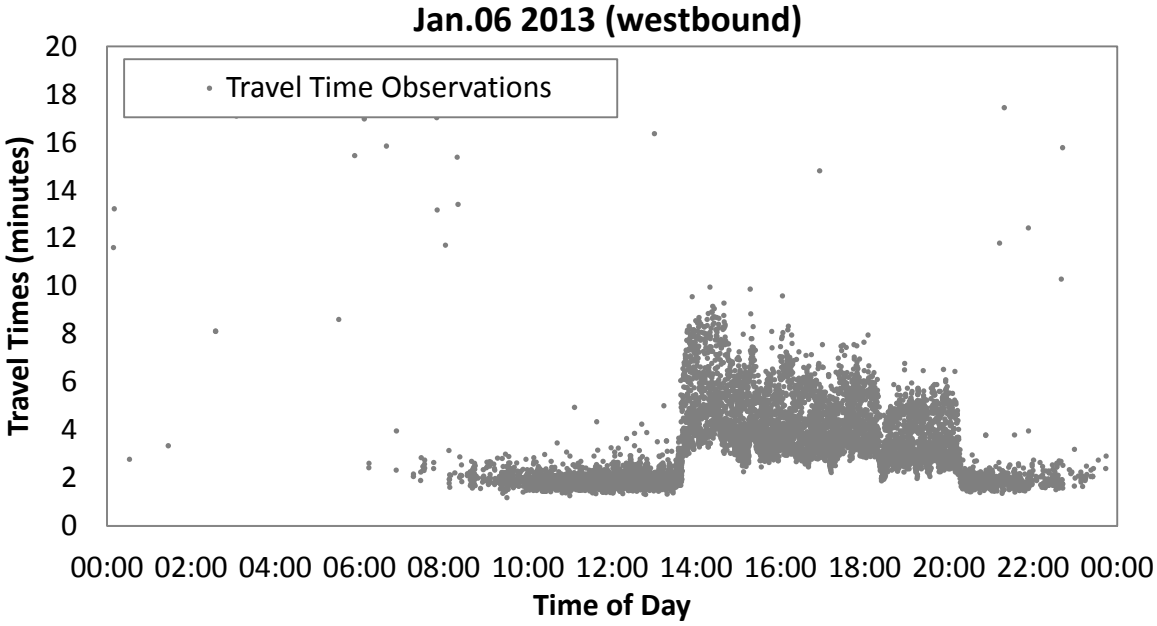

Jul.20 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

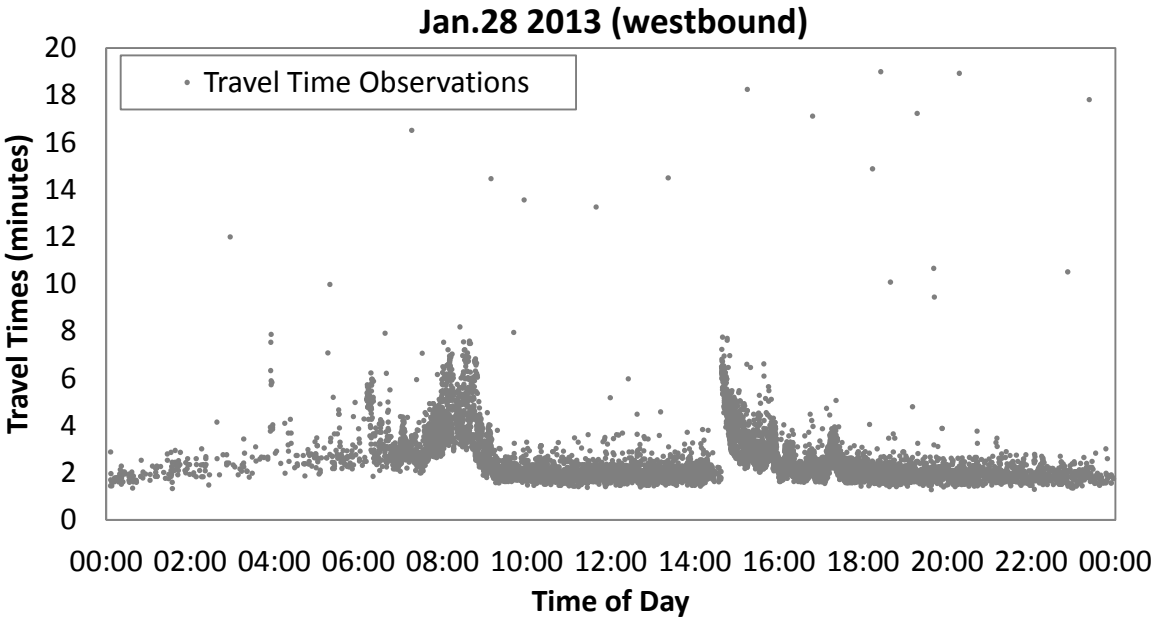

Sep.21 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

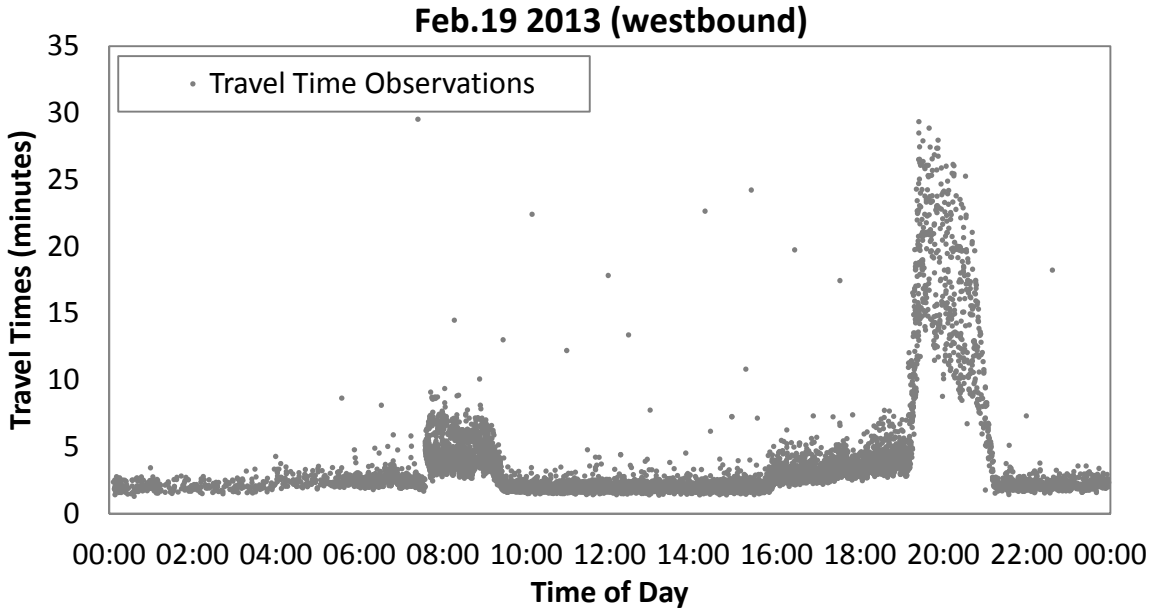

Oct.26 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

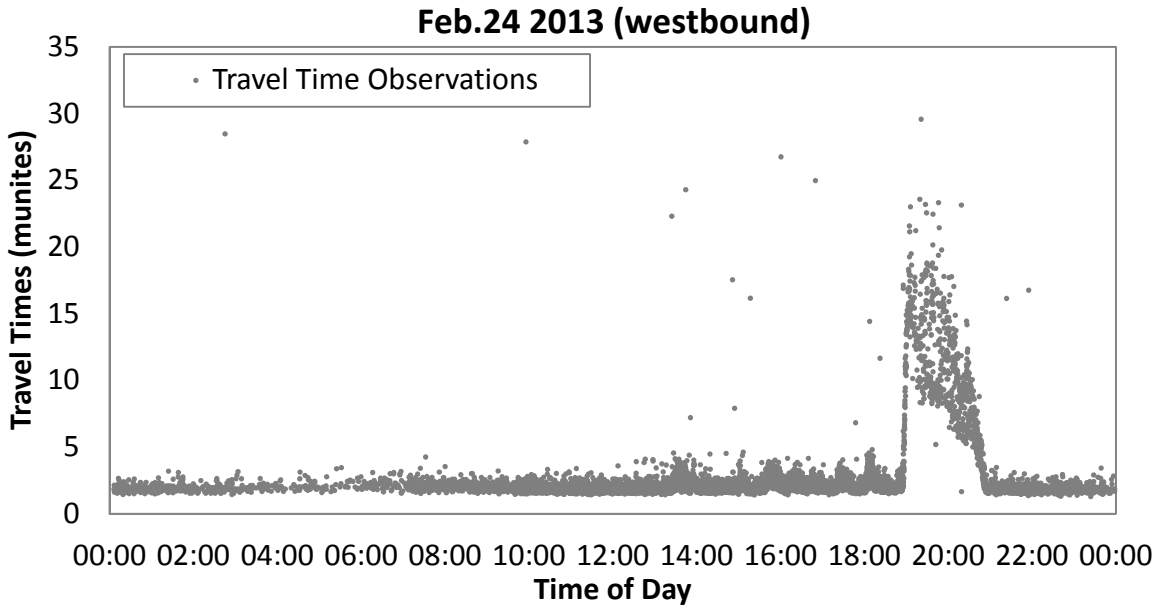

Jun.29 2012 (eastbound)
Travel Time Observations
Travel Times (minutes)
0
5
10
15
20
25
30
35
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day


Jul.27 2012 (westbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
20
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day


Jan.06 2013 (westbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
20
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day


Jan.28 2013 (westbound)
Travel Time Observations
Travel Times (minutes)
0
2
4
6
8
10
12
14
16
18
20
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day


Feb.19 2013 (westbound)
Travel Time Observations
Travel Times (minutes)
0
5
10
15
20
25
30
35
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day


Feb.24 2013 (westbound)
Travel Time Observations
Travel Times (munites)
35
30
25
20
15
10
5
0
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

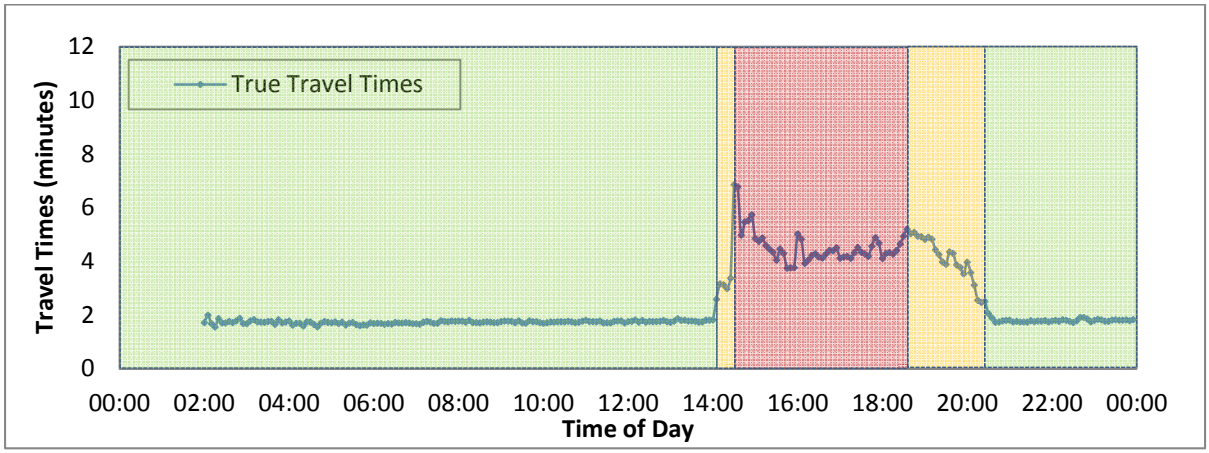
Appendix B

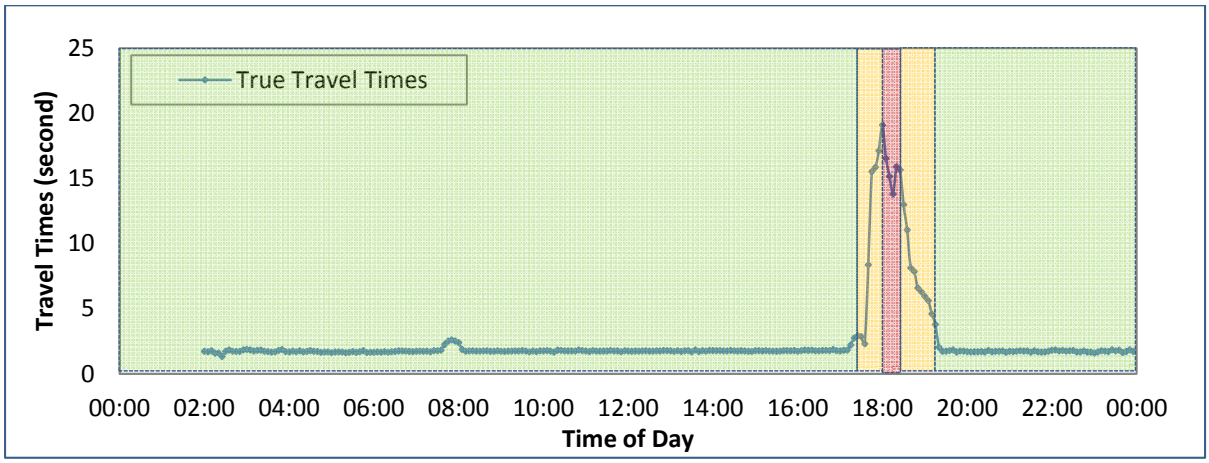
Determination of the traffic states for the tested datasets

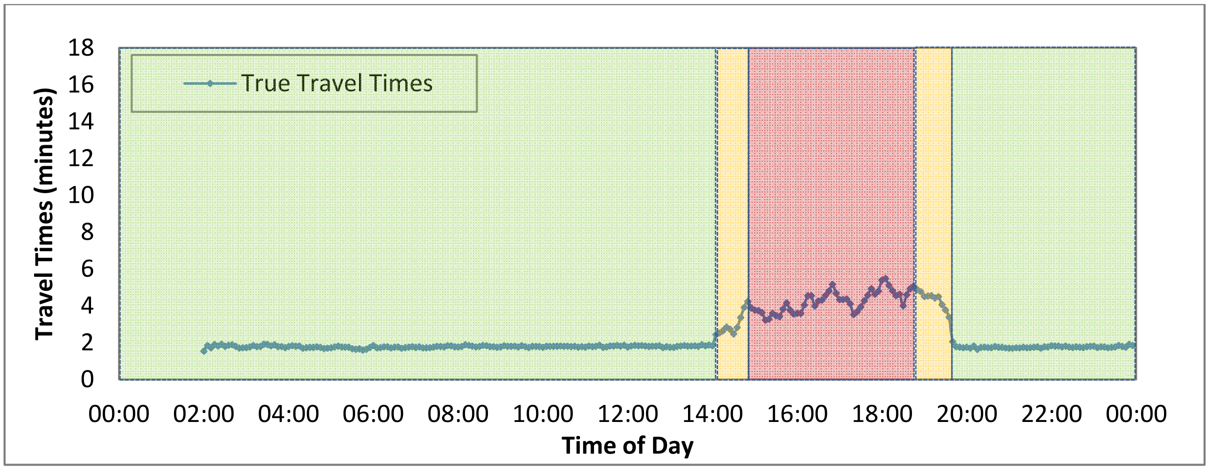
Free Flow

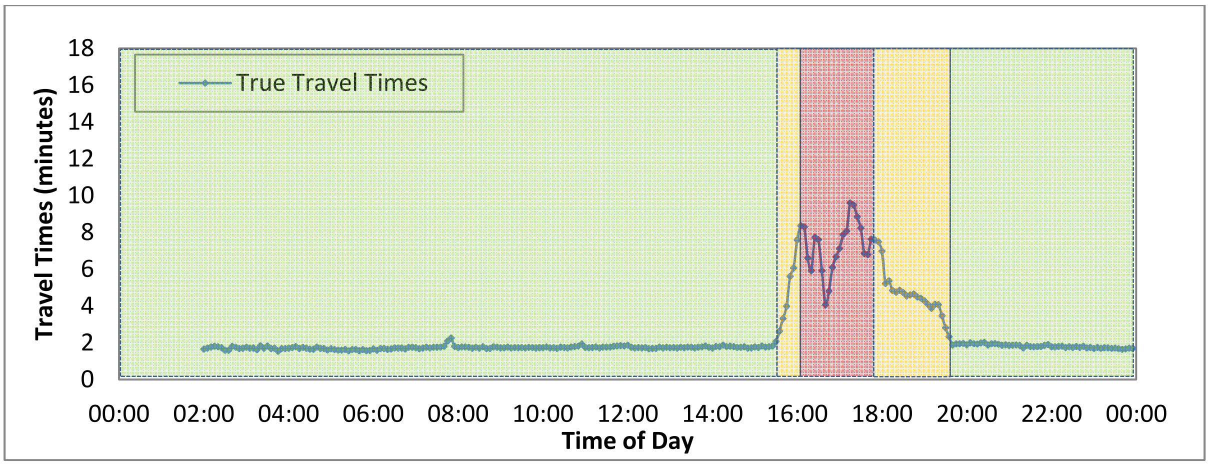
Transition

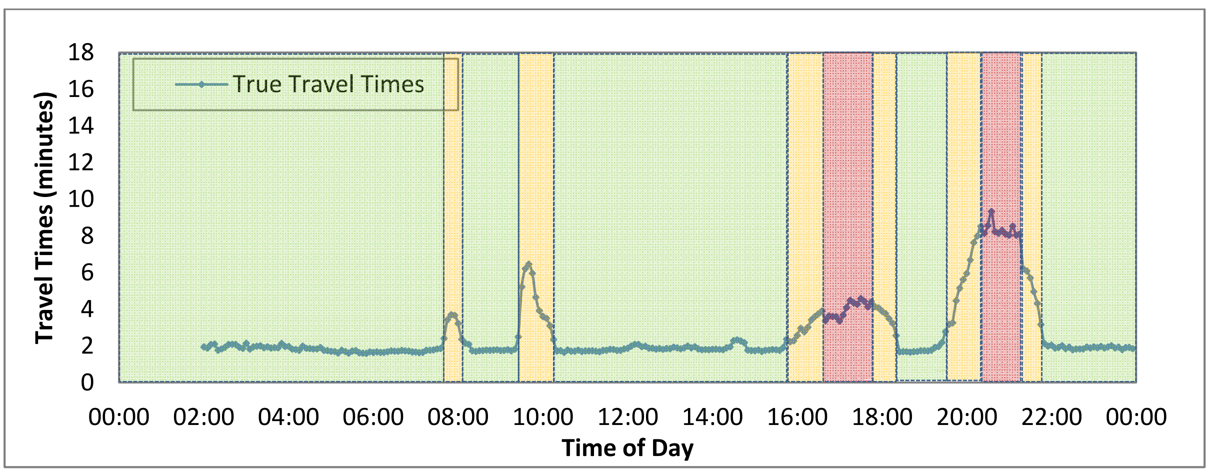
Congestion





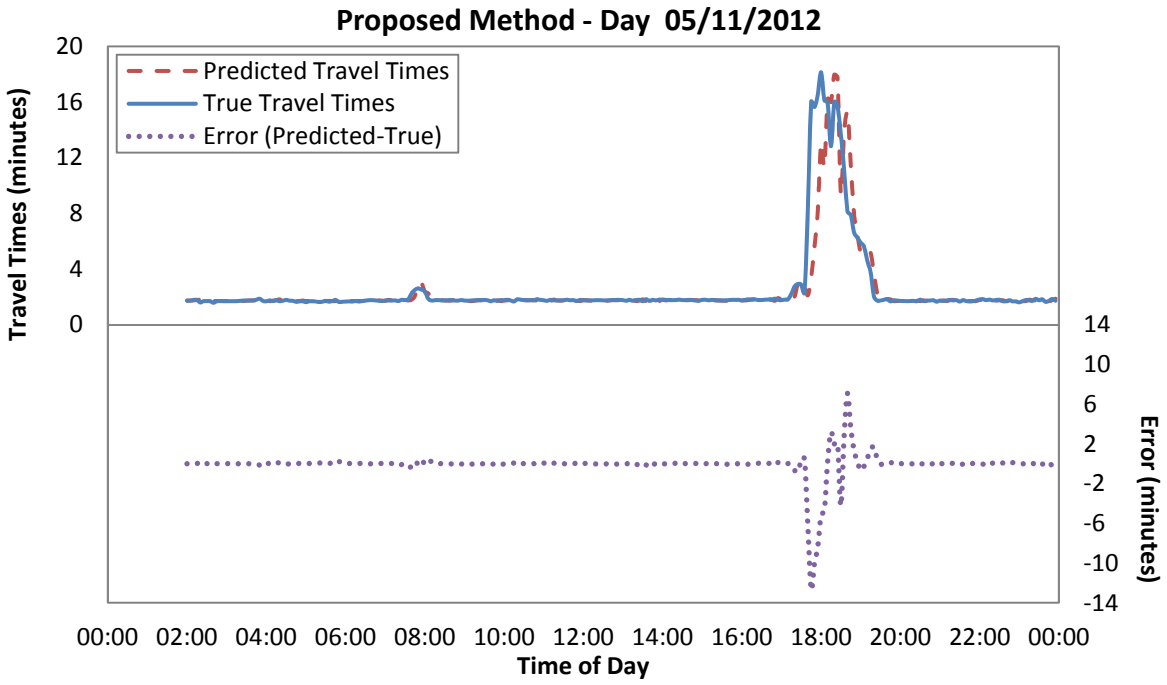

True Travel Times
Travel Times (minutes)
18
16
14
12
10
8
6
4
2
0
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day


True Travel Times
Travel Times (minutes)
18
16
14
12
10
8
6
4
2
0
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

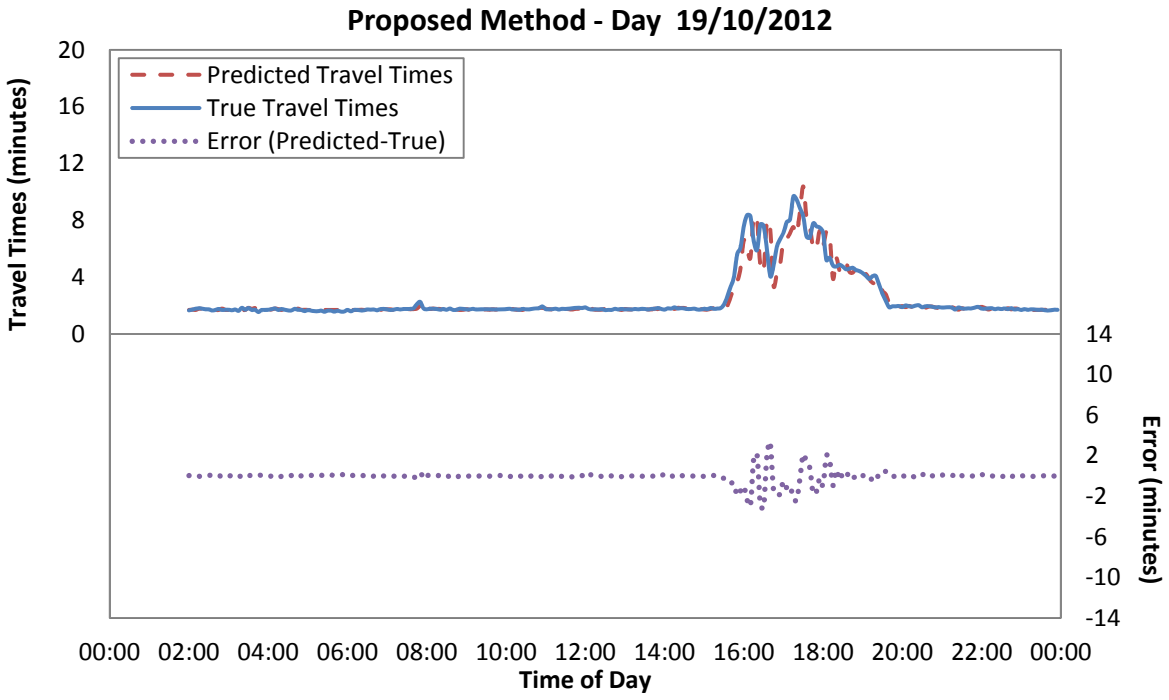

True Travel Times
Travel Times (minutes)
18
16
14
12
10
8
6
4
2
0
00:00 02:00 04:00 06:00 08:00 10:00 12:00 14:00 16:00 18:00 20:00 22:00 00:00
Time of Day

Appendix C

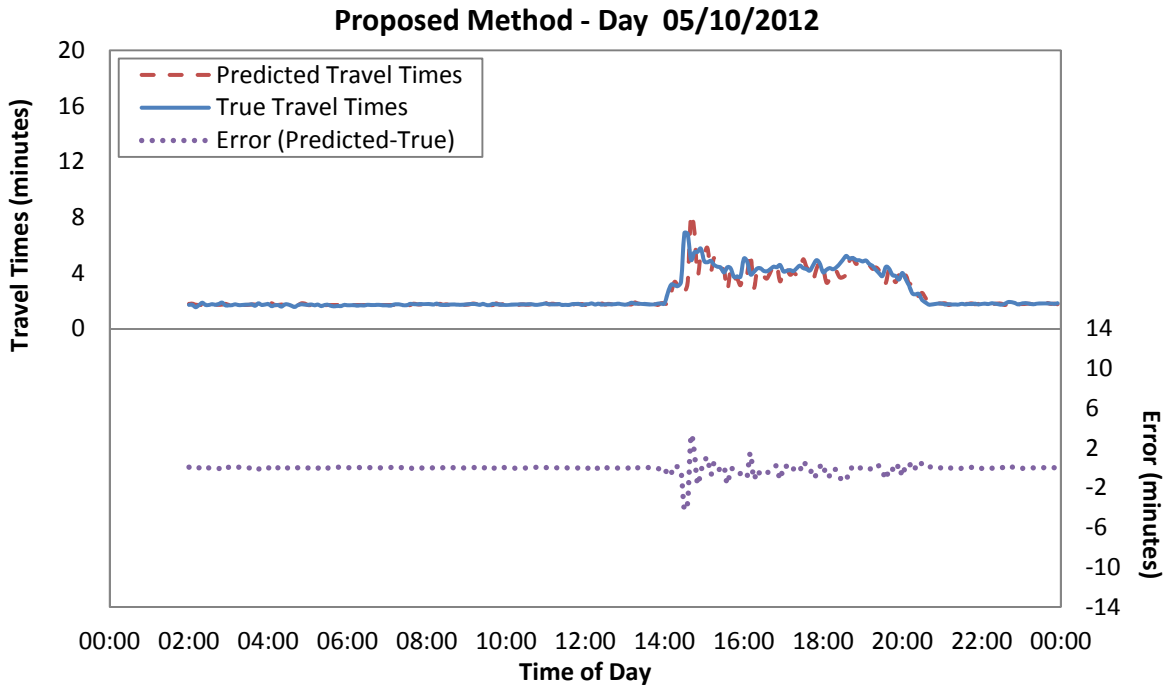
The application results of the proposed model



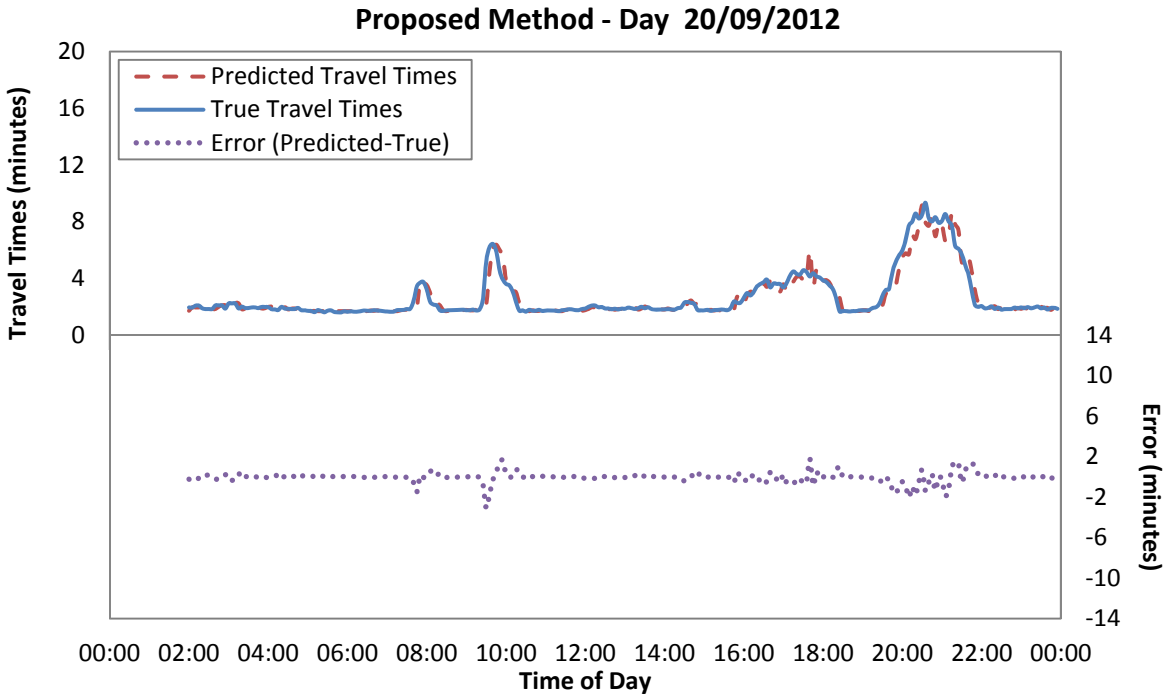
Results of the proposed model applied to data collected at Nov. 11 2012 (eastbound)



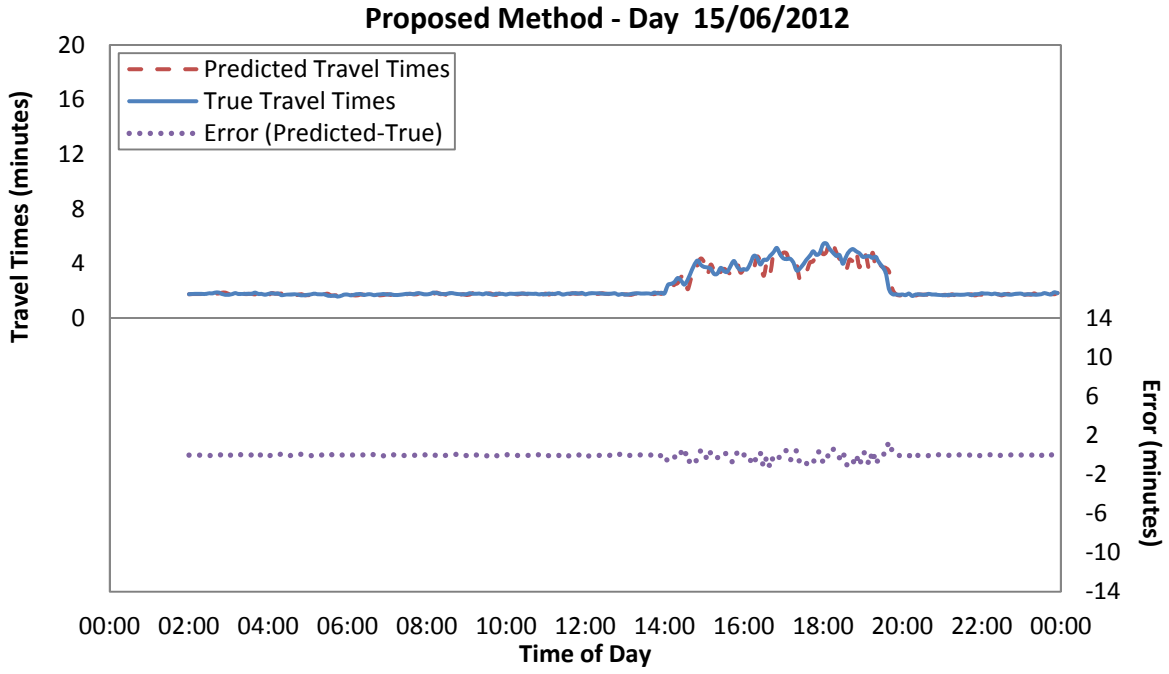
Results of the proposed model applied to data collected at Oct. 19 2012 (eastbound)



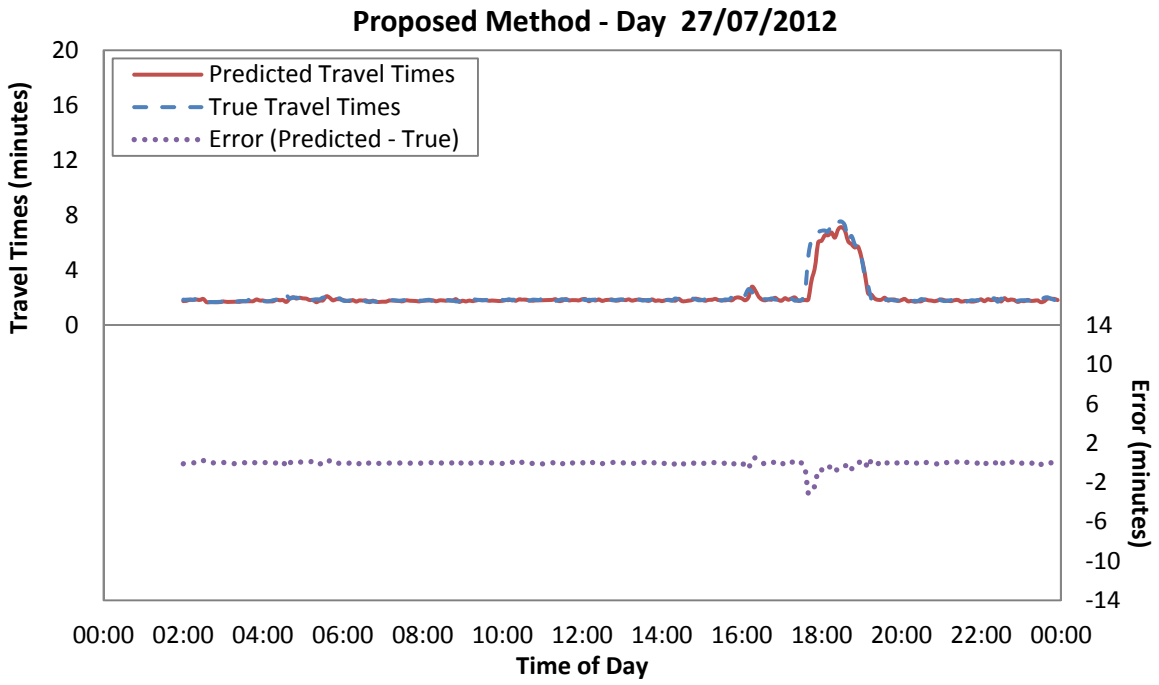
Results of the proposed model applied to data collected at Oct. 05 2012 (eastbound)



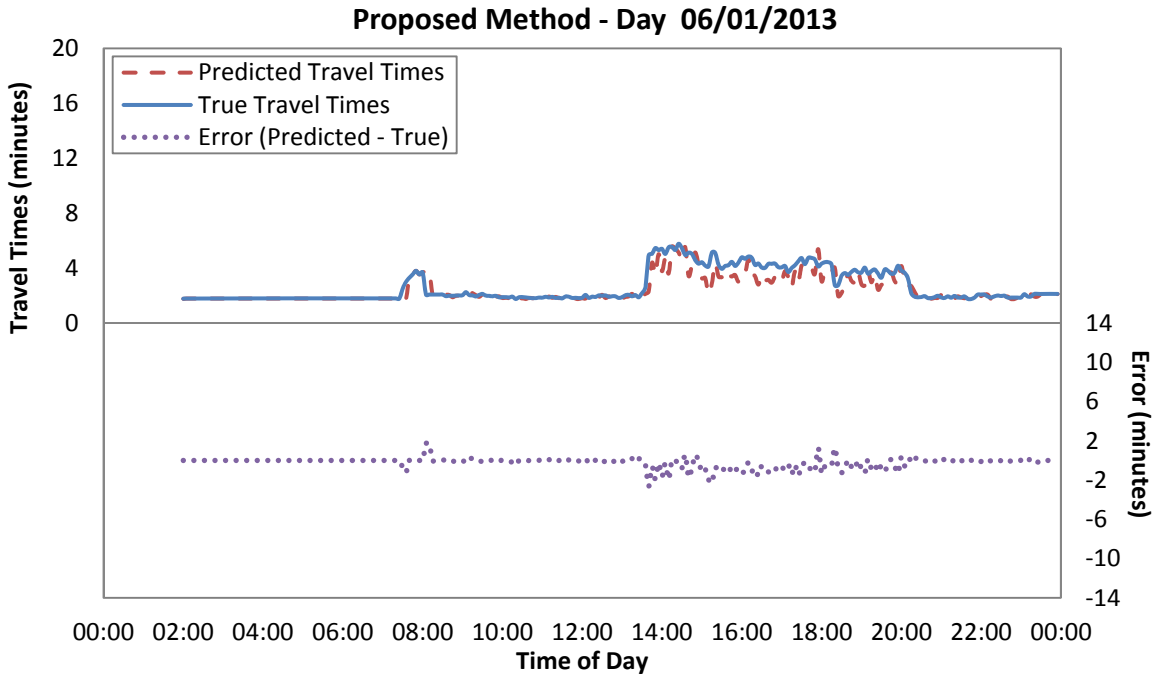
Results of the proposed model applied to data collected at Sep. 20 2012 (eastbound)



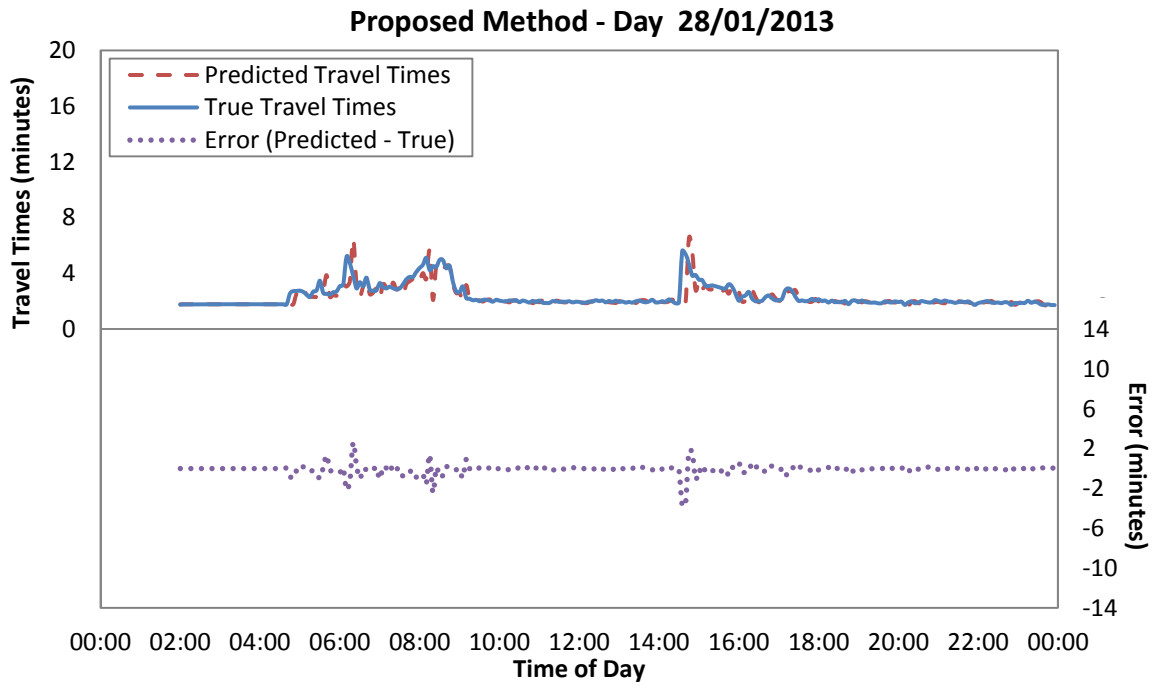
Results of the proposed model applied to data collected at Jun. 15 2012 (eastbound)



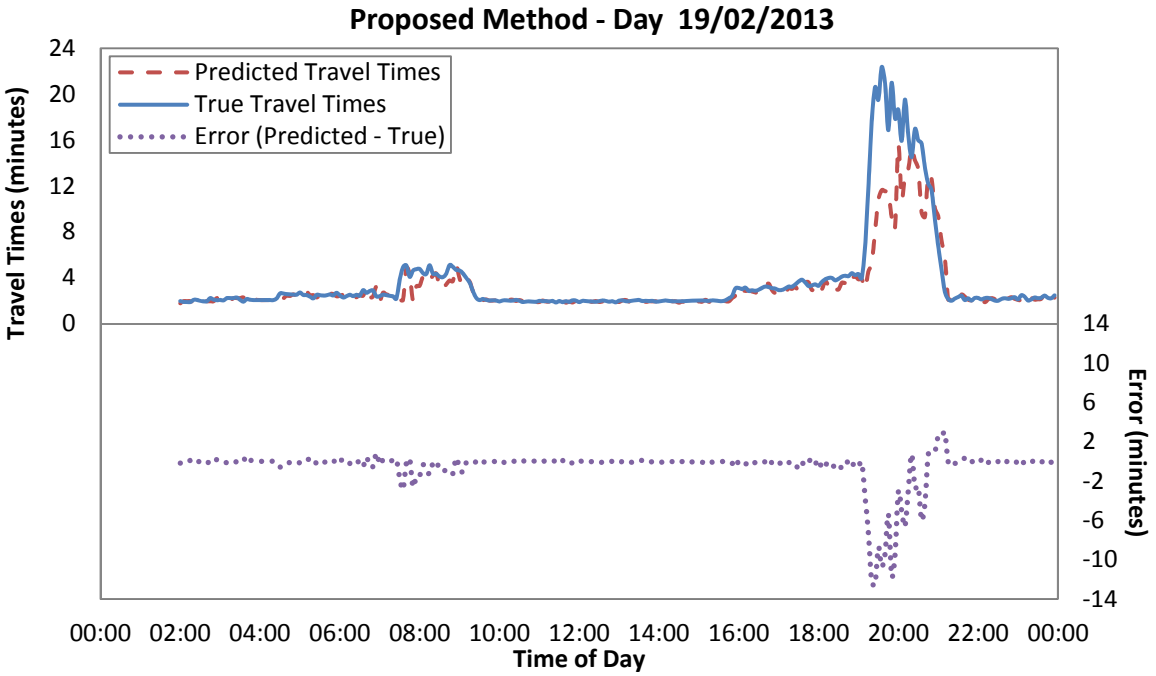
Results of the proposed model applied to data collected at Jul. 27 2012 (westbound)



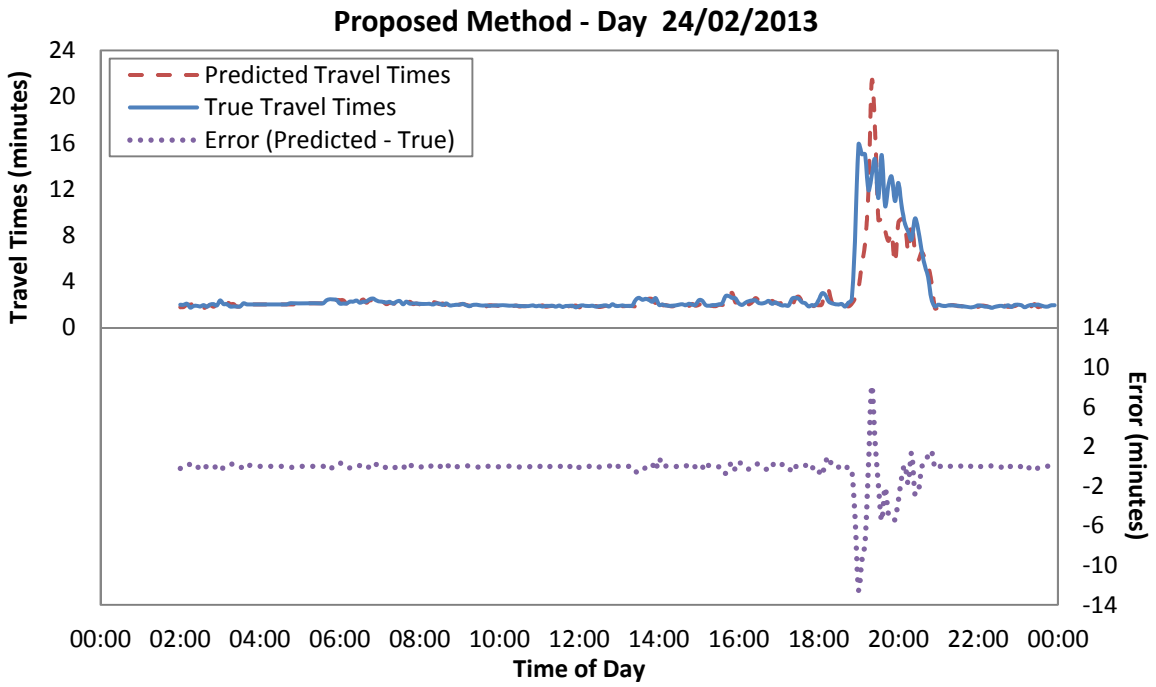
Results of the proposed model applied to data collected at Jan. 06 2013 (westbound)



Results of the proposed model applied to data collected at Jan. 28 2013 (westbound)



Results of the proposed model applied to data collected at Feb. 19 2013 (westbound)

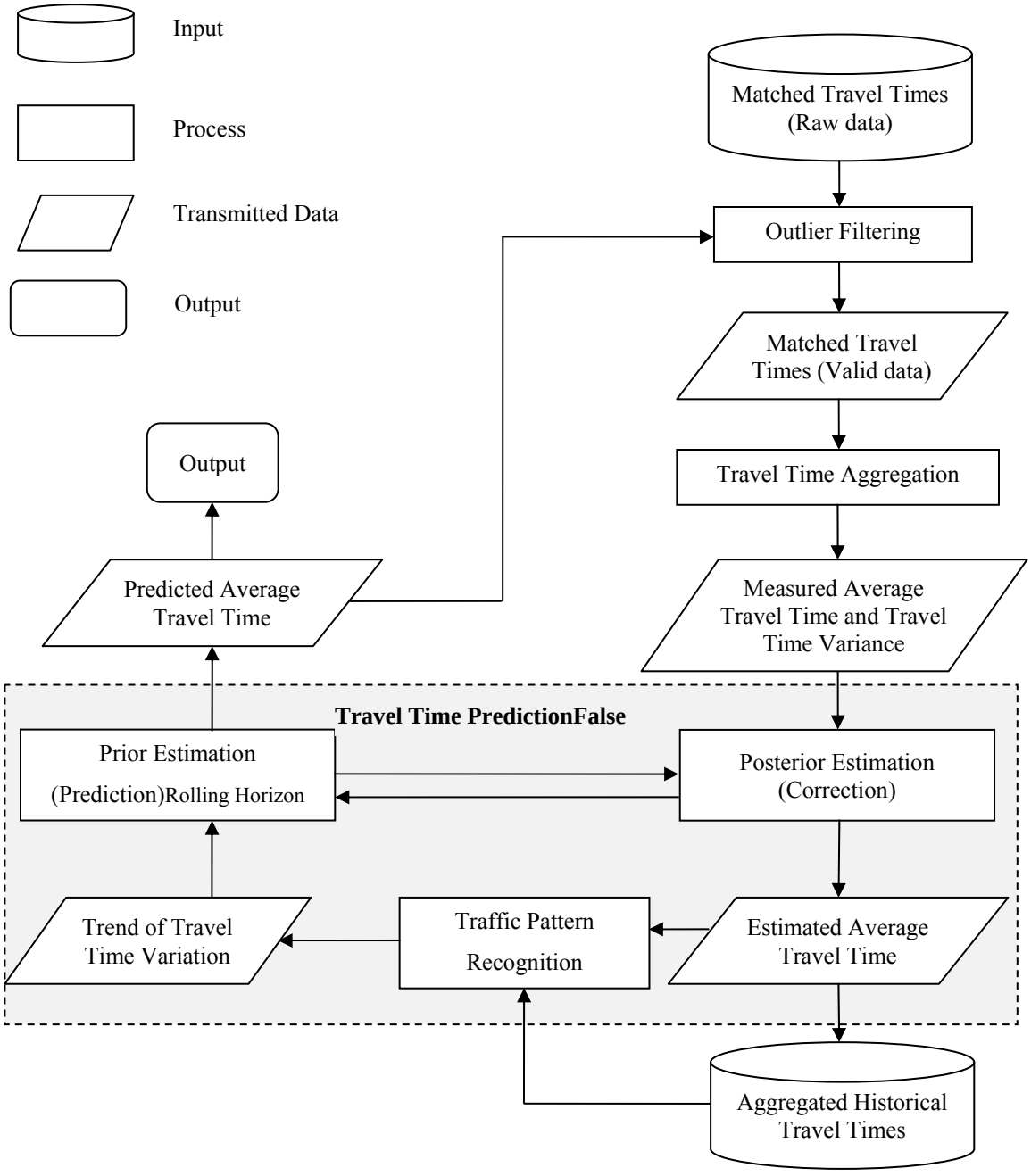


Results of the proposed model applied to data collected at Feb. 24 2013 (westbound)

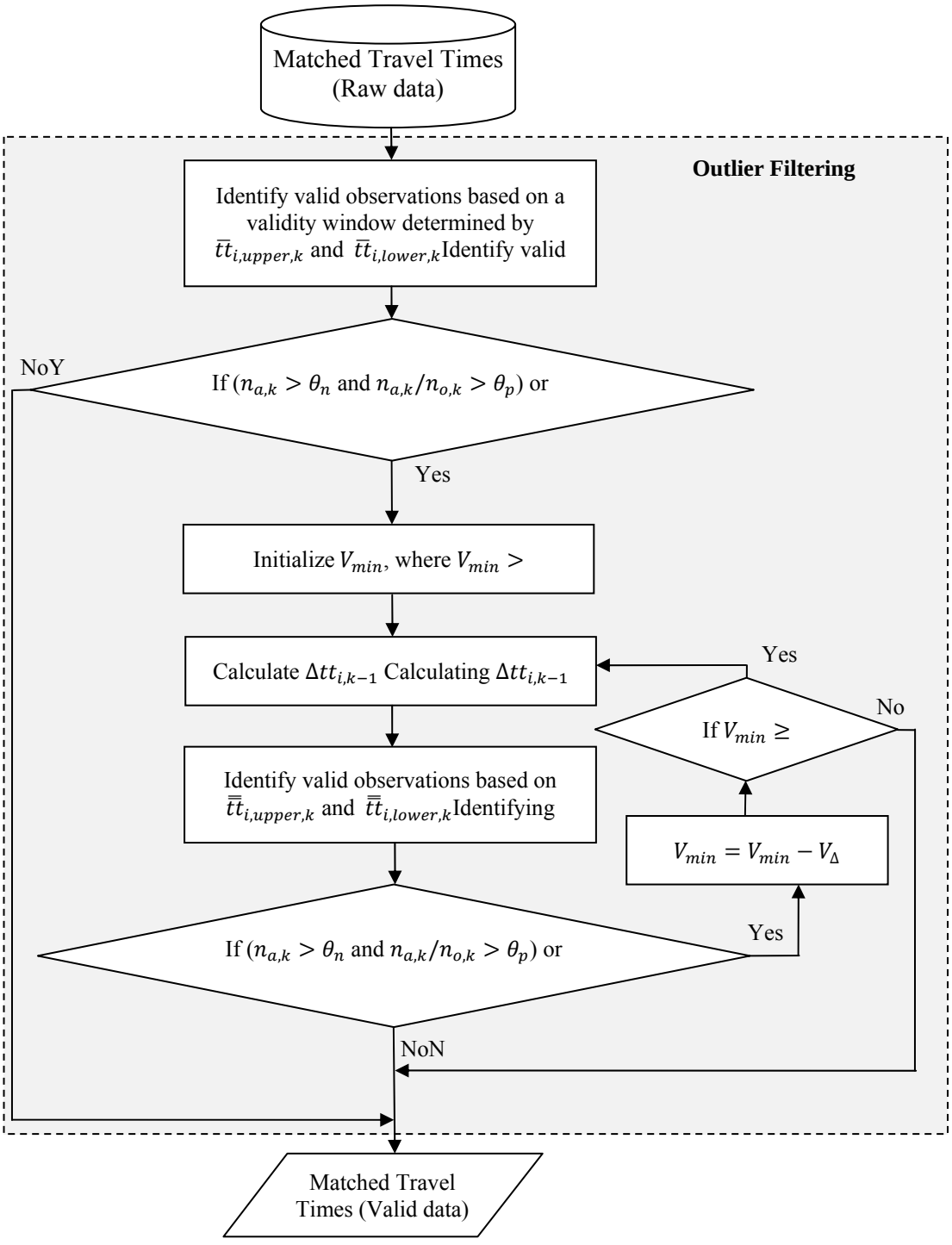
Appendix D

Flow charts for method implementation

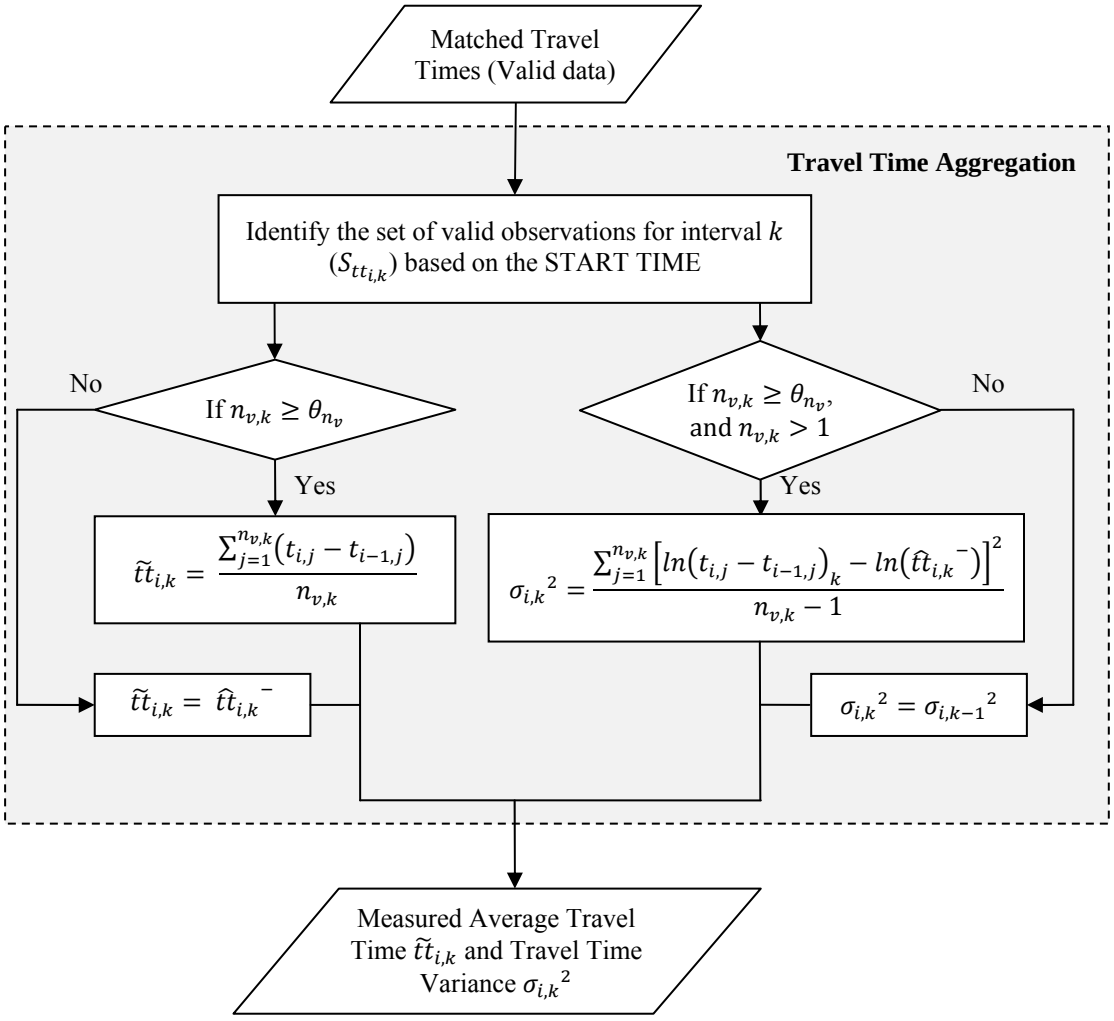
General flow chart



Sub flow chart for outlier filtering



Sub flow chart for travel time aggregation



Sub flow chart for travel time prediction