Truck GPS Data in Freight Planning:
Methodologies and Applications for Measurement and Forecasting

Zun Wang

A dissertation submitted in particular fulfillment of the

requirement for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Anne Goodchild, Chair

Edward McCormack

Mark Hallenbeck

Program Authorized to Offer Degree:

Department of Civil and Environmental Engineering

University of Washington

Abstract

Truck GPS Data in Freight Planning:

Methodologies and Applications for Measurement and Forecasting

Zun Wang

Chair of the Supervisory Committee

Associate Professor Anne Goodchild

Department of Civil and Environmental Engineering

ABSTRACT

Efficient and reliable goods movement via our nation's highway system is critical to the nation's economy and quality of life. Truck mobility is one of the key performance measures for evaluating the conditions of goods movement and supporting freight planning. Truck GPS data can be useful in developing truck mobility measures and providing insights into freight planning.

This dissertation employs truck GPS data and proposes a set of methodologies for measuring and forecasting truck mobility performance, with particular emphases on truck travel time and travel time reliability. It also examines how GPS data can be used to support freight planning, using the analysis of impacts of a tolling project on truck mobility and routing as a case study.

The first part of this dissertation investigates how to measure truck travel time reliability given the characteristics of GPS data. An improved spot-speed distribution based travel time reliability measure is proposed. The proposed approach is compared with a number of commonly applied reliability measures. The correlations among these measures reveal that the reliability measures are not highly correlated, demonstrating that different measures provide different conclusions for the same underlying data and traffic conditions. The author presents recommendations of the appropriate measures for different applications.

Quantitative freight project prioritization processes require both pre- and post-investment truck mobility performance. Therefore, the second part of this dissertation develops quantitative methods for forecasting truck specific travel time and travel time reliability. For travel time prediction, a speed-density based approach is proposed to predict truck travel time associated with segment density changes. Traffic regimes are segmented using a cluster analysis approach. The travel time estimates are compared with two widely applied traditional methodologies. The results demonstrate that the proposed method is able to estimate more accurate travel times. For reliability prediction, we analyze the changes of GPS spot speed distribution in response to different traffic conditions. A relationship between truck spot speed distribution coefficient of variation and segment density is proposed to forecast reliability. The approach is transferrable and sheds a light on forecasting travel time reliability.

The third part of this dissertation focuses on examining how GPS data can be used to assist freight planning. The SR-520 toll bridge in the City of Seattle, Washington is selected as the case study. We quantify the toll project impacts on truck mobility and route choice. Truck GPS data is used to evaluate route choice and travel speed along SR-520 and the alternate toll-free route I-90. A logit model is developed to determine the influential factors in truck routing. The results indicate that travel time, travel time reliability and toll rate are all influential factors during both peak and off-peak periods. The values of truck travel time during different time periods are estimated, and the values vary with the definition of peak and off-peak periods.

This dissertation provides decision makers with useful guidance and information on using GPS data for truck mobility measurement and forecasting. It also demonstrates the capability of GPS data in supporting freight planning.

Contents

List of Tables

List of Figures

ACKNOWLEDEGMENTS

I would like to thank my PhD advisor, Professor Anne Goodchild, for supporting and encouraging me during the past three years. I appreciate all her guidance and contributions of time and ideas to make my graduate school experience a productive, thoughtful and rewarding journey. She is my role model for a mentor, researcher and professional. I would also like to thank my dissertation committee of Edward McCormack and Mark Hallenbeck, for their support and guidance in the past two years as I moved from an idea to a completed study.

I also thank all members of my research team-the Goods Movement Collaborative. I am lucky to be a part of the team. Their friendship, help and support have been continuous fuel during the journal in finishing my Ph.D. The fun and laughs we had together are invaluable memories of mine.

I especially thank my loving parents. I am so grateful for everything that they have taught me and given me. I love them so much and I would not have made it this far without them. I cannot give enough thanks to them for all their advice, support and unconditional love.

Finally, I thank my best friend, soul-mate and husband Haijun. He is always supportive and has a faith in me during my good and bad times. I truly thank him for his patience, encouragement and tremendous love. I feel lucky that I married the best person out there for me.

Chapter 1 INTRODUCTION

Efficient and reliable goods movement via our nation's highway system is critical to the nation's economy and quality of life. Truck mobility measures, defined as travel time and travel time reliability in this dissertation, are designed as a tool to evaluate the efficiency and reliability of freight goods movement. Information derived from mobility measures can support freight planning, including identifying bottlenecks, prioritizing projects, and assessing project impacts. The truck mobility measurements rely upon truck-specific movement data. This dissertation employs truck GPS data to measure and forecast truck mobility and support freight planning. Chapter 1 presents the research motivations, background, research questions, and dissertation organization.

1.1 Research Motivations

Trucking industry plays a significant role in the U.S. economy. It employed 9 million people and generated $659 million in revenue, which represented 5% of the U.S. Gross Domestic Product in 2007(American Trucking Association 2008). In 2011, the U.S. highway system moved 17.6 billion tons of goods worth $10 trillion. Meanwhile, the freight highway demand is projected to grow dramatically by 66% to be worth $21.5 trillion by 2040 (FHWA 2012a). However, it is unlikely the transportation supply side can keep up with the same growth without improvement to freight goods movement networks. The freight highway system's economic significance to the United States and the needs for sustaining the highway system both have been gradually recognized by government and private sectors (NCFRP 2011). The Moving Ahead for Progress in the 21st Century (MAP-21) act signed into law by President Obama in June 2012 emphasizes funding freight-related projects which are able to improve freight movement efficiency and economic vitality (FHWA 2012b). Despite the efforts that have been made to address future transportation needs, the report published by The National Commission suggested that the nation had spent only 40% of what is needed to sustain and improve the highway network (The National Commission 2008).

The increasing demand of the freight highway system and the limited federal and state level transportation budgets require truck performance measures for monitoring the condition of the system and identifying needs for future improvement. In this light, the MAP-21 act requests that all national and state roads be gauged by performance measures. In addition, it encourages all State Freight Plans to include performance measures that will guide the freight-related transportation investment decisions of the state (FHWA 2012b, USDOT 2012).

For the past two decades we have developed increasingly sophisticated approaches for understanding the transportation system performance. Although a number of states have developed performance measure frameworks for monitoring general traffic, only a few of them have truck-specific performance measures. None has the performance measure-based freight planning strategies for freight project prioritization and impact assessment. Among the existing truck performance measures, most agencies only collect or estimate truck traffic volume and truck safety data, without providing truck mobility measures, e.g., speed, travel time and travel time reliability. However, developing truck mobility performance measures is important and necessary as it evaluates current system conditions and identifies problems. Furthermore, the performance measure-based planning is based upon real world data and, therefore, is able to provide a means to a more efficient investment of transportation funds and monitor project efficiency and effectiveness (NCFRP 2011).

The feasibility of using truck GPS data in freight performance measure has been investigated in several studies (Battelle 1999, McCormack and Hallenbeck 2006, FHWA 2003). It is found that GPS data provides reliable truck location and movement information and has potential to support performance measurement and forecasting. However, there are limited studies applying truck GPS data to measure and forecast truck mobility and assist freight planning. Thus this dissertation aims to take advantage of this new data source and investigate how truck GPS data can be used to support truck mobility measurement, forecasting, and freight planning.

1.2 Research Background

1.2.1 Access to Truck GPS Data for Freight Mobility Measurements and Planning

Despite the fact that many truck performance-measure applications have been implemented, most agencies are only able to provide truck volume and safety statistics based on real world observations. They are unable to provide truck mobility measures (e.g., truck travel time and travel time reliability) which are based on expensive field data that is labor-intensive to collect. One of the major challenges faced by transportation agencies is the insufficiency of truck-specific movement data (McCormack and Hallenbeck 2006, Beagan 2007). The traditional means for collecting such information involves stopping trucks and interviewing drivers or giving them a questionnaire. Such methods may not be able to get accurate information and were conducted, at most, once a decade (NCFRP 2011).

The GPS technology has been widely applied in the trucking industry for fleet management since the 1980s, and both shippers and carriers have benefited from its use (Roetting 2003, Baumgartner 2008). Several pilot studies had proved the capability of commercial truck GPS technology in collecting truck movement data for supporting freight mobility research. Greaves and Figliozzi (2008) analyzed the issues and potential applications of collecting commercial truck tour data using GPS technology. The study was based on the commercial vehicle data collection study in Melbourne in 2006. It is found that the capacity of GPS data to provide reliable and detailed time-space information in an economical manner has many potential applications in the transportation field, including constructing origin-destination matrices and developing speed-time profiles. McCormack and Hallenbeck (2005, 2006) conducted tests which involved installing GPS devices on commercial trucks. The results also indicated that it is possible to use GPS devices to collect truck movement data and assess truck trip reliability and identify truck route choices. However, the results also suggested that it is difficult to recruit drivers and install GPS devices due to privacy concerns.

This data paucity issue has been gradually solved as the FHWA started collaborating with the American Transportation Research Institute (ATRI) to collect commercial truck GPS data. They also investigated how data gathering from the GPS devices installed in trucks can be used to measure the mobility along interstate highways since 2002 (Jones et al. 2005). The data was also shared with several research agencies for analysis. In July 2013, FHWA also released the National Performance Measurement Research Data Set (NPMRDS), which is an aggregated travel time dataset in five-minute intervals for both passenger and commercial vehicles. State DOTs and their contractors have access to the data for interstate highway performance measures (ATRI 2012). In addition, some researchers purchased GPS data from GPS vendors or fleet management firms to study truck behaviors (McCormack et al. 2010, Sharman and Roorda 2011).

The GPS data acquisition efforts on which this dissertation is based was a pilot study of investigating how to gather and use existing GPS data collected for truck fleet management to develop performance measures for trucks (McCormack et al. 2010). The GPS data was initially collected by GPS vendors for trucking companies' fleet management. Data used in this dissertation was purchased from GPS vendors directly instead of from trucking companies. Each vendor collects data for many trucking companies. The data includes a unique device ID, vehicle trajectory, spot speed, heading, location, time and date (McCormack et al. 2010, Ma et al 2011, Zhao et al. 2012). Each device ID was scrambled for anonymity to protect the customer information. Data was collected since September 2008 until the present. Originally, there were approximately 3,000 tracked trucks traveling in Washington, which represented around 3% of the total Washington truck population. Because the data was collected for the purpose of fleet management, the average reading frequency is around 15 minutes. Since November 2011, the sample size was enlarged to approximately 5,000 trucks. In addition to all roads throughout Washington, the data also covers 100 miles outside Washington's borders (in British Columbia, Idaho, and Oregon). Some of the GPS reading rates were improved from 15 minutes to 2~5 minutes. Figure 1-1 shows the daily coverage of GPS data collected on October 3rd 2012.

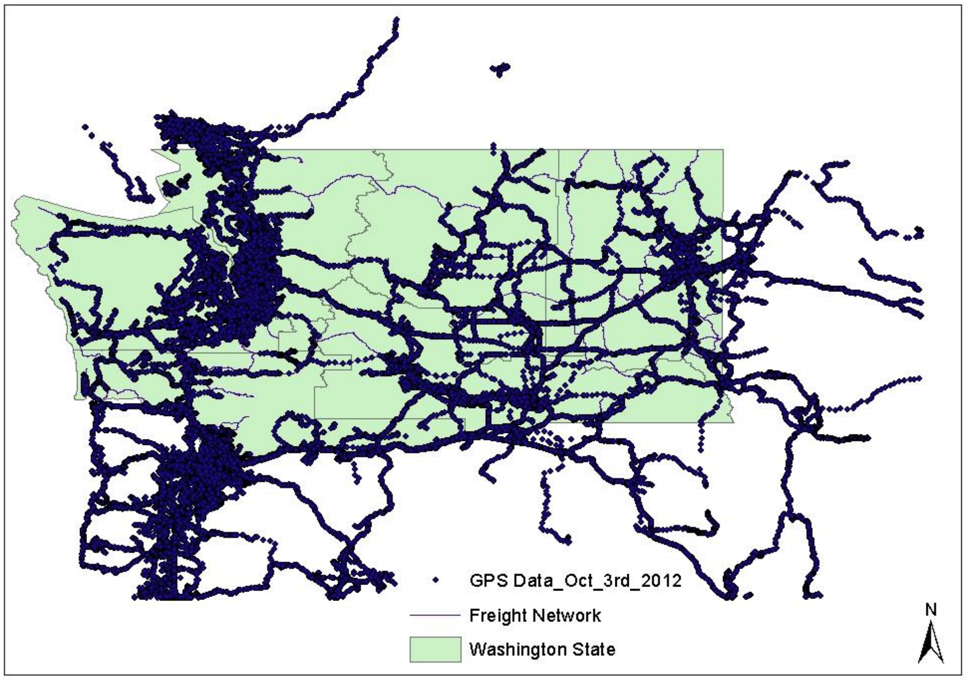


Figure 1-1 Daily Coverage of GPS Data (Collected on October 3rd 2012)

1.2.2 Applications of Truck GPS Data in Freight Mobility Measurements

Due to the increasing availability of truck GPS data to transportation agencies and researchers, there are some emerging studies of using GPS data to support freight performance measures and freight planning.

Since 2002, the Federal Highway Administration (FHWA) started collaborating with the American Transportation Research Institute (ATRI) to investigate how data gathering from the GPS device installed in trucks can be used to measure the mobility and reliability along interstate highways (Jones et al. 2005). A series of reports has been published. Five freight-significant corridors were selected in the initial two phases of study, including I-5, I-10, I-45, I-65 and I-70 (ATRI 2005). The truck GPS data was employed to measure the average travel speed; travel time index, which is defined as the ratio of observed travel speed to free-flow travel speed; and buffer index, which represents travel time reliability and is computed by dividing the extra travel time to ensure on-time arrival by average travel time. The GPS dataset was enlarged in 2006, and ATRI re-examined the truck mobility in the five corridors (USDOT 2006b). In 2009, ATRI and FHWA monitored 100 freight-significant highway locations and provided the congestion ranking based on the congestion index calculated based on truck GPS data (ATRI 2010). The congestion rank index is the total freight congestion value, which is produced based on three steps. The first step is to set the free flow speed of 55 mph. The second step is to calculate the hourly freight congestion value, which is obtained by multiplying the difference between free flow and average truck speed with the hourly number of commercial vehicles. The final step is to sum the 24 hours freight congestion value and rank the bottlenecks accordingly (ATRI 2010). The performance measure was conducted in 2010 and 2012 as well, and 250 freight-significant highway locations were monitored and ranked using truck GPS data (ATRI 2011, ATRI 2013).

FHWA and ATRI also partnered with academia and several state transportation agencies to investigate methods for developing freight performance measures. Figliozzi et al. (2011) utilized the ATRI truck GPS data to examine the travel time reliability on I-5 corridor in Oregon using statistical techniques. They developed an algorithm to segment the I-5 segment to ensure an adequate number of observations to conduct statistically reliable analysis. Truck travel time reliability was evaluated based on the travel time distribution, including 50th, 80th, and 95th percentile travel time. Florida DOT utilized the ATRI GPS data to estimate the travel speed for Florida's highways, as well as identifying truck trips and OD flow table (FDOT 2012). Minnesota DOT sponsored the research to analyze truck travel speed and reliability along I-90 and I-94 based on the ATRI GPS data (Liao 2009). Building on this research effort, Liao developed an analysis methodology using ATRI truck GPS data to study the performance of heavy commercial vehicles along 38 critical freight-significant corridors in Twin City, Minnesota to identify truck bottlenecks (Liao 2014). Truck mobility, delay, and reliability index were evaluated. The mobility was measured as the number of hours in peak periods with average speed below threshold speed. Truck travel time reliability was quantified as an index calculated by dividing the 80th percentile travel time by the travel time at pre-defined threshold speed. Truck traffic bottlenecks were identified and ranked on the average truck delay per mile and number of hours in peak periods with speed less than threshold speed.

In July 2013, the FHWA released the National Performance Measurement Research Data Set (NPMRDS), which is travel time data in every five-minute interval for all national highway facilities for both passenger and freight vehicles. Liao (2014) documented the steps to associate the NPMRDS travel time data with the national highway system GIS shapefile data. The average speed of passenger and commercial vehicles at each segment was computed by dividing the segment distance by the corresponding travel time. The derived speed was plotted on map within the ArcGIS environment.

The GPS data provided by ATRI only covers the interstate highways and major state roads and does not provide much data about arterial roads and routes in rural areas, which also experience severe truck congestions. Instead of accessing data from ATRI, WSDOT purchased the GPS data from GPS vendors directly and used the data to identify truck trips, speed, travel time and reliability (McCormack et al. 2010). The GPS data was collected since September 2011 to present and covers most truck roads in Washington. Based on this data, Ma et al. (2011) developed a web-based benchmark to process commercial GPS data for truck performance measures. An algorithm was proposed to identify truck trip originations and destinations from raw GPS data. Truck performance measures include speed, travel distance, travel time between transportation analysis zones (TAZ), and travel time reliability. Zhao et al. (2011) utilized the same dataset to compare different freeway travel time estimation methods. It is found that the mean GPS spot speed can represent the average truck travel speed along short segments, and travel time is computed by dividing the segment distance by the corresponding mean spot speed. The travel time estimates were comparable with space mean speed-based travel time estimates. In addition, WSDOT has developed a set of criteria for identifying truck bottlenecks using the same dataset (WSDOT 2011). The first criterion is truck slow-speed bottleneck, which is identified preliminarily upon the truck speed estimates retrieved from GPS devices. If there are 50% of trucks traveling below the poor performance threshold defined by WSDOT, the segment is categorized as a bottleneck. The poor performance threshold was defined as 60% of posted speed limit and 35 mph on urban freeways. The second type of bottleneck is defined as reliability bottleneck. Two approaches were employed to identify the bottlenecks. The first approach was proposed by Zhao et al. (2012) based on the truck GPS spot speed distribution. If the speed distribution of the studied segment follows a mixture of two Gaussian distributions, it is recognized as unreliable, otherwise it is reliable. The second approach relies upon the 95th percentile of truck travel time (USDOT 2006a). However, the second approach requires a vast amount of GPS data, and the current dataset used by WSDOT can only support the reliability measure of a few corridors.

While several pilot studies have been implemented, most of them assess the truck mobility by calculating the truck speed based on GPS spot speed only, and one of them deals with the travel time evaluation explicitly. However, the complexity of travel time reliability still requires substantial efforts to investigate. The forecasting of truck mobility performance using the aforementioned GPS datasets has not been investigated in any studies yet.

1.2.3 Applications of Truck GPS Data in Freight Planning

In the past decades, there have been multiple studies seeking to collect truck movement data to support freight planning, including identifying truck travel patterns, supporting travel demand model improvement, and evaluating impacts of freight policies. Sharman and Roorda (2011) developed an automatic processing application to identify truck trip destinations and track frequency of frequently visited destinations. They tested different approaches to cluster trip ends retrieved from GPS devices in order to group trip ends into repeated visits to common destinations. The comparison of a number of clustering approaches indicates that the Ward's method is superior to other methods for this application due to better clustering results and reasonable computation efforts. This research associates GPS stops with common visit destinations, which initiates new research opportunities to identify truck destination patterns. You (2012) developed methodologies for tour-based truck demand modeling. Truck GPS data was employed to gather truck tour data and understand the drayage trucks' trip patterns. Four types of truck pattern were categorized given the corresponding characteristics. The analysis results demonstrate that the traditional trip-based travel demand model cannot address drayage truck behaviors, and a tour-based model is needed.

Puget Sound Regional Council (PSRC) employed truck GPS data to support the improvement of the regional travel demand model (2009). The truck speed retrieved from GPS data along a selected corridor was compared with passenger car speed in the same corridor to improve the understanding of whether trucks travel at a different speed than passenger cars. The study also suggested potential applications of truck GPS data in improving travel demand models, including identifying zone to zone movements; identifying origins, destinations, and tours; providing information on speed, travel time, trip length, congestion, and bottlenecks; and evaluating project effects by modeling both pre- and post-investment travel time using travel demand models.

GPS data has also been used for analyzing project impacts. Golias (2013) employed GPS data to evaluate the effects of the commercial truck hours of service (HOS) rule on traffic congestion. The HOS rule regulates the maximum number of driving and working hours per day for a commercial truck driver before a rest period. The regulation is expected to reduce negative impacts of truck traffic on roadway congestion as well. The effects are evaluated through truck volume, truck trip characteristics, and the corresponding roadway level of service (LOS). A 212 mile long segment of I-40 between Memphis and Nashville, Tennessee was selected as the case study. The truck GPS data was provided by ATRI and collected between September 1st, 2011 and October 31th, 2011. Each GPS record includes truck ID, latitude, longitude, timestamp (date and time), speed, and heading. GPS data was spatially associated with the highway segment being studied within the ArcGIS environment. The analysis results reveal that truck traffic increased significantly during peak period and consequently lead to a worse traffic condition after applying the new HOS rule.

Despite the fact that truck GPS data has been applied to various freight planning studies, none of them has investigated the tolling project impacts on truck travel performance and behaviors. Thus this dissertation intends to fill this gap by exploring the GPS data capability in quantifying tolling project impacts.

1.3 Research Objectives

The overall goal of this dissertation is to use the advantages of the information provided by the truck GPS data to support freight planning. To fill the gap identified in the previous section, three specific research objectives are identified:

The first research objective is quantifying current travel time reliability using GPS data. Accordingly, the set of research questions are:

- What are the existing implementable metrics that can be used to estimate truck travel time reliability given the characteristics of the GPS data we are using?
- What are the pros-and-cons of each method?
- If there are limitations of the existing approaches, how can we improve it?
- What are the appropriate reliability measures to use under different situations?

The second research objective is predicting truck travel time and travel time reliability for freight planning using truck GPS data. The second set of research questions include:

- What are the limitations of using the traditional engineering equations and travel demand models to predict the post-investment truck travel time and travel time reliability for freight planning projects?
- What methods can be applied to predict truck travel time based on GPS data? What data do we need to implement the forecasting? How does the accuracy vary spatially and temporally? How to validate the results?
- What methods can be employed to forecast travel time reliability? Why do we choose this method? What data do we need to support the forecasting?

The third research objective is applying truck GPS data to support freight planning by quantifying the toll road impacts on truck speed and route choice. The third set of research questions are:

- Will the truck travel speed be affected after the SR-520 Bridge toll project? To what extent will they be affected?
- How can GPS data be used to quantify truck route choice?
- What are the factors affecting truck route choice? What are the impacts of these factors?

1.4 Organization

The remainder of this dissertation is organized as follows:

Chapter 2 starts with the literature review of the existing commonly applied travel time reliability measures. It is followed by a discussion of the challenges of employing truck GPS data to evaluate reliability using these approaches. The author then proposes an improvement to the existing spot speed distribution based reliability measure and compares it with a number of existing approaches. Correlations are provided between the improved approach and a number of commonly used reliability measures. The advantages and disadvantages of each measure are discussed and recommendations of the appropriate measures for different applications are presented.

Chapter 3 proposes a speed-density relationship based freeway truck travel time prediction approach based on truck GPS data and loop data. Traffic regimes are segmented using a cluster analysis approach. The proposed method is compared with another two traditional approaches. The results indicate that the proposed method generates more accurate travel time estimates.

Chapter 4 examines the impact of traffic condition on GPS spot speed distribution. The relationship between traffic density and coefficient of variation of GPS spot speed distribution is developed to forecast truck travel time reliability based on truck GPS data and loop data.

Chapter 5 explores how GPS data can be used to evaluate impacts of tolling on truck speed and routing. The City of Seattle SR-520 toll bridge is selected as a case study. Truck GPS data is employed to evaluate route choice and changes in travel speed along the toll route SR-520 and the alternative toll-free route I-90. A logit model is developed to determine the influential factors during both peak and off-peak periods. The values of truck travel time during both periods are also estimated.

The final chapter presents the conclusions and contributions of this research and discusses the potential obstacles and challenges to implement the proposed methodologies.

Chapter 2 MEASURING TRUCK TRAVEL TIME RELIABILITY USING TRUCK PROBE GPS DATA

This chapter focuses on proposing a travel time reliability measure given the characteristics of the GPS data used in this dissertation, and providing recommendations of the appropriate reliability measures to use under different conditions. It starts with a literature review of a number of commonly applied travel time reliability measures. It is followed by an improvement to the recently proposed GPS spot speed based approach. The author then applies the improved GPS spot speed based measure and those widely applied reliability metrics to a case study, and compares the correlations among these reliability measures. In addition, the advantages and disadvantages of each measure are summarized and appropriate methods for different applications are recommended.

2.1 Introduction and Background

Travel time reliability represents the level of consistency in travel times for the same trip for a time period (Lomax et al. 2003). It has been recognized as a critical factor in truck routing and scheduling. A survey conducted by Bogers and van Zuylen (2004) found that truck drivers prefer the more reliable route, even if it involves a longer trip in comparison to other routes with shorter travel time and higher uncertainty. While travel time reliability is a factor for determining truck route choice, it is also becoming an important component of freight mobility performance metrics (Cambridge Systematics 2013, USDOT 2006b). Given the importance of travel time reliability, numerous quantitative approaches have been proposed to measure travel time reliability based on a variety of data sources. The truck probe data collected from GPS devices has gained increased attention as a source of truck travel time reliability input given the growing market penetration of GPS technology, as well as the improved truck specific vehicle location and speed information provided by GPS devices. Meanwhile, the Moving Ahead for Progress in the 21st Century (MAP-21) program will make GPS data from commercial vehicles available to transportation agencies for evaluating regional freight performance, including travel time reliability.

Most truck travel time reliability studies that apply GPS data are based on travel time observations that are retrieved from GPS data (ATRI and FHWA 2005, USDOT 2006b, Liao 2009, USDOT 2010, Figliozzi et al. 2011). The travel time observations require substantial sample size to ensure statistical reliability (NCHRP 2008, Figliozzi et al. 2011), and the major challenges to using GPS data to obtain travel times are small non-random observation sets and low reading frequency. In contrast, using GPS spot speed directly can alleviate the low read rate and read density concerns. In addition, raw GPS data typically provides spot speed (not travel time), and the conversion from spot speed to travel time for a particular segment involves data processing and therefore may cause a loss of data accuracy. Despite the potential of truck GPS spot speed data to support truck specific travel time reliability assessment, there are limited studies investigating reliability metrics based on spot speed data. Zhao et al. (2013) developed the GPS spot speed distribution based approach to evaluate truck travel time reliability and identified bottlenecks based on the hypothesis that the truck speed distribution can be modeled by either unimodal or bimodal probability density functions. They further identified that if truck speed follows a bimodal distribution, the segment is classified as unreliable. Otherwise, it is defined as reliable. However, this reliability metric only classifies reliability into three categories: reliably slow, reliably fast and unreliable. It does not provide a numerical value which would allow for a more quantitative evaluation, e.g. ranking reliabilities on different segments or during different time periods, or quantifying the changes in travel time reliability associated with transportation investments.

In light of this, the objective of this chapter is to improve the current GPS spot speed based reliability metric by proposing a means to support more quantitative analyses. In addition, the authors compare the proposed approach with a number of commonly used travel time based reliability measures: travel time coefficient of variation (COV), buffer time index (BI), skew, and truck reliability index (RI_{80}). The appropriate reliability measures for different applications are discussed. The remainder of this chapter is organized as follows: section two (2) provides a brief review of the commonly used travel time reliability measures that are implementable with truck GPS data, and discusses the sample size constraint associated with travel time based reliability measures; section three (3) proposes the improvement to the recently proposed GPS spot speed based approach, section four (4) applies the improved GPS spot speed based measure and those widely applied travel time based metrics to a case study and compares the correlations among these reliability measures; section five (5) offers findings and conclusions of the analyses.

2.2 Literature Review on Current Travel Time Reliability Measures

There has been substantial effort to develop travel time reliability measures relying upon statistical techniques and probe data collected from GPS devices. Comprehensive overviews of travel time reliability measures can be found in Lomax et al. (2003), NCHRP Report 618 (2008) and Cambridge Systematics (2013). Several commonly applied reliability measures are reviewed in this section since they can be measured and implemented with GPS data, and have been tested and applied in practical projects. The authors classified these measures into two categories according to the data on which these approaches are based: travel time based reliability measure and GPS spot speed based measure.

Travel Time Based Reliability Measures

(1) Standard Deviation and Coefficient of Variation (COV)

The travel time standard deviation is a measure of how spread observations are. The larger the value of the standard deviation, the lower the travel time reliability. In addition to the standard deviation, the ratio of the standard deviation and the mean, also called the coefficient of variation, is defined as a reliability measure. This value is interpreted as the larger the standard deviation relative to the mean, the lower the travel time reliability. One example of the use of COV approach is the research led by the U.S. Department of Transportation (USDOT) on measuring the crossing-border truck travel time and travel time reliability (USDOT 2010). The study location was the Otay Mesa International Border between the U.S. and Mexico. Truck GPS data was collected from January 2009 to February 2010. A large travel time standard deviation from the mean was observed, which ranged from 61% to 81% of the mean value. Therefore the study concluded that carriers crossing the border experienced very low travel time reliability.

(2) Percentile Method

In this method, the 95th percentile travel time was recommended by USDOT as the metric to compare travel time reliabilities on different segments (TTI and Cambridge Systematics 2006). This 95th percentile travel time method is used to measure very long travel times based on observations over a certain time period, e.g. across one year. It estimates the time travelers need to plan in order to meet a desired arrival time. It is also called planning time. It is recommended by the National Cooperative Highway Research Program (NCHRP) as the simplest indicator of travel time reliability (NCHRP 2008). Researchers may also use the 80th or 85th or other percentiles as the base. The SHRP 2 (Second Strategic Highway Research Program) recommended using 80^{th} percentile travel time instead of 95^{th} percentile travel time since they found that events that contribute to the 80^{th} percentile travel time are more common events and are more likely to be influenced by operation strategies, e.g. improvement to transportation infrastructures (Cambridge Systematics 2013). Figliozzi et al. (2011) evaluated travel time reliability along the I-5 corridor through the State of Oregon based upon truck GPS data accessed from the American Transportation Research Institute (ATRI). The 50^{th}, 80^{th} and 95^{th} percentile travel time were selected as metrics to measure travel time reliability along the I-5 corridor.

(3) Buffer Time Index (BI)

Buffer time is defined as the extra travel time travelers must add to the average travel time to allow for on-time arrival, and it is calculated as the difference between the 95% travel time and average travel time (TTI and Cambridge Systematics 2006). The buffer time index (BI) is calculated by dividing buffer time by the mean travel time. Federal and regional transportation agencies have used the BI to evaluate system performance. The Federal Highway Administration (FHWA) and ATRI have evaluated how information retrieved from GPS devices could provide data to support freight travel time reliability measures. The BI measure was employed to evaluate freight travel time reliability along five major freight corridors in the U.S. (USDOT 2006). The Minnesota DOT evaluated freight performance along I-94/I-90 from the Twin Cities to Chicago using archived truck GPS data and freight travel time reliability was evaluated using the BI metric (Liao 2009).

(4) Skew

While standard deviation and COV represent the spread of the travel time distribution, the skew depicts the "leaning" of travel time distribution to one side of the mean. Van Lint and van Zuylen

(2005) examined the travel time distribution along a 19.1 km freeway in Netherlands and found that both width and skew of travel time distribution change with respect to different traffic regimes (van Lint et al. 2008). The travel time distribution is approximately symmetric before congestion, with small values of both width and skew. The distribution tends to be left-skewed with wider breadth (longer tail) during the onset of congestion. During the congested period, the travel time distribution growths wider and becomes right skewed. Finally, while congestion wanes, both the median travel time and the spread of travel time distribution decrease, and the distribution is left skewed again (van Lint et al. 2008). Given the changes in both width and skew, van Lint and van Zuylen (2005) suggested that not only the variance of travel time should be used as reliability measures, but also the skewness. The skewness is quantified by comparing how much of the 90th percentile travel time is greater than the median to how much the 10th percentile travel time is less than the median, as expressed in Equation (2-1) (van Lint and van Zuylen 2005, van Lint et al. 2008).

$$Skew = \frac{T_{90} - T_{50}}{T_{50} - T_{10}} \tag{2-1}$$

(5) Truck Reliability Index(RI_{80})

The reliability measure recommended by The American Association of State Highway and Transportation Officials (AASHTO) for the MAP-21 Program is the RI_{80}, which is defined as the ratio of the total truck travel time needed to ensure on time arrival to the agency-determined congestion threshold travel time (e.g. observed travel time or preferred travel time) (AASHTO 2012, Cambridge Systematics 2013). The 80th percentile travel time is chosen to represent the total truck travel time needed. The congestion threshold travel time is determined by each transportation agency and should account for various reasons to slowing trucks, e.g. weather, congestion, accident and work zone.

Sample Size Constraint

The reliability measures discussed above rely upon travel time observations. Two common approaches are utilized to compute the travel times on a specific roadway segment using GPS data. One is the vehicle location based approach. Two buffers are created at the segment start and end points respectively, and truck trips that have GPS reads in both buffers are identified. The difference between the two timestamps in the two buffers is viewed as the travel time along the segment (Figliozzi et al. 2011). Another approach is the "estimated link speed" method (Zhao et al. 2011). This method is based upon the assumption that averaged GPS spot speed is able to approximate the travel speed along the segment when the segment is short, and consequently the travel time can be approximated by dividing the segment length by the average spot speed. Both approaches require sufficient travel time observations to ensure the estimated travel time can represent the link travel time with reasonable accuracy. The minimum number of travel time observations is proposed to ensure statistical reliability, and it is determined by the precision desired by the analysts and the variability of the dataset (NCHRP 2008). If analysts need to know the average travel time very precisely and the variability of the observations is high, e.g. during peak period, a large number of observations will be required. The minimum required number of observations is shown in Equation (2-2) (NCHRP 2008).

$$N = 4 \times \left[t_{(1-\alpha/2),N-1} \times \frac{S}{CI_{1-\alpha\%}} \right]^2 \qquad (2\text{-}2)$$

where N = minimum required number of observations,

$CI_{1-\alpha\%}$ = confidence interval for the true mean with probability of (1-α)%, where α equals to the probability of the true mean not lying within the confidence interval,

$t_{(1-\alpha/2),N-1}$ = the t statistic for the probability of two-sided error summing to alpha with N-1 degrees of freedom,

S = the standard deviation in the measured travel times.

If the number of minimum observations is not reached, analysts need to either extend the time period, e.g. from a 30-minute to a one hour interval, or increase the length of the segment being studied. However, for some segments with sparse GPS datasets and low data reading frequency, the minimum sample size cannot be achieved even if the analysis time period is extended to 3 hours. Also, the length of the segment should not be too long since roadway segmentation is mainly determined by the changes in truck volume, roadway geometric design and traffic control to ensure similar roadway characteristics.

GPS Spot Speed Based Approach

Sample size is often a challenge when producing travel time based reliability metrics of statistical strength. To alleviate the challenges, Zhao et al. (2013) developed a GPS *spot speed* distribution based approach, which provides a reliability measure with a sparse GPS dataset. WSDOT has evaluated the truck reliability performance and identified freight bottlenecks using this approach (McCormack et al. 2011). The probe data used in Zhao's research is sparse for most segments, and is not sufficient to provide a travel time distribution to support travel time reliability analyses using the travel time based reliability methods reviewed above. Instead of examining the travel time distribution, they plotted the spot speed on each segment during certain time periods. It was found that a mixture of two Gaussian distributions provided the best fit for the truck speed observations. Zhao et al. (2013) assessed the reliability by evaluating the speed distributions with the assumption that the travel time is unreliable if bimodal distributions are observed. Otherwise (a unimodal distribution), it is classified as reliable. The probability density function of a mixture of two Gaussian distributions is shown in Equation (2-3). The parameters are fitted based on the maximum likelihood rule.

$$f(x) = w \cdot n(x, \mu_1, \sigma_1) + (1-w) \cdot n(x, \mu_2, \sigma_2)$$
$$n(x, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right] \quad (2\text{-}3)$$

where w = the proportion of the first normal distribution,

μ_1 and μ_2 = mean of the first and second Gaussian distribution,

σ_1 and σ_2 = standard deviation of the first and second Gaussian distribution.

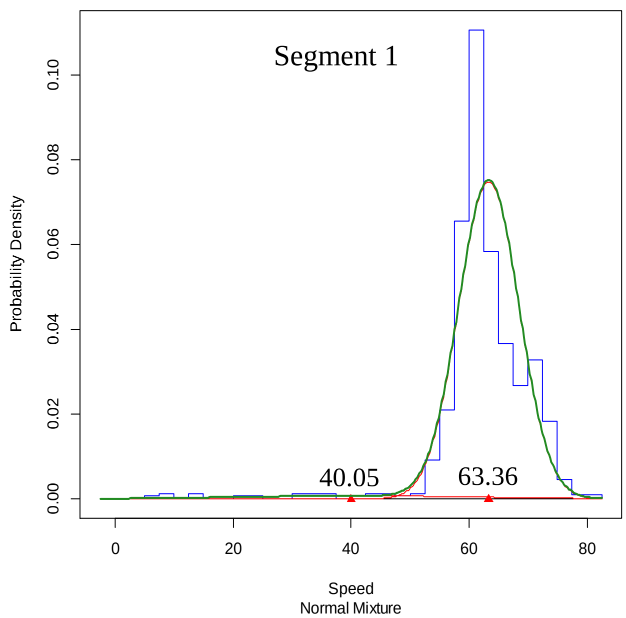
The approach defines the travel condition as unreliable if and only if

$|\mu_1 - \mu_2| \geq |\sigma_1 + \sigma_2|, w \geq 0.2,\ and\ \mu_1 \leq 0.75 \times V_p$ (V_p is the posted speed), otherwise, it is viewed as reliable. For the reliable performance, it is subdivided into reliably fast and reliably slow depending on the average speed. The major advantage of this methodology is that the reliability evaluation does not require a large number of travel time observations, rather only spot speed. However, the current method does not provide a numerical value which would allow for a more quantitative evaluation and ranking.

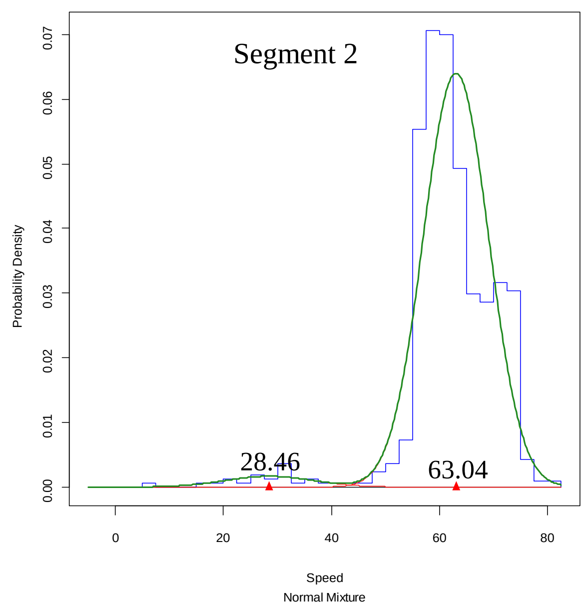
The literature review section recalls a number of commonly applied travel time reliability measures. In contrast to the travel time based reliability measures, the reliability metric proposed in the next section is based on GPS spot speed data. It is an improvement to the newly proposed spot speed based approach discussed above, which allows more quantitative analyses.

2.3 Improvement to the GPS Spot Speed Based Reliability Measures

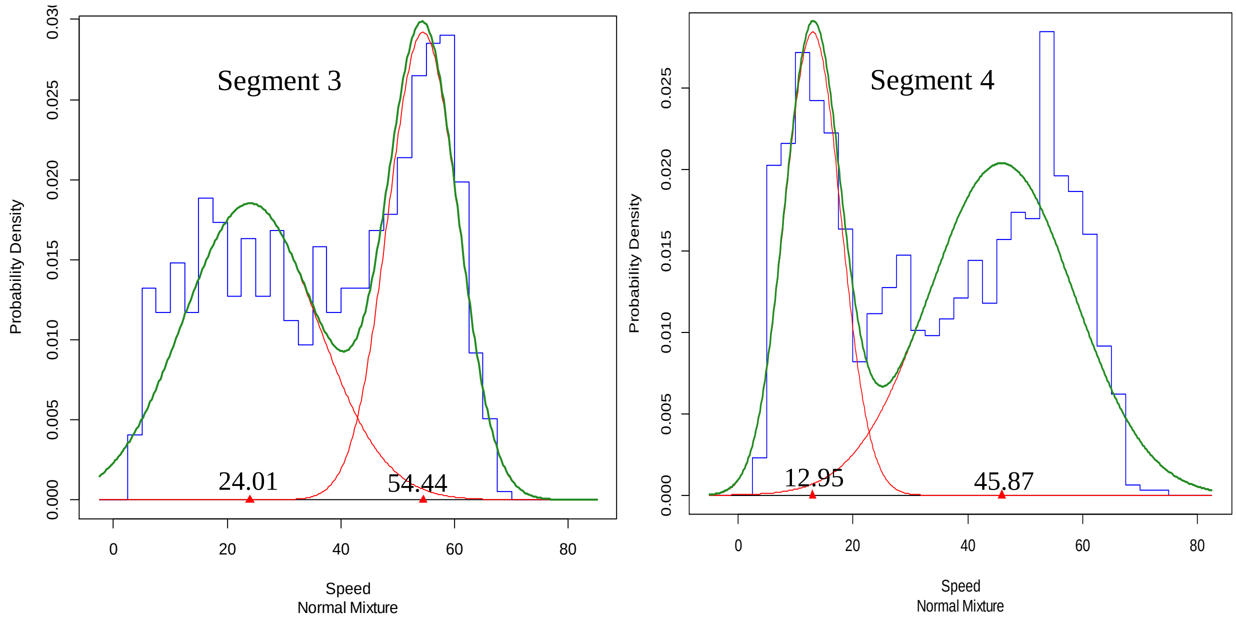
As discussed in the previous sections, the current GPS spot speed based approach can only classify segment travel time reliability into three categories. For example, Figure 2.1 shows the fitted truck spot speed distributions of four segments during the AM peak period (6:00 AM – 9:00 AM) based on GPS observations collected in May 2012. Segment 1 and Segment 2 are the stretch of 9 miles of eastbound and westbound of Interstate 90 (I-90) near Spokane, WA. Segment 3 and Segment 4 are the stretch of 3.5 miles of southbound and northbound of Interstate 5 (I-5) near downtown Seattle, WA. The corresponding fitted parameters are given in Table 2-1. The distribution fitting was accomplished using the R software package "mixdist" (Du 2002). Taking segment 1 as an example, the fitted result can be interpreted as follows. The GPS spot speed distribution is composed of two traffic regimes. The average truck travel speed of the first traffic regime is 40.05 mph, with standard deviation of 21.6 mph. The average truck travel speed of the second traffic regime is 63.36 mph, with standard deviation of 5.11 mph. The probability of truck travel speed falling within the first traffic regime is 4%, which indicates that the probability of truck travel speed falling within the low-speed regime is very small. Since the fitted parameters do not meet the rule of "$|\mu_1 - \mu_2| \geq |\sigma_1 + \sigma_2|$ *and* $w \geq 0.2$", the travel time distribution of segment 1 follows a unimodal distribution (as shown in Figure 2-1 (a)). In addition, the average travel speed is 62.29 mph, which is greater than the 75% of the posted speed limit. Thus travel time on segment 1 is further defined as reliably fast. Similarly, the truck travel time on segment 2 is also defined as reliably fast. For segment 3, it is also composed of two traffic regimes. The average truck travel speed of the first traffic regime is 24.01 mph, with standard deviation of 11.78 mph. The average truck travel speed of the second traffic regime is 54.44 mph, with standard deviation of 6.19 mph. The probability of truck travel speed falling within the low-speed traffic regime is 55%. The fitted parameters meet the predefined rule, and therefore the travel time on segment 3 is defined as unreliable during the AM peak period. Similarly, segment 4 is defined as unreliable. However, this approach can only identify the reliability category, but it is not able to rank the reliabilities to identify the most unreliable segment.



(a)



(b)



(c) (d)

Figure 2-1 GPS Spot Speed Distribution Fittings of Four Segments during AM Peak Period

Table 2-1 Estimated Parameters for GPS Spot Speed Distribution Fittings of Four Segments

	Segment 1	Segment 2	Segment 3	Segment 4
w	0.04	0.03	0.55	0.35
μ_1	40.05	28.46	24.01	12.95
σ_1	21.60	8.16	11.78	4.94
μ_2	63.36	63.04	54.44	45.87
σ_2	5.11	6.02	6.19	12.65
V_p	60	60	60	60
Average speed	62.29	61.84	37.52	34.16
$if\ \lvert\mu_1-\mu_2\rvert \geq \lvert\sigma_1+\sigma_2\rvert, w \geq 0.2$ $and\ \mu_1 \leq 0.75 \times V_p$	No	No	Yes	Yes
$if\ \ average\ speed \leq 0.75 \times V_p$	No	No		
Reliability category	Reliably Fast	Reliably Fast	Unreliable	Unreliable

A coefficient of variation (COV) method is proposed to improve the current approach. Since it has been proven that the GPS spot speed distribution follows a mixture of two Gaussian Distributions, the mean and standard deviation of the spot speed distribution can be calculated based on the parameters of the two distributions, as shown in Equation (2-4). The COV which is computed by dividing the standard deviation by the mean is employed as travel time reliability measure, as shown in Equation (2-5).

$$\mu = \sum_{i=1}^{n} w_i \mu_i$$
$$\sigma^2 = \sum_{i=1}^{n} w_i ((\mu_i - \mu)^2 + \sigma_i^2) \tag{2-4}$$

$$Coefficient\ of\ Variation\ (COV) = \frac{\sigma}{\mu} \tag{2-5}$$

where μ = mean of the mixture of Gaussian distributions,

w_i = weight of the ith Gaussian distribution,

μ_i = mean of the ith Gaussian distribution,

σ = standard deviation of the mixture of Gaussian distributions,

σ_i = standard deviation of the ith Gaussian distribution,

n = number of Gaussian distributions, n=2 since it has been proved that spot speed follows of the mixture of two Gaussian distributions.

Using Equation (2-5), the corresponding COV of the four segments can be computed, as displayed in Table 2-2. The authors ranked the travel time reliability based on the values of the COV, where 1 represents the least reliable segment and 4 represents the most reliable segment. The larger the standard deviation relative to the mean, the lower the travel time reliability. According to the calculation, segment 4 was identified as the most unreliable segment during AM peak period based on the one month GPS spot speed observations, and segment 1 was the most reliable segment. This information can be used to support resource allocation and planning.

Table 2-2 Reliability Measurements and Ranking Results of the Four Segments

	Segment 1	Segment 2	Segment 3	Segment 4
Mean	62.43	62.00	37.70	34.35
Standard deviation	8.04	8.48	17.97	18.95
COV	0.13	0.14	0.48	0.55
Reliability Ranking	4	3	2	1

2.4 GPS Data Based Travel Time Reliability Measures Comparison

This section provides a case study to compare various reliability measures by ranking the reliabilities on the same segment during different times-of-day and days-of-week, and computing the correlation among these measures.

Study Area and Description of the Probe Data Used

A stretch of 3.5 miles of southbound Interstate 5 (I-5) through downtown Seattle was selected for the case study. Travel time reliability was examined during two time periods: off-peak period (12:00 AM – 6:00 AM) and AM peak period (6:00 AM – 9:00 AM). The GPS data acquisition efforts on which this dissertation is based was a pilot study of investigating how to gather and use existing GPS data collected for truck fleet management to develop performance measures for trucks (McCormack et al. 2010). The GPS data was initially collected by a GPS vendor for trucking companies' fleet management. The data is purchased from the GPS vendor directly and it includes a unique device ID, vehicle trajectory, spot speed, heading, location (latitude and longitude), time and date. Each device ID was scrambled for anonymity to protect the customer information. Data was collected from January 2012 to December 2012. The average GPS reading frequency ranges from 2 to 15 minutes. More details of data collection efforts can be found in McCormack et al. (2010), Ma et al. (2011) and McCormack et al. (2011). The GPS data processing consists of three steps: (1) cleaning data to filter out problematic and duplicated data, (2) geocoding GPS data to road segments, and (3) estimating travel time from GPS spot speed (if travel time based reliability measures are selected). More details of the data processing and travel time estimation can be found in McCormack et al. (2011) and Zhao et al. (2011).

The traffic performance information retrieved from the GPS dataset represents the performance of trucks equipped with GPS devices. Previous research by Zhao et al. (2011) has demonstrated that the mean truck travel speed computed from the GPS data compared well with the mean mixed traffic speed recorded by loop detectors deployed in the right-most lane (the absolute differences between the two values are less than 6%). The MAP-21 program will provide State DOTs with GPS for performance measures. Although the GPS data may be provided by different vendors than that described in this dissertation, the data formats are consistent with those identified in this dissertation.

Reliability Ranking Results

Truck travel time reliability on the selected segment was measured using a number of reliability metrics: COV, BI, skew, RI_{80} and the improved GPS spot speed based method. Travel time standard deviation method was not included in the case study since it is highly correlated to the COV. The RI_{80} was calculated by dividing the 80^{th} percentile travel time by 60% of posted speed (Washington State Department of Transportation's congestion threshold (WSDOT 2010)). The 80^{th} percentile travel time measure was not included in the analysis as it is highly correlated with the RI_{80} metric.

Table 2-3 shows the reliability ranking results of the same segments during different times-of-day and days-of-week, each based on one of the reliability measures listed above.

Table 2-3 (a) Reliability Ranking Results during Off-peak Period (12:00 AM – 6:00 AM)

Measures	Mon	Tue	Wed	Thu	Fri
COV	2	5	4	3	1
BI	3	5	2	4	1
Skew	3	5	4	2	1
RI_{80}	3	5	2	4	1
Improved GPS spot speed based method	3	5	2	4	1

Table 2-3 (b) Reliability Ranking Results during AM Peak Period (6:00 AM – 9:00 AM)

Measures	Mon	Tue	Wed	Thu	Fri
COV	1	4	2	3	5
BI	4	5	3	1	2
Skew	1	4	5	3	2
RI_{80}	4	3	2	1	5
Improved GPS spot speed based method	4	2	1	3	5

The travel time reliability ranking results vary depending on the measures used. During the off-peak period, all measures identify travel time on Friday as the least reliable and travel time on Tuesday as the most reliable. However, the rankings of the rest of three days are different. The rankings differ significantly during the AM peak period. The COV and skew metrics indicate travel time on Monday is the least reliable compared to other days, the BI and RI_{80} method show that truck drivers experienced the most unreliable travel time on Thursday, and travel time on Wednesday is defined as the most unreliable by the improved GPS spot speed based approach. The differences stem from the fact that different measures capture different components of reliability.

Correlations among Travel Time Reliability Measures

The above ranking results indicate that different measures get different conclusions even if the same data is used. To further explore the relationship among these measures, the correlations among each measure were calculated, as displayed in Table 2-4. The values represent the degree to which these measures are related.

Table 2-4 (a) Correlations among Reliability Measures during Off-peak Period

	COV	BI	Skew	RI_{80}	Improved GPS Spot Speed
COV	1.000				
BI	0.639	1.000			
Skew	0.666	0.408	1.000		
RI_{80}	0.695	0.735	0.446	1.000	
Improved GPS spot speed	0.556	0.433	0.420	0.769	1.000

Table 2-5 (b) Correlations among Reliability Measures during AM Peak Period

	COV	BI	Skew	RI_{80}	Improved GPS Spot Speed
COV	1.000				
BI	0.679	1.000			
Skew	0.471	0.418	1.000		
RI_{80}	0.508	0.595	0.135	1.000	
Improved GPS spot speed	0.322	0.196	-0.223	0.821	1.000

Although the results above are based on a specific freeway segment truck travel time and speed data, it reveals a general finding that there are large deviations among the travel time based reliability measures, and between the travel time based reliability measures and the improved GPS spot speed based approach. What's more, the deviations are more significant during peak period compared to off-peak period.

The COV and Skew are not highly related during the off-peak period (with correlation of 0.666), and they are even more weakly related during peak period (with correlation of 0.471). By examining the definitions of the two measures, we see that they capture different characteristics of the travel time distribution. The COV evaluates the width or spread of travel time distribution, while the Skew depicts the leaning of travel time distribution. It is not necessarily the case that a small variance is associated with small skew, especially when the travel time distribution is highly left-skewed (during congestion onset and congestion dissolve regimes).

The COV is not closely related to the BI either. This is because the BI is computed based on the difference between the extreme travel time (80th percentile travel time) and the average travel time. The smaller difference between the extreme travel time and average travel time is not necessarily related to a small COV since a few extreme values affect the mean more significantly than the extreme travel time, e.g. 80th, 90th and 95th percentile travel time (Cambridge Systematic 2013). This is explained in Figure 2-2, which displays two travel time distributions. The first distribution contains some extreme travel time values, and the distribution is left-skewed. The second one represents the distribution after removing those extreme values. The traffic performance of the second condition is more reliable than the first one and generates smaller COV. However, as shown in Figure 2-2, the corresponding BI of the second condition is greater than the first one. This may explain the week correlation between COV and BI. As a result, several studies suggested computing BI by using median travel time instead of mean travel time (Cambridge Systematics 2013, Pu 2010).

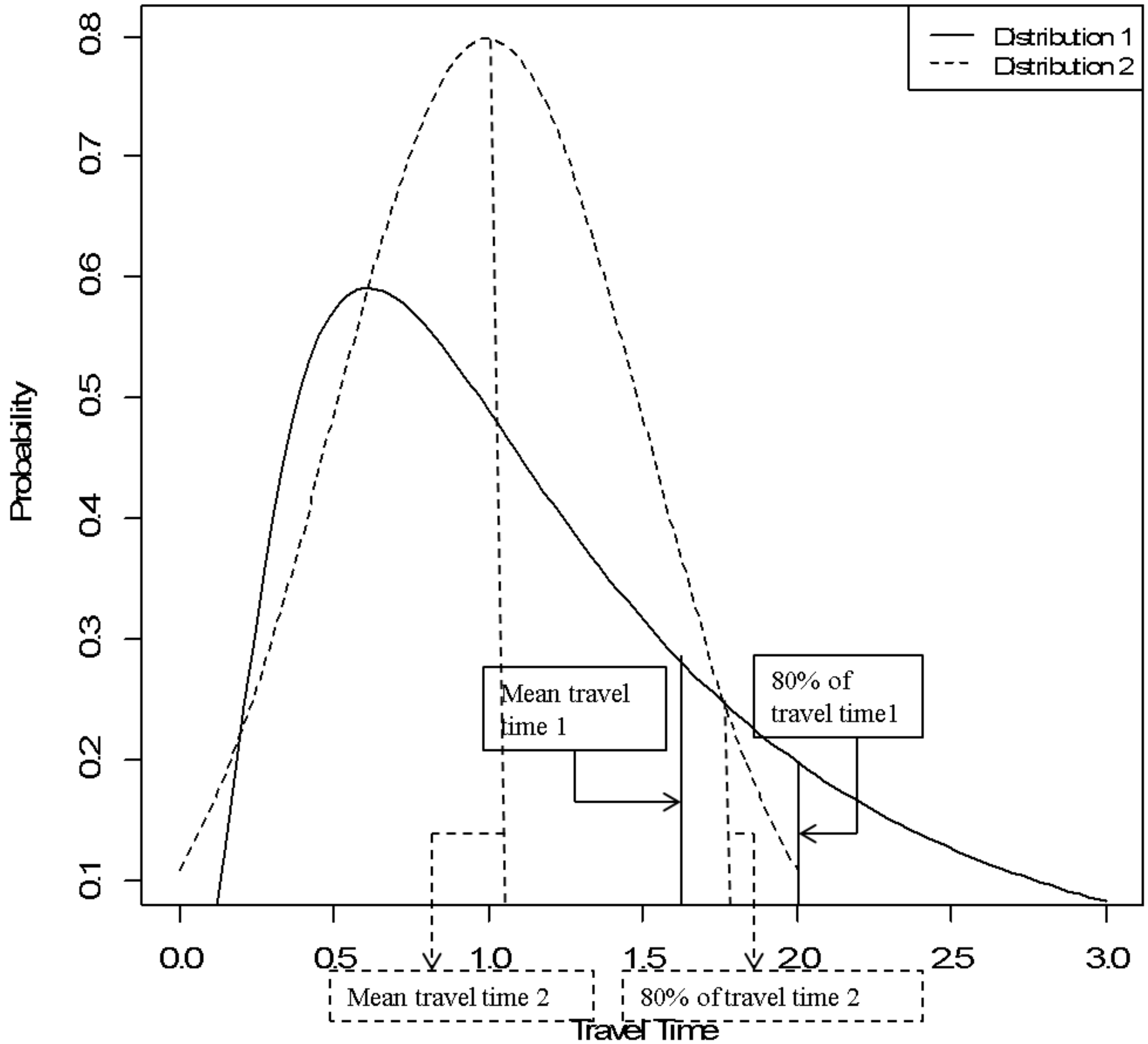


Figure 2-2 Mean and 80th percentile Travel Time of Two Distributions

The improved GPS spot speed approach is computed as the coefficient of variation of the speed distribution. However, the travel time based COV metric is not highly related to the improved GPS spot speed approach either. This may due to the loss of data accuracy during the conversion from GPS spot speed to travel time estimations.

The correlation analysis reveals that different measures provide different conclusions for the same underlying data and traffic conditions, and judgment as to whether or not a particular segment is reliable depends on the reliability measures used. Table 2-5 presents the comparison of each measure. The selection of the appropriate measures depends on the characteristics of each measure discussed above, as well as the available dataset, potential users and analysis purposes.

Table 2-5 Comparison of Travel Time Reliability Measures

	SD and COV	Percentile Method	BI	Skew	RI_{80}	GPS spot speed based approach
Does not require conversion of original data into travel time estimate.						×
Smaller sample size requirement.						×
Has been widely applied.	×	×	×			
Easy to compute when historical travel time observations are available.	×	×	×	×	×	
Easy to interpret to non-technical users.		×	×			
Ability to be applied for daily trip planning.		×	×			
Ability to compare reliability across trips and segments.			×		×	×
Ability to indicate if congestion is increasing or decreasing.				×		

As summarized in Table 2-5, if raw GPS data is readily available, the GPS spot speed approach is preferred as it does not require additional data processing efforts to retrieve travel time estimates from raw data. It is an asset to avoid this task as it requires resources, and provides opportunities for introducing error. For State DOTs and other transportation agencies, when travel time observations are readily available, and if the intent of the analysis is to rank reliability on different segments or identify truck bottlenecks, the SD, COV and RI_{80} are recommended as they are easy to compute and provide quantitative measures to rank reliability. If the real time data is available and the analysts aim to propose efficient traffic operation strategies to alleviate traffic congestion, the Skew could be considered due to its capability in capturing short-term traffic trends. For a nontechnical audience, the Percentile Method and the mean value-based BI are ideal measures for trip planning, e.g. determining departure time and vehicle routing for freight industry.

2.5 Findings and Conclusions

This article recalls a number of reliability measures that are implementable with truck GPS data. These measures are classified into two categories: the travel time based measures and spot speed based measures. GPS data sample size is one of the major concerns of implementing the travel time based measures with sparse GPS observations. In addition, truck GPS data usually provides instantaneous speed and the conversion from spot speed to travel time estimates requires additional efforts, and therefore may cause loss of data accuracy. The recently proposed GPS spot speed based approach can alleviate the sample size constraint, and it does not require data processing from spot speed to travel time estimates. However, it is not able to provide a quantitative means for ranking and comparing reliabilities. Thus an improvement was made to provide a means for more quantitative analyses by calculating the spot speed distribution COV. The improved spot speed based reliability measure is able to provide numerical values which allow for quantitative analyses.

The improved method was compared with a number of travel time based reliability measures. It is found that the assessment of whether or not travel time reliability on a particular segment during specific times-of-day and days-of-week periods should be regarded as reliable depends on the reliability measures used. Different measures may get different conclusions for the same underlying data. Therefore, the ranking of travel time reliability bottlenecks varies depending on the reliability measures used, and even just those travel-time based reliability measures do not obtain the same conclusion. The correlation calculation indicates that there are large deviations among reliability measures, and it is mainly due to the fact that these measures capture different components of reliability. For instance, the COV represents how spread the observations are, the BI captures the impacts of extreme values, and the Skew reflects the leaning of travel time distribution to one side of the mean. Given the different characteristics of each measure, the selection of the appropriate measures for different applications are determined by the available datasets, potential users and analysis purposes.

Chapter 3 FREEWAY TRUCK TRAVEL PREDICTION

3.1 Background

Predicting truck travel time is a principle component of freight planning. For instance, most freight prioritization tools, which are required in an era of state and regional budget constraints, count travel time reduction as one of the key project benefits associated with a freight investment. Travel time changes are also an input into other calculations, for example, vehicle operating cost. Historically, vehicle performance functions have been used to predict travel time in planning tools, e.g. project prioritization tools and travel demand models. However, most vehicle performance functions are designed to represent passenger travel and do not consider truck performance separately. As a result, in these planning tools, trucks performance is either treated the same as passenger travel or approximated by simply applying an adjustment factor to passenger travel. For instance, the Puget Sound Regional Council travel model converts truck volume to passenger car equivalents for trip assignment and applies an additional 25% factor on travel time of trucks traveling on freeways during model calibration (Cambridge Systematics, 2007). Similarly, the Atlanta Regional Commission model assigns trucks to the network with a time-penalty value in relative to passenger travel (Atlanta Regional Commission 2011). Although there are considerable truck specific models, they are designed for modeling truck demand generation and distribution (Cambridge Systematics and Jack Faucett Associates, 2001) and no truck specific performance function for predicting travel time is found in the literature. The reason is due to the deficiency of truck specific movement data as customer privacy issues and strategic concerns prevent companies from sharing their truck movement data. What's more, many passenger trip travel time prediction models relying upon out-of-date data or limited samples. For instance, the widely applied travel time prediction model developed by the U.S.

Bureau of Public Roads, called as the BPR function, was proposed based on data collected on uncongested highways, and therefore is not able to capture the travel time under congestion condition (U.S. Bureau of Public Roads,1964).

Fortunately, the needs for quantitative freight performance measures and planning have been recognized and several studies have investigated how truck GPS data can be used to support freight performance measure and planning. The U.S. Federal Highway Administration (FHWA) collaborated with the American Transportation Research Institute (ATRI) to investigate how data gathering from GPS devices installed in trucks can be used to measure the mobility and reliability along interstate highways (ATRI and FHWA 2005). In June 2013, the FHWA released the National Performance Management Research Data Set, a 5-minute aggregated truck GPS speed data set covering the national highway system for truck performance measure (FHWA 2013). As truck GPS data is increasingly available to transportation agencies and researchers, there are emerging studies evaluating truck mobility and behaviors using truck GPS data. Ma et al. (2011) implemented a trip identification algorithm to identify truck trip originations and destinations, and retrieved truck trips from raw GPS data which was reported from approximately 5,000 trucks traveling in the entire Washington State. They also developed an online platform to measure and report truck trip performance including speed, trip distance, travel time, and travel time reliability. Zhao and Goodchild (2011) employed the same dataset to measure truck travel time on freeways. The segment being studied was divided into several sub-segments with shorter distances. The travel time of each sub-segment was obtained by dividing the sub-segment distance by the average truck speed along the sub-segment. Travel time of the entire link was the sum of the travel time of each sub-segment. The result was compared with both empirical GPS observations and estimates based on loop detector data. It is found that the approach is sufficiently accurate to estimate truck travel time on freeways. Furthermore, truck behaviors were investigated using observed GPS data. Wang and Goodchild (2014) studied the impacts of tolling on truck speed and routing using truck GPS data. They found that truck speed along the toll road increased considerably after tolling while speed along the alternative free road decreased. A logit model was developed to identify the influential factors in truck routing, and the results reveal that drivers are willing to pay for a toll for a faster and more reliable route. Despite that there are studies using truck GPS data to study truck mobility and behaviors, none of them provides insight into how truck GPS data can be used to predict future truck travel time. To bridge this gap, this chapter proposes a pragmatic approach to estimate future truck travel time in response to traffic changes using truck GPS data and loop data. The logic of this approach is based on multi-regime relationships between truck speed and segment density. Cluster analysis was employed to segment traffic regimes. Future truck travel time could be estimated in response to segment density changes. The predicted travel time can be used to estimate travel time changes associated with freight investments or other planning practices.

3.2 Existing Travel Time Prediction Approaches

There is considerable research being done on predicting future travel time. These approaches can be classified into two categories based on their applications: short-term (real time) travel time prediction for traffic operation purposes and long-term travel time estimation for transportation planning purposes. A great deal of recent research has been targeted at developing short-term travel time prediction models using statistical techniques and mathematical modeling approaches, including time series (D'Angelo 1999), Kalman filtering (Chien 2003), artificial neural networks (Van Lint 2005), and Markov chain (Yeon 2008). Most of these approaches require current traffic conditions and historical observations, as well as considerable computing resources to develop predictions for real-time traffic operations. The objective of this chapter is to propose an approach that can support long-term freight project prioritization and planning, not real-time operations, and therefore, the literature review emphasizes travel time prediction over a longer time horizon.

One of the most straightforward methodologies for longer time travel time estimation is the use of speed and volume-capacity ratio (V/C) relationship. It has been applied extensively in various project benefit-cost analysis tools (McFarland 1993, Dowling Associates 2000). The speed is predetermined and changes in response to various V/C, facility type and speed limit. This engineering relationship is simple but not always accurate. In addition, it does not capture any network effects when additional traffic is attracted to the improved segments from other roads. Equilibrium traffic assignment methods address this issue, by assigning traffic to the network based on the predefined cost functions. The entire system reaches an equilibrium status assuming all vehicles travel along the minimum cost path. For instance, the Freight Analysis Framework version 3 (FAF3) freight traffic analysis developed by Battelle (2011) uses this method to assign freight traffic flow to the national highway network. The FAF3 employs the BPR function as the cost function for the stochastic user equilibrium traffic assignment procedure, as shown in Equation 3-1.

$$TT_{BPR} = TT_{ff} \times [1 + \alpha (x)^{\beta}] \qquad \text{Equation (3-1)}$$

where TT_{BPR} = segment traversal time estimated using the BPR function,

TT_{ff} = segment vehicle travel time at free flow speed,

x = volume-capacity ratio,

α and β are determined by facility type, free-flow speed and speed at capacity.

According to the Highway Capacity Manual (HCM 2000), the freeway free flow speed is calculated based on the information of number of lanes, lane width, shoulder width, and interchange density. Segment capacity is defined as number of vehicles during one hour under free-flow condition, and determined by facility type and free-flow speed.

The parameter α of the BPR function influences the ratio of free-flow speed to the speed at capacity. The parameter β determines how sensitive the speed drops when v/c is close to 1.0 (Dowling et al 1998). Given the characteristics of the two case studies of this chapter, α and β are assigned to 0.15 and 4 respectively.

The BPR function assumes travel time has a linear relationship with volume-capacity ratio. The model was developed by fitting data collected on uncongested freeways, and does not capture the travel time under congestion condition. To overcome the inaccurate prediction of oversaturated condition, Ak çelik developed a time-dependent travel time prediction function based on the steady-state delay equation for a single channel queuing system, and this model was recommended by the HCM 2000 for predicting vehicle travel time for planning purposes (Ak çelik et al 1991, HCM 2000). The model is shown in Equation 3-2.

$$TT_{Ak\c{c}elik} = TT_{ff} + 0.25T[(x-1) + \sqrt{(x-1)^2 + \frac{16JL^2}{T^2}}] \qquad \text{Equation (3-2)}$$

where $TT_{Ak\c{c}elik}$ = segment traversal time predicted using the Ak çelik function,

T = expected duration of demand (typically 1 hour),

L = segment length (mile)

J = calibration parameters determined by facility type, signal per mile, free-flow speed and speed at capacity (exhibit 30-4, HCM 2000)

Although the above two travel time prediction equations are extensively employed in travel demand models to estimate vehicle speed and travel time in response to various traffic volumes, neither of them is a truck specific model. As a result, there exist considerable deviations between the truck travel estimates and actual truck travel times. To solve this issue and predict reasonably accurate truck travel time for freight planning, this chapter proposes an approach to forecast truck travel time based on empirical truck GPS observations.

3.3 Proposed Methodology

This section discusses data on which this research is based and the proposed freeway truck travel time prediction approach. The proposed approach predicts future truck travel time based on the relationship between truck speed and density, which were retrieved from truck GPS data and dual-loop detectors respectively. The k-means cluster analysis algorithm was selected to partition data into homogeneous groups based on the characteristics of different traffic regimes.

4.3.1 Data Preparation

Two traffic datasets from different locations were collected to demonstrate the proposed approach: Interstate-5 (I-5) northbound between milepost (MP) 158 and 161 in the City of Seattle, WA and Interstate-405 (I-405) northbound between MP 8 and 10 in the City of Bellevue, WA.

Truck Speed

Truck speed used in this research was retrieved from GPS devices equipped on commercial vehicles traveling along the two selected segments. Data was collected anonymously between

May 2012 and July 2012. The GPS data was reported every 2-15 minutes. Information provided by GPS data includes a unique device ID, latitude and longitude, instantaneous truck speed, truck heading direction, and timestamp (time and date). Data was cleaned and geocoded to the freeway network in the ArcGIS environment. More details of data processing can be found in (McCormack 2011). GPS data was aggregated into 1 hour bins for each freeway segment to get average truck speed along the link.

Roadway Density

Roadway density was obtained by dividing traffic volume by truck speed. Traffic volume was collected by dual-loop detectors deployed in the right-most lane. The raw loop data provides traffic count every 20 seconds. Traffic count data was also aggregated into every 1 hour. Case study I contains six loop detectors deployed at MP 158.21, 158.92, 159.2, 159.96, 160.4 and 160.97. Case study II contains five loop detectors deployed at MP 8.03, 8.4, 8.9, 9.36 and 9.75. Traffic volume was estimated as the averaged value of loop detector collections along the segment.

4.3.2 Travel Time Predication Approach

The approach consists of 4 major steps:

1. Classify clusters based on the characteristics of truck speed and segment traffic volume using k-means algorithm,
2. Fit speed-density relationships,
3. Estimate freeway truck travel time, and
4. Evaluate estimation accuracy.

To evaluate the results, the estimates are compared with travel time calculated based on empirical truck speed observations, BPR function outputs and Akçelik model outputs.

Identify Clusters

Existing traffic flow studies have observed that traffic data shows two clear phases: free-flow and congested phases. In the free-flow phase, vehicles move at their desired speed and there is little influence/interaction between vehicles. In the congested phase, the traffic volume on the segment approaches capacity, and vehicles speed declines. Recent studies have also identified a transitional phase, called the intermediate phase (Kerner 1996). In the intermediate phase, vehicles experience stop-and-go driving conditions and are forced to drive as part of the overall traffic. Both two-regime and three-regime traffic models have been proposed in the literature. The first two-regime traffic flow model was proposed by Edie (1961), in which, the free-flow regime was fitted using the Underwood model and the congestion-flow regime was represented by Greenberg model, as shown in Equation 3-3.

$$u = \begin{cases} 54.9\exp(-k/163.9) & for \quad k \le 50 \\ 26.8\ln(162.5/k) & for \quad k \ge 50 \end{cases} \qquad \text{Equation (3-3)}$$

where u = vehicle speed (mph)

k = traffic density (vehicles per lane per mile)

Drake et al. (1967) developed a three-regime traffic model based on the Greenshields-type linear model for all three regimes, as given in Equation 3-4.

$$u = \begin{cases} 50 - 0.098k & for \quad k \le 40 \\ 81.4 - 0.913k & for \quad 40 \le k \le 65 \\ 40 - 0.265k & for \quad k \ge 65 \end{cases} \qquad \text{Equation (3-4)}$$

While these multi-regime models substantially improve the capability to capture different traffic characteristics under various traffic conditions, one of the major challenges of proposing such models is to determine the breakpoints between regimes (Sun and Zhou 2005). In the literature, most density breakpoints were determined by the researchers' engineering experience, which is subjective and biased by the judgment of model developers. Sun and Zhou (2005) employed the Cluster Analysis method to determine the breakpoints automatically given the fact that data belongs to the same cluster share similar features and data with different features belong to different groups. This chapter also employs a cluster analysis method to determine the breakpoints. Cluster analysis is a methodology to classify samples into a number of groups using a quantitative measure of association. The k-means algorithm is chosen in this study to identify traffic clusters. This algorithm is a centroid-based clustering algorithm, which aims to find the k cluster centers and assign the data to the nearest cluster center whose mean yields the least within-cluster sum of squares (Hartigan 1975). The k-means algorithm requires that the number of clusters is predetermined by modelers. The cluster analysis was accomplished using the R software package "cluster" (R Software 2014).

Fit Truck Speed-Density Relationships

For each cluster, the corresponding speed density relationship is fitted by minimizing squared errors. According to empirical observations, the speed-density relationships usually follow linear, logarithmic and exponential relationships (Sun and Zhou 2005), and the appropriate format to fit the data is determined based on the adjusted R-squared values. The one with the greatest R-squared value is chosen to represent the speed-density relationship of the empirical observations.

Estimate Truck Travel Time

Truck travel time is estimated by dividing segment distance by speed predicted on the speed-density relationships. It is assumed that trucks travel at a constant speed along the segment. This assumption is reasonable when the segment is short and maintains similar features, including both traffic volume and roadway geometric characteristics. This approach has been proved to be a reliable method by comparing the travel time estimates with empirical observations (Zhao 2011).

Evaluate Results

Mean absolute percentage error (MAPE), which is widely used as a measure to quantify the difference between the estimated value and the observed value, is chosen to evaluate the accuracy of the prediction, as shown in Equation 3-5. In this study, the observed travel time is defined as the estimates obtained by dividing segment distance by average truck speed from GPS data. The MAPE value of the proposed approach is compared with the MAPE values of the BRP function and the Akçelik function. A lower MAPE value represents more accurate prediction of truck travel time.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{TT_i - TT_i'}{TT_i}\right| \times 100\% \qquad \text{Equation (3-5)}$$

where n = total number of examples,

TT_i = observed travel time,

TT_i' = model predicted travel time.

3.4 Case Studies

4.4.1 Case Study I

A 3-mile stretch of northbound I-5 in the City of Seattle, WA between MP 158 and MP 161 was selected as case study I. Both truck GPS data and loop data were collected between May 2012 and July 2012. The data set was divided into a training set (May 2012 and June 2012) and a testing set (July 2012). Truck speed along the segment was retrieved from GPS data. Traffic volume was calculated as the averaged traffic volume recorded by the six dual loop detectors. Density was obtained by dividing traffic volume by truck speed. Figure 3-1 displays the truck speed-density plot of the training dataset.

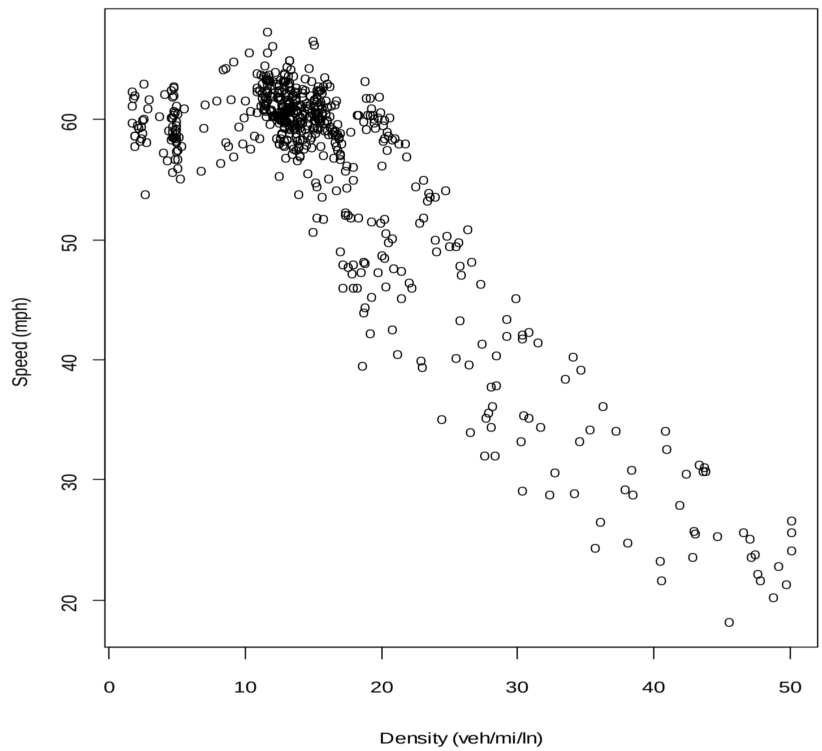


Figure 3-1 Case Study I Truck Speed-Density Plot

As shown in Figure 3-1, trucks maintain a constant speed around 60 mph when segment density is less than 10 vehicles/mile and speed drops significantly while density increases, with the lowest observed speeds of approximately 20 mph. The K-means algorithm was employed to classify dataset into different clusters representing various traffic regimes. It is clear from Figure

3-1 that there are at least two traffic regimes, and may be more as the speed decreases at different rates with the increase of density. The appropriate number of clusters is often ambiguous, and depends on the distribution of observations in a dataset and the desired resolution of the user. Meanwhile, the number should not be too many for convenient use of the model. Thus the authors conducted the cluster analysis with two clusters and three clusters respectively, and compare the results in the following sections.

Two Clusters

Figure 3-2 and Table 3-1 present the clustering results when there are two clusters. The first cluster characterizes the free-flow traffic regime, in which trucks travel at around 60 mph when segment truck density is less than 10 vehicles/mile. The clustering result shows that the average truck speed of cluster 1 is equal to 60 mph, and average density is 4.87 vehicles/mile. The second cluster represents the non-free flow condition where truck speed starts to decrease when density is greater than 10 vehicles/mile and drops continuously with the increase in segment density. The average truck speed and segment density of the second cluster are 54.63 mph and 18.46 vehicles/mile respectively. It should be noted that the cluster numbers here are only used to identify each specific cluster.

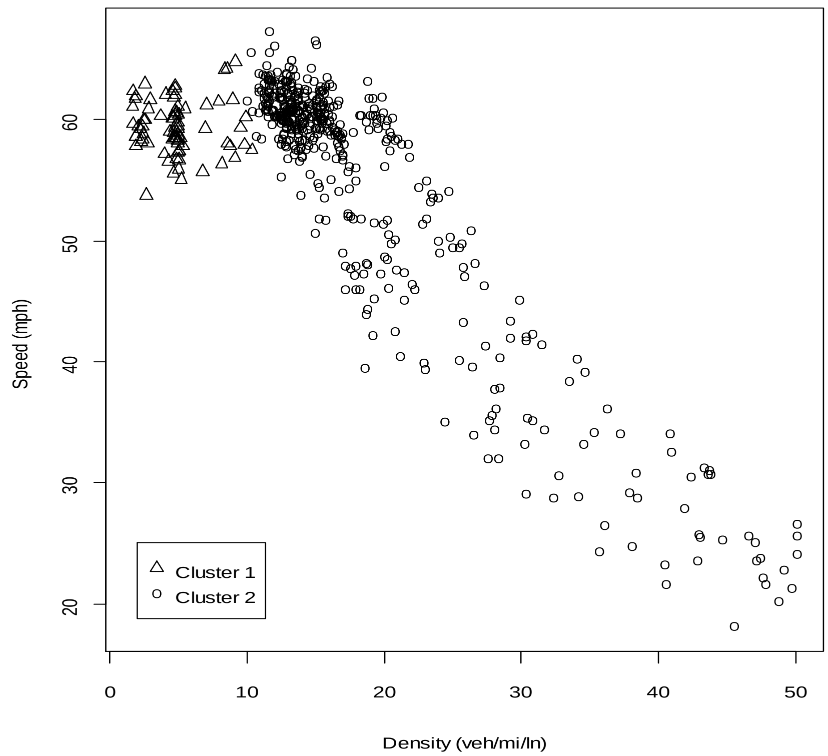


Figure 3-2 Case Study I Two Clusters Truck Speed-Density Plot

Table 3-1 Case Study I Cluster Centers of Two Clusters Analysis

	Cluster 1	*Cluster 2*
Truck Speed (mph)	60	54.63
Density (vehicles/mile)	4.87	18.46

For cluster 1, trucks travel at the average of 60 mph regardless of the segment density. For cluster 2, truck speed is a dependent variable of density. The authors fitted the data using linear, logarithmic and exponential models which were applied in the rational speed-flow relationships shown in Equation 3-3 and 3-4. It is found that the exponential function provides the best fit of the observed data with the greatest R-squared value, and the regression results are summarized in Table 3-2. All parameters are significant with P-values less than 0.0005. The truck speed-relationship of the test dataset is given in Equation 3-6.

Table 3-2 Case Study I the Second Cluster Fitted Results of Two Clusters Analysis

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.478	0.009	469.177	<0.0005
Density	-0.027	0.000	-58.459	<0.0005

$$\begin{cases} u = 60 & \textit{for } k \leq 10 \\ u = \exp(4.478 - 0.027k) & \textit{for } k > 10 \end{cases} \qquad \text{Equation (3-6)}$$

Data collected in July 2012 was used to evaluate the proposed approach. Hourly traffic volume was retrieved from loop detector data and averaged hourly truck speed was calculated from truck GPS data. Truck travel time obtained from dividing the segment distance by observed truck GPS speed was used as the ground truth travel time to evaluate the accuracy of the proposed approach. The authors also employed the BPR function and Ak çelik function to estimate travel time, and compared with the ground truth travel time to calculate the corresponding MAPE values and evaluate the accuracy of each method. As shown in Table 3-3, the MAPE value of the proposed speed-density based approach is 6.16%, less than the MAPE values of BPR and Ak çelik methods of 11.52% and 11.60% respectively. This result indicates that the proposed approach generates less deviation between travel time estimates and observations, and therefore performs better than the existing BPR method and Ak çelik method.

Table 3-3 Case Study I MAPE Values of Each Travel Time Prediction Method

	MAPE value
Speed-density method (two clusters)	6.16%
BPR method	11.52%
Ak çelik method	11.60%

Three Clusters

Figure 3-3 and Table 3-4 show the clustering results with 3 clusters. Similar to the two clusters results, cluster 1 represents the free-flow traffic regime, in which traffic density is low and truck travel at about 60 mph when density is less than 11 vehicles/mile. The speed is constant and not affected by density. The average truck speed and density of cluster 1 is 60mph and 4.94 vehicles/mile respectively. Truck speed in cluster 2 and cluster 3 decreases considerably with the increase of density. Cluster 2 features a high speed and intermediate density phase when density is between 11 and 25 vehicles/mile, and cluster 3 characterizes a low speed and high density congested phase when density is greater than 25 vehicles/mile. For cluster 2, the average speed and density are 58.77mph and 15.26 vehicles/mile. For cluster 3, the average speed and density are 32.33 mph and 35.71 vehicles/mile.

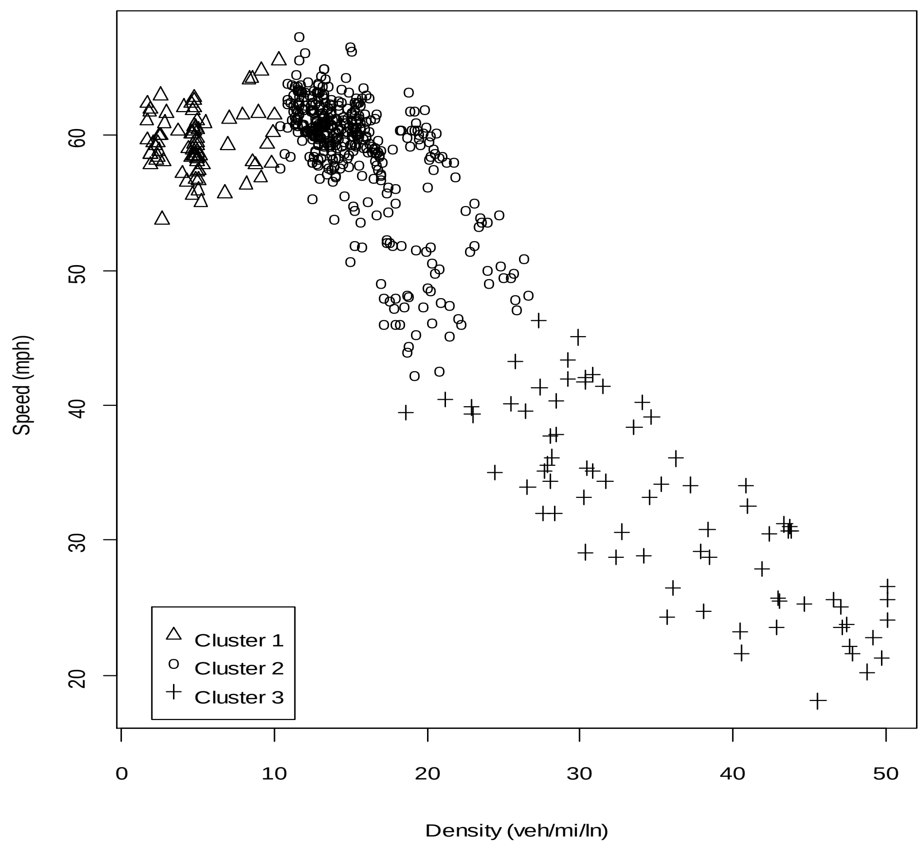


Figure 3-3 Case Study I Three Clusters Truck Speed-Density Plot

Table 3-4 Case Study I Cluster Centers of Three Clusters Analysis

	Cluster 1	*Cluster 2*	*Cluster 3*
Truck Speed (mph)	60	58.77	32.33
Density (vehicles/mile)	4.94	15.26	35.71

As shown in Figure 4-3, the rate at which speed decreases differs between cluster 2 and 3. The linear, logarithmic and exponential models were tested to fit the cluster 2 and 3 data. It is found that the linear function fits cluster 2 data best and exponential function fits the cluster 3 data best, the fitting results are presented in Table 3-5. All parameters are statistically significant. Truck speed-density relationships are given in Equation 3-7.

Table 3-5 (a) Case Study I Second Cluster Fitted Results and (b) Third Cluster Fitted Results

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	72.709	0.794	91.569	<0.0005
Density	-0.914	0.051	17.975	<0.0005

(a)

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.238	0.068	61.926	<0.0005
Density	-0.022	0.002	11.804	<0.0005

(b)

Truck speed and density relationship:

$$\begin{cases} u = 60 & for\ k \le 10 \\ u = 72.709 - 0.914k & for\ \ 11 < k < 25 \\ u = \exp(4.238 - 0.022k) & for\ k \ge 25 \end{cases} \qquad \text{Equation (3-7)}$$

Similar to the previous analysis, the authors evaluated the proposed model using the test dataset and calculated the MAPE values. The MAPE value of the proposed approach is 5.55% as shown in Table 3-6. This value is less than the corresponding values of the BPR method and the Akçelik method, which are 11.52% and 11.60% respectively. This result reveals that the proposed approach generates, by a substantial margin, more accurate results than the other two methods.

Table 3-6 Case Study I MAPE Values of the Selected Travel Time Prediction Methods

	MAPE value
Speed-density method (three clusters)	5.55%
BPR method	11.52%
Akçelik method	11.60%

By comparing Table 3-3 and Table 3-6, the two clusters and three clusters analysis results show that the MAPE value is improved from 6.61% to 5.55%. While the three clusters approach provides a slightly more accurate result, it also requires a greater data analysis effort. While the user is entitled to choose the number of clusters appropriate for their study, for this case study, no significant improvement is observed when using three clusters instead of two clusters, and the case study is carried forward with the two clusters approach.

4.4.2 Case Study II

Case study II is a 2-mile segment of I-405 northbound between MP 8 and 10. Traffic volume was the averaged value of data collected by the five loop detectors deployed along the rightmost lane. The speed-density plot is displayed in Figure 3-4. Similar to case study I, trucks travel at a constant speed in free-flow traffic pattern. Truck speed decreases when density is greater than 20 vehicles/mile. Both two clusters and three clusters analyses were performed to identify the appropriate number of clusters for this dataset.

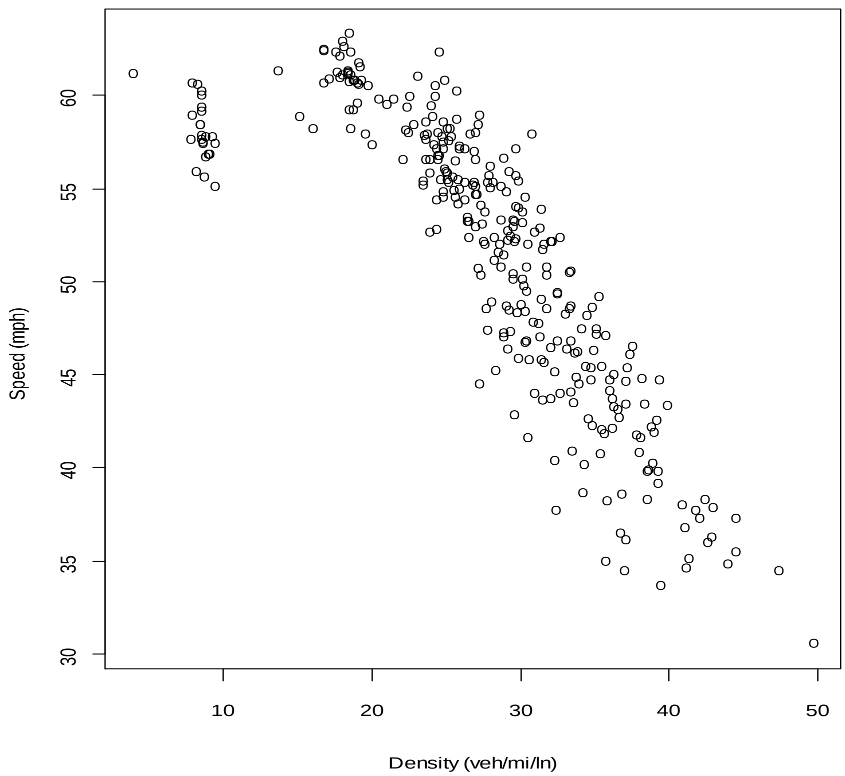


Figure 3-4 Case Study II Truck Speed-Density Plot

Two Clusters

Figure 3-5 and Table 3-7 present the clustering results with two identified clusters. Cluster 1 features free-flow phase, in which trucks travel at a constant speed. According to the cluster analysis result, the average speed and density of cluster 1 is 58 mph and 9.12 vehicles/mile. For cluster 2, truck speed starts to decline when density is greater than 16 vehicles/mile. The average speed and density are 50.88 mph and 29.39 vehicles/mile respectively.

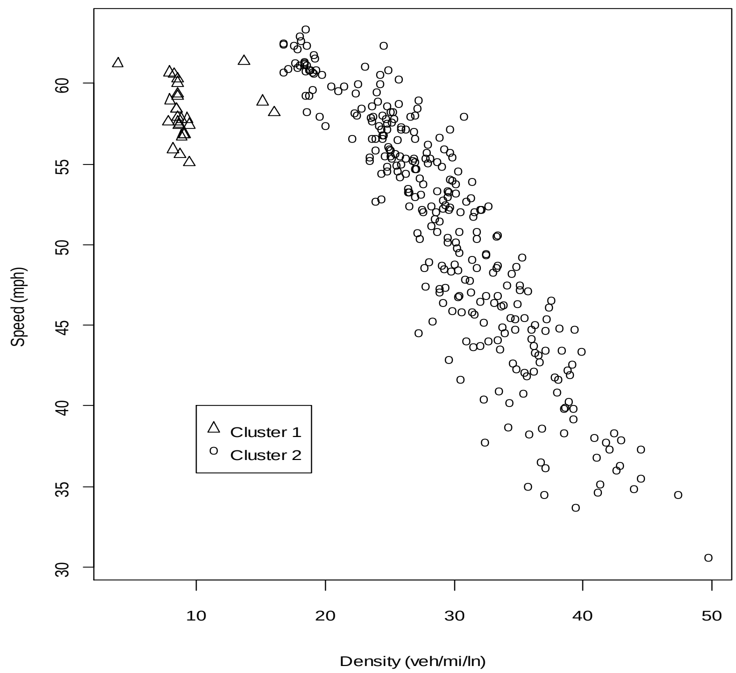


Figure 3-5 Case Study II Two Clusters Truck Speed-Density Plot

Table 3-7 Case Study II Cluster Centers of Two Clusters Analysis

	Cluster 1	*Cluster 2*
Truck Speed (mph)	58	50.88
Density (vehicles/mile)	9.12	29.39

To fit the data of cluster 2, the linear, logarithmic and exponential models were tested, and the adjusted R-squared values of each model indicate that the linear model provides the best fit. The model results are presented in Table 3-8. The truck speed-density relationship is given in Equation 3-8.

Table 3-8 Case Study II the Second Cluster Fitted Results of Two Clusters Analysis

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	81.88	0.82	99.70	<0.0005
Density	-1.05	0.03	-38.65	<0.0005

$$\begin{cases} u = 58 & for\ k \le 16 \\ u = 81.88 - 1.05k & for\ k > 16 \end{cases} \qquad \text{Equation (3-8)}$$

The MAPE values of the proposed approach, BPR function and Ak çelik function were calculated and the results are summarized in Table 3-9. The proposed approach generates the least MAPE value, and therefore performs better than the other two approaches.

Table 3-9 Case Study II MAPE Values of the Selected Travel Time Prediction Methods

	MAPE value
Speed-density method (two clusters)	7.33%
BPR function	13.75%
Ak çelik method	14.60%

Three Clusters

The three clusters were tested, and the speed-flow plot is displayed in Figure 3-6 and Table 3-10. While cluster 1 characterizes the free-flow traffic regime, cluster 2 represents the intermediate phase and cluster 3 features the congested phase. The authors tested linear, logarithmic and exponential models and find that none of them is able to delineate the dataset of cluster 2 and 3 well given the R-squared values are all less than 0.5. Thus the authors concluded that two clusters analysis is better than three clusters for this specific data of case study II.

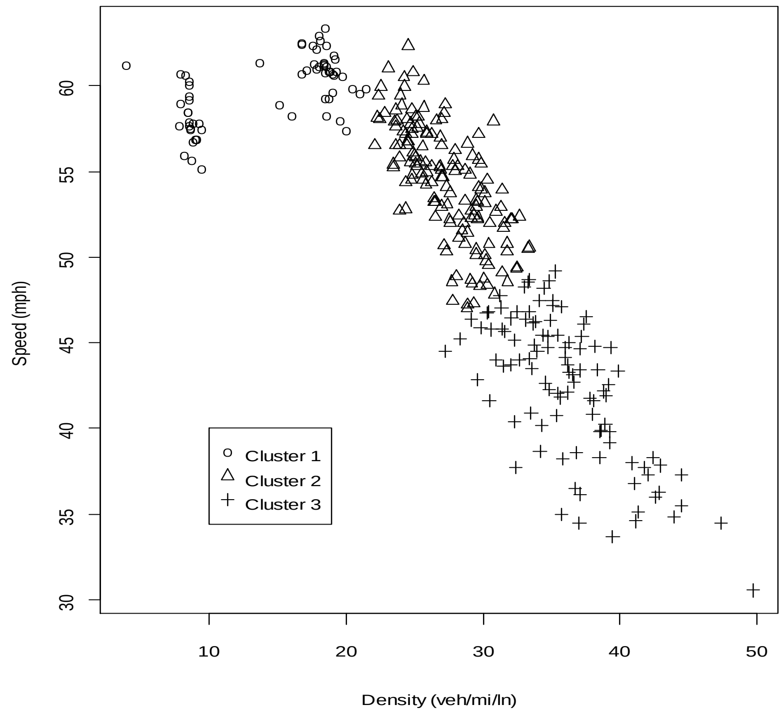


Figure 3-6 Case Study II Three Clusters Speed-Density Plot

Table 3-10 Case Study II Cluster Centers of Three Clusters Analysis

	Cluster 1	*Cluster 2*	*Cluster 3*
Truck Speed	59.73	54.35	42.46
Density	14.49	27.31	36.06

The above two case studies illustrate how the proposed multi-regime speed-density based approach can be used to forecast truck travel time. The analysis results indicate that the proposed approach is superior to the traditional BPR method and Akçelik method, and is able to forecast more accurate travel time. The number of clusters can be determined by both the distribution of data and the desired resolution of the user. The increase of number of clusters is able to improve the travel time prediction accuracy, but will involve additional data processing efforts and model application complexity. For both case studies, two clusters are able to provide substantial improvements over current methods used to predict truck travel time.

Despite the fact that both case studies considered freeways in the Puget Sound, the speed-density relationships shown in Equations 3-6 and 3-8 are different. Figure 3-7 shows the speed-density plot of the two datasets. It is noted that two case studies have distinct speed-density relationships. For case study I, speed starts to decline when density is greater than 10 vehicles/mile, while the breakpoint of case study II is around 16 vehicles/mile. Further, when density exceeds the breakpoint, dataset 1 has a convex shape and an exponential model provides the best fit, while dataset 2 displays a straight and linear relationship. The deviation of the speed-density distributions is associated with several characteristics of each segment, including roadway geometric features and travel demand distribution. Thus, it is less accurate to select one model to fit both datasets. Given this, and the simplicity of the approach, we recommend users to apply the clustering and best-fit modeling approach, and develop their own equations for different locations.

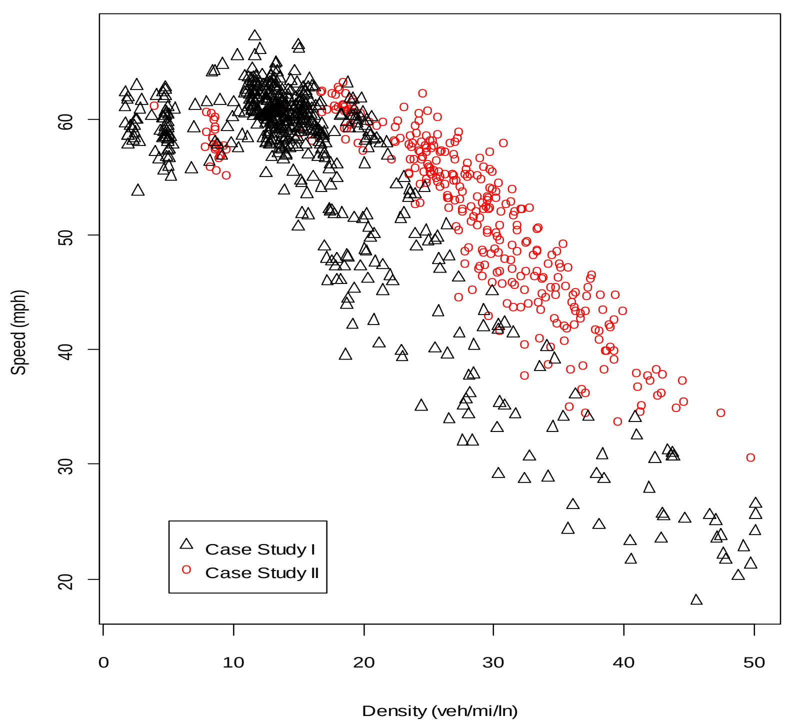


Figure 3-7 Speed-Density Plot of Case Study I and Case Study II

3.5 Conclusions

This chapter proposes a multi-regime speed-density relationship based approach to predict freeway truck travel time using empirical truck probe GPS data and loop detector data. The K-means cluster analysis algorithm was employed to determine the breakpoints of different traffic regimes. Each cluster was fitted using linear, logarithmic and exponential models, and the model with the highest R-squared value was selected. The parameters of the best models for both cases are all statistically significant. The travel time estimates were compared with estimates calculated based on empirical truck GPS speed data, and the mean absolute percentage error was calculated. This was compared with the BPR model and the Akçelik model. It is found that the new approach is able to estimate more accurate travel times than traditional methods given it generates the least MAPE values for both case studies.

The authors investigated the appropriate number of clusters when segment the data using the K-means algorithm. For case study I, the two-cluster identification is recommended since it is easier to use and still provides reasonably high accuracy of estimates. For case study II, the two-cluster identification is recommended as well since the commonly applied speed-density formats do not fit the three-cluster clustering results. The analysis reveals that the number of clusters is determined by the distribution of data and the resolution desired by users. In the case studies evaluated, the more clusters are classified, the less deviation is obtained. The two clusters analysis is recommended when the improvement from two to three clusters is small.

The predicted travel time can support freight prioritization and planning. For instance, to forecast truck travel time associated with traffic density changes resulting from freight investments, one can apply this approach to generate the multi-regime speed-density relationships based on GPS and loop data, and estimate the corresponding post-project traffic density. The fitted speed –

density relationship was not tested with other segments. We recommend users applying this approach to develop their own equations as speed-density relationships vary depending on roadway geometric features, traffic demand distribution, traffic operation strategy and other factors.

Chapter 4 FREEWAY TRUCK TRAVEL TIME RELIABILITY FORECASTING

Despite the fact that there are considerable quantitative project prioritization applications, only a few of them consider travel time reliability. None of them includes truck-specific travel time reliability forecasting. Some applications employ regional travel demand models to quantify both pre- and post-investment performance. These travel models rely upon fixed inputs, and therefore cannot capture transportation system variability/reliability. Travel time reliability is rarely considered in project prioritization tools due to the deficiency of truck specific movement data and the complexity of both measuring and forecasting reliability. Due to the fact that truck GPS data is increasingly available to transportation agencies and researchers, there now exist opportunities to better measure and forecast truck travel reliability. The measurement of reliability has been discussed in Chapter 2. This chapter focuses on forecasting travel time reliability using truck GPS data.

4.1 Background

There have been multiple studies investigating how GPS data can be used to measure travel time reliability. Zhao et al. (2013) proposed a truck GPS spot speed distribution based approach. This work demonstrated that the mixture of two Gaussian distributions provides the best fit for the truck GPS spot speed observations. The probability density function of mixture of two Gaussian distributions is shown in Equation 4-1. The parameters are fitted based on the maximum likelihood rule.

$$f(x) = w \cdot n(x, \mu_1, \sigma_1) + (1-w) \cdot n(x, \mu_2, \sigma_2)$$
$$n(x, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right] \quad (4\text{-}1)$$

where w = the proportion of the first normal distribution,

μ_1 and μ_2 = mean of the first and second Gaussian distribution,

σ_1 and σ_2 = standard deviation of the first and second Gaussian distribution.

Furthermore, they developed a rule to classify reliability into three categories: reliably fast, reliably slow and unreliable. Travel condition is defined as unreliable if and only if $|\mu_1 - \mu_2| \geq |\sigma_1 + \sigma_2|, w \geq 0.2, \textit{ and } \mu_1 \leq 0.75 \times V_p$ (V_p is the posted speed), otherwise, it is viewed as reliable. If travel condition is defined as reliable and the average speed is less than 75% of the posted speed ($v \leq 0.75 \times V_p$), it is defined as reliably slow, otherwise, it is viewed as reliably fast. Wang et al. (2014) further improved the approach by computing the coefficient of variation (COV) of the fitted mixture of two Gaussian distributions. The calculation is given in Equation 4-2 and 4-3. The improved approach is able to provide a numerical value which would allow for a more quantitative evaluation and reliability ranking.

$$\mu = \sum_{i=1}^{n} w_i \mu_i$$
$$\sigma^2 = \sum_{i=1}^{n} w_i((\mu_i - \mu)^2 + \sigma_i^2) \quad \text{(Equation 4-2)}$$

$$Coefficient\ of\ Variation\ (COV) = \frac{\sigma}{\mu} \quad \text{(Equation 4-3)}$$

For forecasting travel time reliability, understanding how reliability changes in response to different traffic conditions is the key since speed and reliability are associated with traffic condition closely. Under free-flow condition, traffic volume is low and most of the trucks travel at a reliable desired speed, which is close to free-flow speed. During congestion phase, trucks stop and go due to the influence of other vehicles, and results in unreliable travel time.

The following section presents the changes of speed distribution in response to different traffic density along a selected segment.

4.2 Data and Analytical Process

The idea of forecasting reliability is to understand how speed changes with respect to different traffic situations. Future traffic conditions can be obtained from engineering experience, e.g. the reduction in traffic volume/density by adding one lane. We started with gathering historical spot speed distributions and roadway traffic data.

The same freeway segment utilized in Chapter 2 was selected as an example to show how speed distribution changes under different traffic conditions. The segment is a stretch of 3.5 miles of southbound Interstate 5 (I-5) through downtown city of Seattle, Washington. Both GPS data and loop data were collected between January 2012 and December 2012. GPS data was collected by GPS devices installed in commercial vehicles traveling along the segment of interest. The data was cleaned to remove duplicated and problematic records. The cleaned data was geocoded to the corresponding network to reflect truck travel speed along the segment within ArcGIS environment. Details of the data cleaning and preparation can be found in Zhao et al. (2011). The monthly GPS data was aggregated for one hour interval to generate hourly truck spot speed distribution. For each one hour dataset, spot speed was fitted using the mixture of two Gaussian distributions. The distribution fitting was accomplished within the R software, using the package called “mixdist” (Du 2002). The fitting process generates the mean values and standard deviations of the two normal distributions, and the probability of each distribution. Details can be found in Chapter 2.

Traffic condition was implied using traffic density computed based on traffic volume and speed information. Both traffic volume and speed were collected by the loop detector deployed along the segment being studied. The data was recorded at every 20-second interval. Similar to the GPS data, monthly loop data was aggregated for one hour interval to generate the average hourly traffic volume and hourly speed. Traffic condition was quantified using traffic density obtained by dividing traffic volume by speed.

As proved in Zhao et al. (2013), truck spot speed distribution follows a mixture of two Gaussian distributions, and it is either a unimodal or bimodal distribution. Truck travel time is classified as unreliable if spot speed distribution follows a bimodal distribution; otherwise travel time is reliable and the spot speed distribution follows a unimodal distribution. When travel time is reliable, most of the trucks travel at a constant and desired speed, and therefore the spot distribution follows a unimodal distribution. However, during the traffic congestion condition, trucks stop and go, and therefore the presence of another speed distribution emerges. The fluctuated speed generates a bimodal speed distribution representing two traffic regimes, the low speed regime and high speed one. The more unreliable the system is, the wider the speed distribution and the greater possibility of spot speed falling within the low speed regime.

Figure 4-1 displays the hourly spot speed distributions between 5 AM and 11 AM in May 2012 along the segment being studied. The fitted distribution parameters and the corresponding calculated hourly traffic density values are given in Table 4-1.

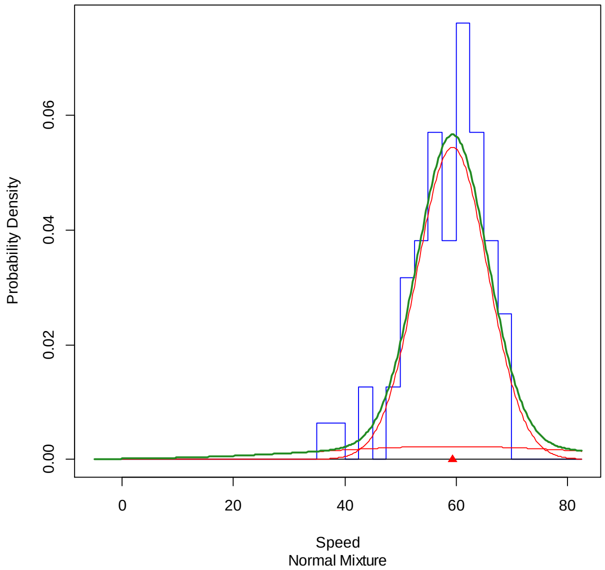
As shown in Figure 4-1 (a) and Table 4-1, between 5 AM and 6 AM, the average segment density was 11.4 veh/mi/ln, and therefore it was classified as level of service (LOS) A according to the Highway Capacity Manual (HCM) (HCM 2000). The spot speed distribution followed a unimodal distribution, and the predefined reliability rule identified the travel time as reliably fast. Most of the trucks traveled at a free-flow speed, and the average truck speed was around 60 mph. According to the Highway Capacity Manual, the traffic condition is classified as Level of Service (LOS) A since density is less than 15 veh/mi/ln. The description of LOS A indicates that this is the traffic condition of free flow. Traffic density is low, with uninterrupted flow speeds controlled by driver desires, speed limits, and physical roadway conditions. Drivers can maintain their desired speeds with little or no delay. The truck travel performance observed from the speed distribution is consistent with the description of LOS A.

Traffic density increased to 17.75 veh/mi/ln between 6 AM and 7 AM, and therefore fell into LOS B category. The presence of the second distribution emerged. Namely, not all trucks were able to travel at the desired speed, but some of them were influenced by others and traveled at a lower speed. The averaged truck speed of the first traffic regime was 39.99 mph, with standard deviation of 13.31 mph; the average speed of the second traffic regime was 58.55 mph, with standard deviation of 3.85 mph. The probability of truck travel speed falling within the low speed traffic regime was 17%. Travel time was still classified as reliable based on the predefined rule. These changes of speed distribution and truck performance are consistent with the transition from LOS A to LOS B. According to the definition of LOS B, drivers still have reasonable freedom to travel at their desired speeds, but there is a low probability that traffic flow will be restricted.

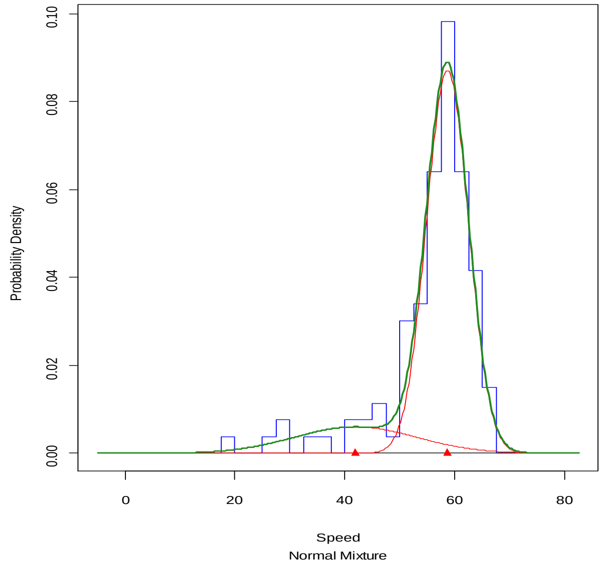
From Figure 4-1 (b) to Figure 4-1 (c), the probability of falling within the low speed regime increased to 47% with the increased traffic density of 22.55 veh/mi/ln, and a clear bimodal distribution was presented. Meanwhile, the average speed of the first traffic regime dropped to 20.76 mph, with standard deviation of 9.92 mph; the average speed of the second traffic regime decreased to 51.63 mph, with standard deviation of 7.87 mph. Travel time was changed to unreliable as evaluated by the predefined rule. The travel time reliability was reduced with the growth of traffic density. The density of 22.55 veh/mi/ln indicates the LOS C condition. The definition of LOS C pointed out that most drivers are restricted in selecting their own speed and driving behaviors, but closely controlled by the higher volume. The speed observation is consistent with the LOS C definition.

Traffic density did not change significantly between 8 AM – 9 AM and similar speed distribution was observed. The probability of falling within low speed regime is equal to 58%. Travel time was identified as unreliable.

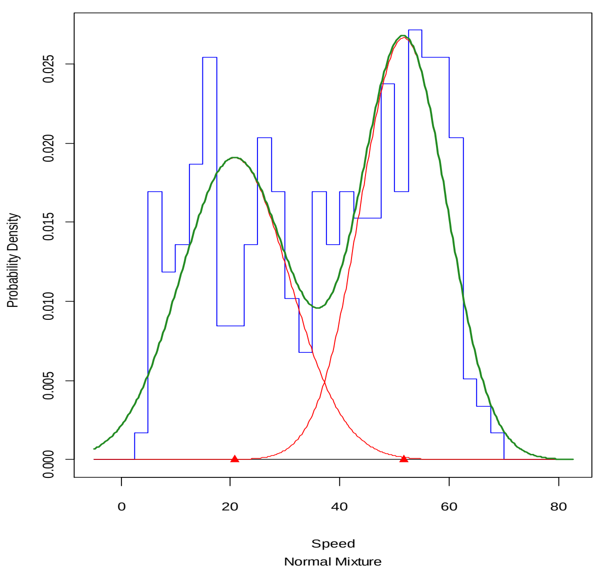
The low speed traffic regime started to wane since 9 AM while average traffic density decreased to 18.91 veh/mi/ln. The spot speed distribution showed the trend to change back to unimodal distribution. Travel time between 9 AM and 10 AM, and 10 AM and 11 AM were both reliable.



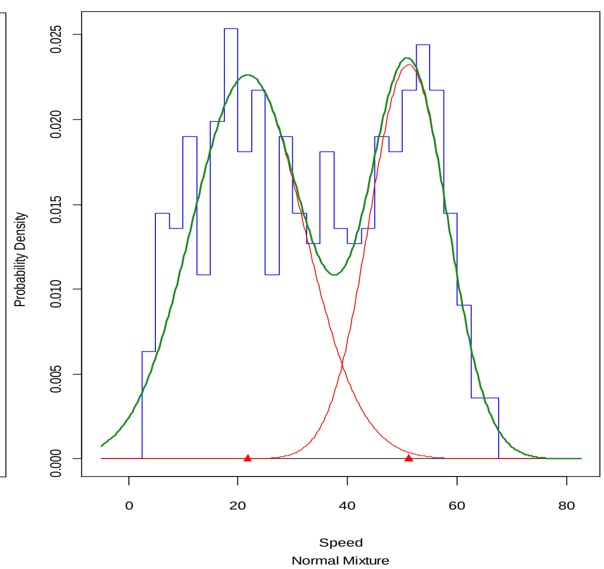
(a) 5 AM-6 AM



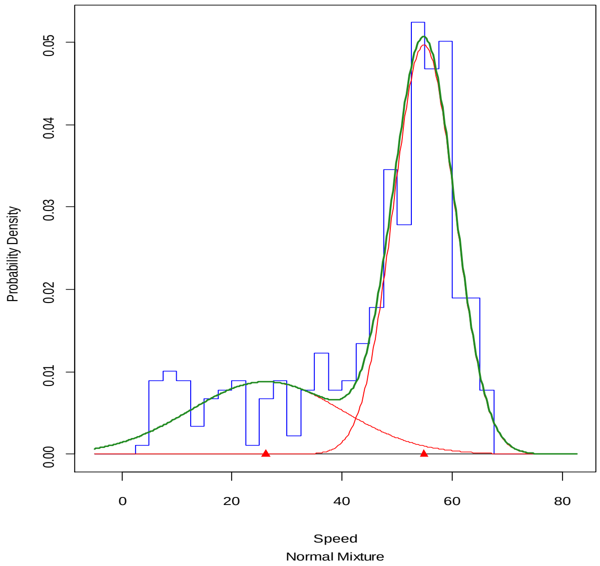
(b) 6 AM-7 AM



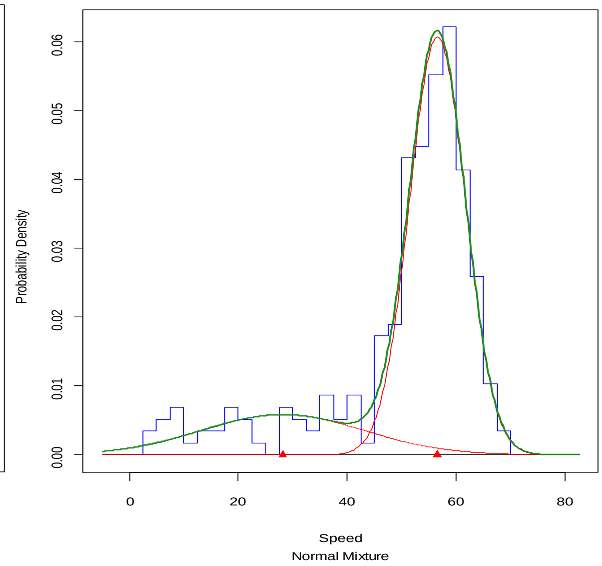
(c) 7 AM-8 AM



(d) 8 AM-9 AM



(e) 9 AM-10 AM



(f) 10 AM-11AM

Figure 4-1 Truck Spot Speed Distribution during Different Time-of-Day

Table 4-1 Truck Spot Speed Distribution Fitted Results

	5 AM	6 AM	7 AM	8 AM	9 AM	10 AM
Traffic density (veh/mi/ln)	11.40	17.75	22.55	22.51	18.91	19.21
w	0.14	0.17	0.47	0.58	0.30	0.22
μ_1	59.34	39.99	20.76	21.75	26.15	28.13
σ_1	24.33	13.31	9.92	10.28	13.77	14.93
μ_2	59.34	58.55	51.63	51.08	54.82	56.52
σ_2	6.32	3.85	7.87	7.16	5.59	5.14
V_p	60	60	60	60	60	60
Average speed	59.35	55.36	37.00	33.98	46.10	50.33
$if\ \lvert\mu_1-\mu_2\rvert \geq \lvert\sigma_1+\sigma_2\rvert, w \geq 0.2$ $and\, \mu_1 \leq 0.75 \times V_p$	No	No	Yes	Yes	No	No
$if\ \ average\ speed \leq 0.75 \times V_p$	No	No	Yes	Yes	No	No
Reliability category	Reliably Fast	Reliably Fast	Unreliable	Unreliable	Reliably Fast	Reliably Fast

The above analyses indicate that truck spot speed distributions vary in response to different traffic conditions. Traffic congestion causes fluctuated truck speed and consequently leads to unreliable travel time. Greater roadway density is associated with lower travel speed and lower reliability.

This analysis also reveals that the distribution COV, which is utilized to quantify reliability in this study, is strongly associated with segment density. As a result, travel time reliability could be forecasted based on the relationship between COV and density when future roadway density is available or predictable. The next section presents a case study to illustrate the process of establishing the relationship between COV and density to support reliability forecasting.

4.3 Case Study

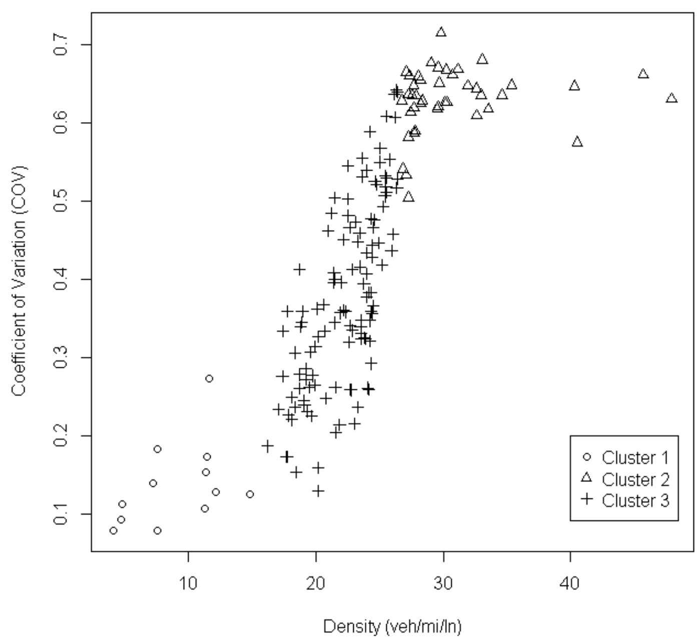


Figure 4-2 K-Means Analysis Results of the Segment Density and Spot Speed Distribution COV Plot

Figure 4-2 displays the hourly segment density and hourly speed distribution COV based on one year observations between January 2012 and December 2012. Higher COV indicates lower travel reliability. The figure illustrates that higher traffic density is associated with lower travel reliability. The impacts of density on COV and reliability vary depending on the value of density.

When density was between 15 veh/mi/ln and 28 veh/mi/ln, COV increased considerably with the growth of density. COV did not change substantially when density was greater than 28 veh/mi/ln.

There were three clear clusters representing different relationships between density and COV. Thus a cluster analysis was conducted to identify the breakpoints and segment data into three clusters, denoted as 1st, 2nd and 3rd cluster as shown in Figure 4-2. The K-means cluster algorithm was selected and the clustering process was accomplished using the R software. The three clusters represent different traffic regimes.

For the first cluster, density was less than 18 veh/mi/ln and the corresponding COV was less than 0.2 (excluding one outlier). According to the rule developed to evaluate travel time reliability by Zhao et al., the truck travel time of the first cluster was assessed as reliable. This result is consistent with the LOS definition, in which LOS A and B are defined as free flow and near free flow conditions when density is less than 18 veh/mi/ln. Travelers travel at the desired or near-desired speed and are not influenced by other travelers under LOS A and B. When density reaches the value of 28 veh/mi/ln, it is classified as the third cluster. The COV does not change significantly with the growth of density, and stays between 0.6-0.7.

The COV increased considerably with the growth of density in the second cluster. The author fitted the data using linear regression, and obtained the relationship between COV and density as presented in Equation 4-4 and Table 4-2. Both the intercept and density are significant. The sign of density is positive, which confirms the adverse impact of density on reliability. The adjusted R-squared value is 0.68.

$$\begin{cases} COV = -0.442 + 0.037 \times density & 15veh/mi/ln < density < 28veh/mi/ln \\ COV = 0.631 & density \geq 28veh/mi/ln \end{cases}$$ Equation 4-4

Table 4-2 The Second Cluster Fitted Results

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.442	0.049	-8.937	<0.0005
Density	0.037	0.002	17.624	<0.0005

For this case study, if density is less than 15 veh/mi/ln, vehicles experience reliable travel time and travel at a speed of 60 mph. The system is unreliable when density is greater than 15 veh/mi/ln. Reducing segment density can improve reliability considerably when density is between 15 veh/mi/ln and 28 veh/mi/ln. When density is greater than 28 veh/mi/ln, moderate density reduction does not affect reliability substantially; notable improvement can be observed only until density is reduced to less than 28 veh/mi/ln.

This cluster analysis based approach is expected to be applied for forecasting travel time reliability and support project prioritization, and to be conducted individually for each segment of interest using data from the segment itself. Potential changes of segment density resulting from highway investments can be estimated based on engineering experience and established models. The project prioritization process should also include sensitivity analysis to assess impacts of different density changes on future travel time reliability.

4.4 Conclusion

This chapter proposes an approach to forecast freeway vehicle travel time reliability for planning purposes using GPS data. The proposal is a methodology that can be applied to any segment of interest for which GPS data is available. The authors analyze the changes in truck GPS spot speed distribution in response to different traffic conditions. The analysis reveals that traffic has considerable impact on speed distribution, and a bimodal distribution emerges with increasing traffic density. Greater speed distribution COV is closely associated with greater segment density, but the relationships are dependent on the traffic regime. A cluster analysis based approach is proposed to segment the COV and density dataset into several groups and identify the breakpoints. The authors further quantify the impact of density on COV by fitting the data using a linear model. The developed equation is able to forecast segment reliability with changes of density.

Given the simplicity of this approach and the increasingly available vehicle GPS data, we recommend users to apply this proposed approach, and develop their own equations for different locations for project prioritization and planning.

Chapter 5 GPS DATA ANALYSIS OF THE IMPACT OF TOLLING ON TRUCK SPEED AND ROUTING

In the past decades, there have been multiple studies seeking to collect truck movement data to support freight planning, including identifying truck travel patterns, supporting travel demand model improvement, and evaluating impacts of freight policies. This chapter explores the capability of truck GPS data in evaluating tolling impacts on truck travel performance and routing. Truck GPS data contains the empirical responses to tolling following the implementation of a toll on the State Route 520 (SR-520) bridge in Seattle. Thus it is used to evaluate route choice and mobility performance along SR-520 and the alternate toll-free route I-90.

5.1 Background

Roadway tolls are designed to raise revenue for funding transportation investments and manage travel demand (AASHTO 2010). The tolling strategies might affect both truck speed and route choice. On one hand, tolling adds additional costs to goods movement, which may result in truck traffic diversion from toll roads to toll-free roads. On the other hand, tolling may reduce travel time and improve reliability along the toll road, which in turn saves logistics costs, and therefore may attract trucks to toll roads. Thus trucks may switch routes between toll roads and free roads based on the trade-offs among a set of time- and cost-related attributes.

Extensive studies of tolling impacts on passenger travel have revealed that travelers' responses include rescheduling trips, canceling trips and consolidating trips, as well as changing routes. However, commercial trips have less flexibility for rescheduling, canceling and consolidation trips due to customers' requirements (Roth 2003). A survey developed to examine the impacts of the Port Authority of New York and New Jersey's time-of-day pricing found that 68.9% of for-hire carriers and private carriers cannot change their schedule due to customer requirements, and only 0.5% of trucks would switch to off-peak hour delivery (Holguin-Veras et al. 2006). Meanwhile, the results showed even less ability to consolidate trips, with only 0.1% of carriers able to increase shipment size, 0.4% of carriers able to wait for longer for mixing pickups and 0.1% of carriers able to cut some of their runs (Holguin-Veras et al. 2006).

The diverted truck traffic has significant impacts on regional traffic safety, travel performance (e.g. travel speed), the environment (emissions), as well as toll revenue and the regional economy. For instance, the Ohio Turnpike caused a 30% to 50% increase in truck traffic on local routes, and imposed significant safety and environment costs (Swan and Belzer 2007). Thus it is critical to understand the effects of tolling on truck performance and route choice. Yet limited research has quantified the impacts of tolling on truck speed and trip diversion due to the lack of detailed truck data. In light of this, this dissertation uses truck GPS data to observe truck speed and route choice following the implementation of a toll on SR-520 in Seattle, WA.

The remainder of this chapter is organized as follows: section two provides a brief review of the state of practice in evaluating the effects of tolling on truck routing; section three introduces the study area; section four details the process by which the tolling effects are analyzed, section five applies the discussed methods and provides the results; section six offers conclusions from the study.

5.2 Literature Review

Truck travel demand elasticity modeling is a common approach to study the effects of tolling on truck trip diversion. Several freight demand elasticity models were developed to model the impacts of tolling on truck route choice. One example is the Ohio Turnpike truck traffic diversion analysis (Swan and Belzer 2007). The toll rates of the Ohio Turnpike were increased substantially in the 1990s for funding new construction projects and were later reduced once the projects were completed. Therefore it is an ideal dataset to estimate the truck demand elasticity with respect to toll rates. The truck VMT (Vehicle Miles Traveled), toll rates, and roadway speed limit data were collected between 1973 and 2005. Truck trip diversion from the Ohio Turnpike to other free alternatives was estimated using the ratio of Ohio Turnpike truck VMT to the U.S. total truck VMT. The results indicated that a 10% increase in toll rate per VMT increased the truck traffic diversion to a toll-free route by 4.7%.

Another study is the I-81 tolling in Virginia (VirginiaDOT 2004). In contrast to the Ohio Turnpike truck diversion research, which relied on historical observations, the I-81 project predicted the truck demand elasticity based on a customized non-linear Reebie’s Truck Cost Allocation Model (TCAM) developed by Oak Ridge National Laboratory. The model assumes that route choice is determined by travel cost only, and the trucking industry behaves in an economical manner, meaning that truckers always take the route along which the travel cost was the minimum. The TCAM estimated the travel cost using a non-linear model with the independent variables of travel distance, travel time, toll, expected congestion, equipment type, driver type and size of carrier. The route choice was translated into traffic diversion at each given toll rate to estimate the truck demand elasticity with respect to toll rates. The traffic diversion is expected to increase from 16% to 67% when the toll rate increases from 12 to 30 cents per mile. What’s more, the study revealed that the commodity type does not matter to the truck VMT diversion on I-81, with the exception of coal.

In addition to the truck VMT modeling, other researchers attempted to model the impacts of tolling and other influential factors on truck route choice based on stated-preference surveys. Arentze et al. (2006) conducted a stated preference survey to understand truck route choice behaviors. The survey was designed to examine how drivers trade off road accessibility characteristics against travel time and travel cost factors. Road accessibility refers to roadway geometry, e.g. sharp curves and intersections. There were 78 respondents who completed the on-line questionnaires. A mixed logit model was formulated based on the responses. It is noted that travel time has the strongest impact on route choice, and congestion and road category are significant factors as well. Sun et al. (2007) conducted interviews with truck drivers at three rest areas and truck stop locations along major highways in Texas, Indiana and Ontario to identify the factors that affect truck routing. Truck drivers were asked to choose between two hypothetical route alternatives. Two scenarios were created: the Bypass scenario and the Turnpike scenario. In the Bypass scenario, truckers were asked to choose between an urban freeway passing through the downtown area, and a bypass alternative which charges additional tolls and involves a longer travel distance, but shorter travel time. Under the Turnpike scenario, drivers were asked to choose between the tolled freeway and free local roads. A logit model was developed to analyze the factors that determine route choice. It is found that toll cost and travel time are the most significant factors affecting truck routing.

Studies of the value of truck travel time also involve the examination of roadway toll impacts on truck routing. Kawamura (2000) estimated truck value of time for different types of carriers based on a stated preference survey conducted in California. Truck drivers were asked to choose between an existing free road and a toll road for different combinations of travel time and costs. In addition, other truck operator characteristics including business type (for-hire or private fleet), shipment size, and pay scale (pay-by-hour, fixed salary or commission-based) were examined as well. The results reveal that the value of truck travel time is closely associated with the time and toll cost trade-off, as well as the business type and pay scale. More specifically, the for-hire carriers have higher value of travel time than private ones. Similar results were found by Zhou et al. (2009) that travel time and toll costs are the significant factors determining route choice. Meanwhile, smaller carriers are more likely to avoid a toll road when compared to larger companies.

The literature review summarizes the commonly applied methodologies to model the impacts of roadway tolls on truck trip diversion. Both the truck travel demand modeling based approach, and the survey and interview based approach are costly to collect data to support the analyses. However, GPS data is able to alleviate such concerns given the decreased GPS devices costs and increased market penetration of GPS technology used for truck fleet management. In addition, the traditional surveys and interviews consist of hypothetical routes and estimated attributes (e.g. travel time and speed), and therefore are not able to reflect the real world truck traffic performance. Since respondents are given these abstract and hypothetical situations, the route preferences indicated by the respondents may not be the same as their actual choices. In contrast, the data retrieved from GPS devices is able to provide details of truck specific movement (speed and travel time) and trajectory (route choice) information, and therefore can support realistic tolling impact analyses, including quantifying changes in truck speed, identifying influential factors in truck routing and estimating value of truck travel time based on utility functions. In light of this, this chapter investigates how truck GPS data can be used to implement the aforementioned tasks.

5.3 Study Area and Data Acquisition

Study Area

The 6.8-mile-long SR 520 is recognized as a critical corridor carrying traffic across Lake Washington between Interstate Highway 5 (I-5), the City of Seattle on the west, and I-405 and the Cities of Bellevue, Redmond, and Kirkland on the east (Figure 5-1). The alternate routes are I-90 and SR-522. The SR-520 bridge was built in 1963, and is approaching the end of its useful life. Thus Washington Department of Transportation (WSDOT) started tolling in December 2011 to fund the SR-520 replacement projects (WSDOT 2012). Toll rates are predetermined, and chang based on time-of-day, truck size, and payment method (WSDOT 2012). According to WSDOT, tolling caused some traffic which used to travel along SR-520 to divert to the alternative free road, I-90. Through March 2012, the traffic on SR 520 bridge dropped by 35% to 40%, while traffic on I-90 increased by 5% to 10% (Craig et al. 2012). No significant change along SR-522 was observed since SR-522 involves a longer detour, intersections and signal delays (Craig et al. 2012). Therefore, this study considers only I-90 as the alternative toll-free road for lake crossing commercial trips.

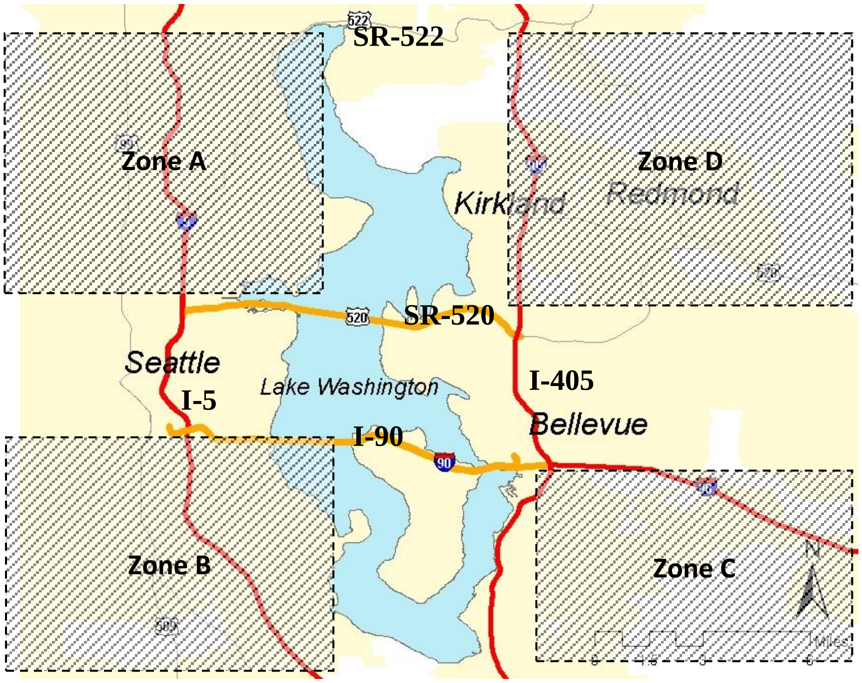


Figure 5-1 Study Area and Surrounding Highways

The study area is subdivided into four zones, denoted as A, B, C, and D, as shown in Figure 1. Areas outside of the four zones are not considered as the detour length is exceedingly long. It is less likely for truck drivers traveling between Zone A and Zone D to choose the toll-free route I-90 since it involes longer detour. That is, if SR-520 bridge is chosen, the crossing lake travel distance is the length of EB (eastbound) SR-520 shown in Figure 5-1. However, if I-90 is chosen, the crossing lake travel distance is the sum of SB (southbound) I-5 between SR-520 and I-90, EB I-90, and NB (northbound) I-405 between I-90 and SR-520. However, drivers traveling between Zone A and Zone C, and Zone B and Zone D face two comparable alternatives (similar travel distances): whether SR-520 or I-90 is chosen depends on the trade-offs between a set of time- and cost-related attributes. If SR-520 is chosen, drivers need to pay for the toll, but may experience shorter and more reliable travel time. In contrast, if free road I-90 is chosen, drivers do not need to pay for the toll, but may experience longer and less reliable travel time since increased traffic was observed on I-90 after tolling. Therefore this study examines the impacts of a set of time- and cost-related attributes on truck route choice based on empirical observations of truck trips between Zone A and Zone C, and Zone B and Zone D during both peak period (6:00 AM – 9:00 AM and 3:00 PM to 6:00 PM), and off-peak period (9:00 AM – 3:00 PM).

GPS Data Acquisition

This research used the GPS data collected in November 2011 (prior to the toll), and April 2012 (following the toll started on December 29th 2011). Data was provided on the condition of anonymity from trucks equipped with GPS devices traveling through the Puget Sound region. The GPS data was reported every 2-15 minutes and at every stop. Information provided by the GPS data includes a unique device ID, location (latitude and longitude), spot (instantaneous) speed, truck heading direction, time and date (McCormack et al. 2012). The truck fleet and commodity information is unknown. Before conducting any analysis, the raw GPS data was cleaned to remove problematic and duplicated data. More details of the GPS data collection and processing can be found in (Zhao et al. 2013, McCormack et al. 2012). In addition, an algorithm was developed to automatically identify discrete truck trips from raw GPS data based on truck dwell time (Ma et al. 2011).

5.4 Methodology

Impact of Tolling on Truck Travel Speed

The truck travel speeds on SR-520 and I-90 before and after tolling are compared to analyze the impacts of tolling on truck performance. The speed on SR-520 and I-90 before and after tolling were estimated using the *estimated link speed* method every 15-min interval (Zhao et al. 2011). The segment being studied is divided into several subsegments. For each subsegment, the speed is calculated by averaging the spot speed over the subsegment, and the corresponding travel time is computed by dividing the subsegment distance by the average spot speed. The total travel time (for the entire segment) is the sum of the travel times on each subsegment. The truck travel speed of the entire link, called the *estimated link speed*, is computed by dividing the total distance by the total travel time. The calculation is given in Equation 5-1. The outcome has been compared with space mean speed and it has been demonstrated that the *estimated link speed* approach is a reliable method to estimate truck speed (Zhao et al. 2011).

$$V = \frac{\sum_{i=1}^{n} l_i}{\sum_{i}^{n} \frac{l_i}{\bar{v}_i}} \qquad \text{Equation (5-1)}$$

where V = estimated link speed,

i = number of subsegments,

l_i = length of the ith subsegment,

$\bar{v}_i$ = average GPS spot speed on the ith subsegment.

Impact of Tolling on Truck Route Choice

A number of studies have investigated the influential factors in truck routing. A thorough review can be found in Cullinane and Toy (Cullinane and Toy 2000), in which they reviewed seventy-five articles and identified the five most common categories. These are travel cost/price/rate, travel speed, transit time reliability, characteristics of the goods and service. In addition, other studies have found that roadway geometric features, safety/security, accessibility, type of carriers, and drivers’ payment method are significant as well (Arentze et al. 2012, Sun et al. 2013, Kawamura 2000, Zhou et al. 2009). Not all of the aforementioned variables are readily available or can be retrieved from the truck GPS data, and therefore, this study chooses travel time, travel time reliability and toll rates as the influential attributes in truck routing. These factors are also identified as the most significant variables in most studies.

A logit model is used to quantify the impact of tolling on truck route choice based on the assumption that trucking industry behaves in an economical manner and always maximizes utility while choosing travel routes. Two alternative routes were considered: (1) toll bridge SR-520, and (2) toll-free route I-90. The utility functions are:

$$\begin{aligned} U_{90} &= \alpha + \beta_1 TT_{90} + \beta_2 TR_{90} \\ U_{520} &= \beta_1 TT_{520} + \beta_2 TR_{520} + \beta_3 Toll \end{aligned} \qquad \text{Equation (5-2)}$$

where U_{90} = utility function of choosing I-90,

U_{520} = utility function of choosing SR-520,

α = constant,

TT = Lake crossing time, as shown in Table 5-1,

TR = Lake crossing travel time reliability, defined as standard deviation of travel time,

Toll = toll rates.

Table 5-1 Truck Travel Time on Each Alternative

Truck trips	Route	Lake crossing time
Zone A->Zone C	SR-520	$TT_{SR\text{-}520\_EB}+TT_{I\text{-}405\_SB}$*
	I-90	$TT_{I\text{-}5\_SB}+TT_{I\text{-}90\_EB}$
Zone C-> Zone A	SR-520	$TT_{I\text{-}405\_NB}+TT_{SR\text{-}520\_WB}$
	I-90	$TT_{I\text{-}90\_WB}+T_{I\text{-}5\_NB}$
Zone B-> Zone D	SR-520	$TT_{I\text{-}5\_NB}+TT_{SR\text{-}520\_EB}$
	I-90	$TT_{I\text{-}90\_EB}+TT_{I\text{-}405\_NB}$
Zone D-> Zone B	SR-520	$TT_{SR\text{-}520\_WB}+TT_{I\text{-}5\_SB}$
	I-90	$TT_{I\text{-}405\_SB}+TT_{I\text{-}90\_WB}$

*$TT_{SR\text{-}520\_EB}$ represents travel time on eastbound SR-520, and $TT_{I\text{-}405\_SB}$ represents travel time on southbound I-405 segment between SR-520 and I-90

As illustrated in Equation 5-2, to construct the logit model, six variables are required for each lake crossing trip: actual route choice, actual travel time, potential travel time on the alternative route, actual travel time reliability, potential travel time reliability on the alternative route, and toll cost on SR-520. The following sections discuss how these variables are observed and estimated.

Truck Route Choice

The selection of lake crossing trips and identification of route choice consists of three steps. First, the processed GPS data is geocoded to the network, and GPS reads on SR-520 and I-90 are selected separately. Second, the truck trips containing GPS reads on SR-520 and I-90 are identified based on the unique GPS device ID. Third, the truck trips that are between Zone A and Zone C, and Zone B and Zone D are selected as the input to support the modeling process. The corresponding route choice of each truck trip is identified depends on if it contains GPS reads on SR-520 or on I-90. The complete process is accomplished inside the ArcGIS environment.

Truck Travel Time

The travel time considered in the model is the lake crossing travel time, as presented in Table 1. The actual lake crossing time is calculated by dividing travel distance by the truck GPS spot speed. To formulate the logit model, the potential travel time on the alternative route is needed as well. The travel time on the alternative route is estimated by dividing travel distance of the alternative route by the *estimated link speed* (calculated using Equation (5-1)) on that alternative route during the corresponding 15-min interval. Travel time is measured as minutes in the logit model.

Travel Time Reliability

Travel time reliability represents the level of consistency in travel times during a time period (Lomax et al. 2013). Numerous approaches have been developed to quantify travel time reliability, including the travel time standard deviation, 95th percentile of travel time, buffer time, probability of on-time arrival, etc. (Lomax et al. 2013). In this study, the travel time standard deviation is chosen as the reliability metric, which is consistent with the measure used in the highway pricing study SHRP 2 (Second Strategic Highway Research Program) Project C04 (Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand) (Parsons Brinckerhoff 2012). The travel time reliabilities on both the actual selected route and the alternative route are calculated for each 15-min interval.

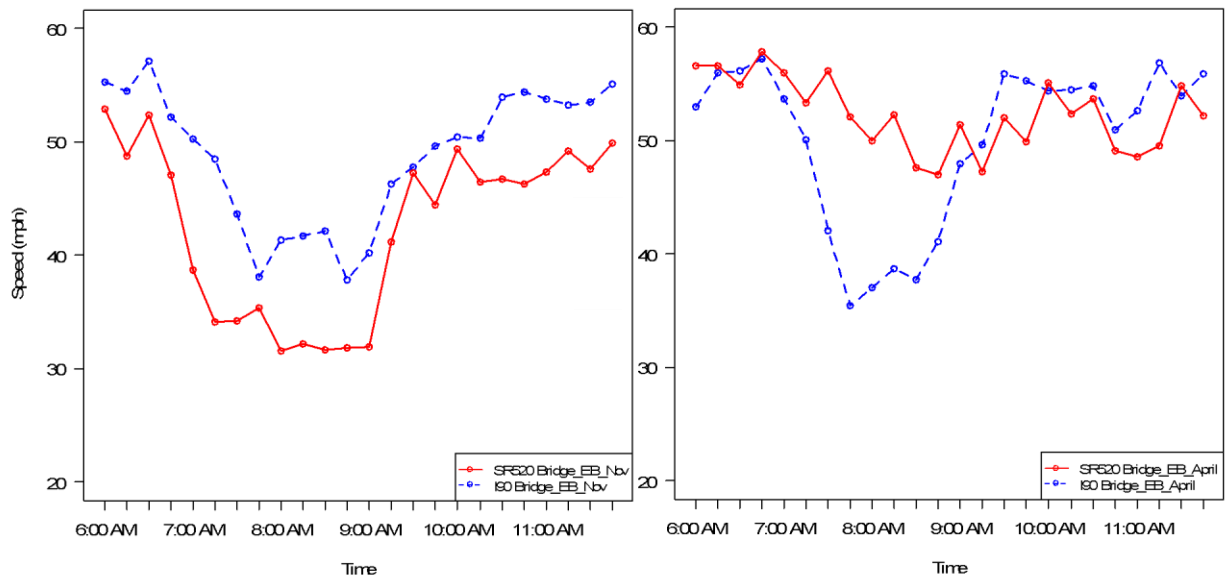
Toll Rate

Toll rates on SR-520 vary depending on the time-of-day, truck type and payment method. The rate ranges from $0 between 11 PM and 5 AM for all types of trucks to $10.50 between 7 AM to 9 AM for six-axle truck paying by the electronic tolling system. We assume all trucks on SR-520 use the electronic tolling system. Higher toll rates are expected if other payment methods, e.g. pay by mail, are used (WSDOT 2012). Since the truck size is unknown, the toll rate on SR-520 during each time period is calculated as a weighted average using WSDOT truck counts by number of axles.

5.5 Results

Impacts of Tolling on Truck Speed

Figure 5-2 presents the weekday average truck speed on EB SR-520 and EB I-90 before and after tolling, between 6:00 AM and 12:00 PM. The speed was aggregated every 15-min interval. According to Figure 5-2 (a), prior to tolling, truck speed on SR-520 was always lower than on I-90. The lowest speed was about 30 mph between 8:00 AM and 9:00 AM. However, as illustrated in Figure 5-2 (b), travel speed on SR-520 improved significantly following the toll. The truck speed was around 50 mph during the AM peak period, which exceeded the speed on I-90. In addition to the improvement to travel speed, the travel time on SR-520 was more stable with reduced fluctuation—the difference in travel speed was greater than 20 mph in November 2011 and was less than 10 mph in April 2012. Changes in travel speed on I-90 were much less pronounced than on SR-520. The truck travel speed on I-90 decreased slightly between 7:30 AM and 9:00 AM. The changes in truck speed on both bridges were mainly resulted from the traffic diversion from SR-520 to I-90. According to the data collected by the traffic count devices deployed along SR-520 bridge in April 2012, the EB number of passenger car trips reduced by 32% compared to the volume recorded in November 2011, and the commercial truck trips dropped by 26%.

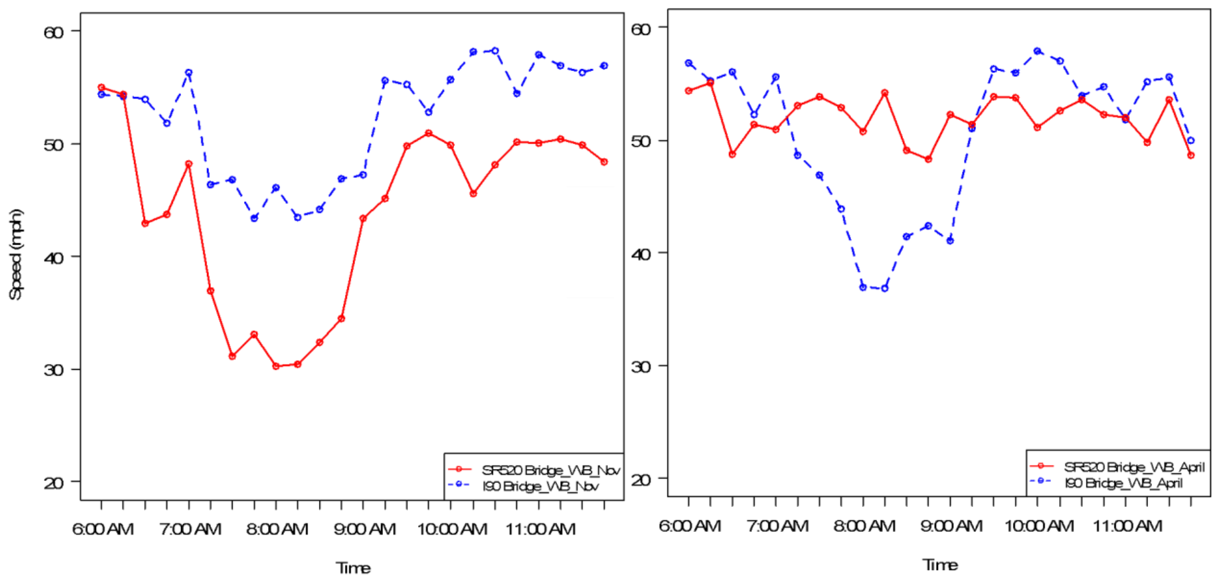


(a) (b)

Figure 5-2 (a) Truck speed on EB SR-520 and EB I-90 in November 2011 and (b) Truck speed on EB SR-520 and EB I-90 in April 2012

The truck speeds over WB SR-520 and I-90 before and after tolling were also compared, and are presented in Figure 3. Similar impacts on WB truck speed were observed. As displayed in Figure 3 (a), before tolling, average truck speed on SR-520 was lower than the speed on I-90 (except between 6 AM and 6:15 AM), with lowest speed of 30 mph around 8 AM. The speed increased dramatically after tolling and was greater than the speeds on I-90 between 7:15 AM and 9 AM, as shown in Figure 3 (b). Meanwhile, the variation of speed on SR-520 was reduced after tolling. The truck speed on WB I-90 after tolling did not change considerably between 6:00 AM and 7:30 AM. However, the speed declined significantly between 7:30 AM and 9:15 AM. This may be due to the fact that toll rate increases by 25% during 7 AM to 9 AM compared to the rate during 6 AM to 7 AM, and results in considerable traffic diversion from SR-520 to I-90. The monthly traffic volume collected by the traffic count devices shows that passenger car trips on

WB SR-520 bridge decreased by 31% in April 2012 compared to the data collected in November 2011, and the truck traffic declined by 26%.



(a) (b)

Figure 5-3 (a) Truck speed on WB SR-520 and WB I-90 in November 2011 and (b) Truck Speed on WB SR-520 and WB I-90 in April 2012

Impacts of Tolling on Truck Route Choice

There are 185 truck trips observed during the peak period, among which 25 trips selected SR-520 and 160 trips selected I-90. For the off-peak period in April 2012, there are 203 trucks trips observed, among which 37 trips chose SR-520 and 166 trips chose I-90. The logit model was implemented using R software. The results are shown in Table 2. Two traffic phases were modeled: the peak period (6: 00 AM – 9:00 AM and 3:00 PM – 6:00 PM), and off peak period (9:00 AM- 3:00 PM). Travel time and toll rate are significant variables for both phases. The signs of travel time and toll rate are both negative, which means that the increase of travel time and toll rate of a specific route will reduce the utility of that route as well as the probability that it will be chosen. The constant in the I-90 logit function is positive, and it indicates that everything being equal, the utility of choosing I-90 is greater than choosing SR-520, which reveals that I-90 is more preferable to truck drivers. I-90 is an interstate highway with four-lanes each direction and sufficient shoulder width, while SR-520 has two-lanes each direction and insufficient shoulder. These geometric features may make I-90 more attractive to trucks than SR-520. In addition, there are weight and size restrictions on SR-520, and some heavy or large trucks may be required to choose I-90. The constants of the two logit functions also reveal that if there is zero toll, the SR-520 and I-405 combination will be more attractive only if it is on average 23 minutes faster during the peak and 19 minutes faster during the off peak. It should be noted that travel time reliability is not significant at the 95% confidence level, and therefore is eliminated from both models. The utility functions for the two time periods are written as follows:

Peak period

$$U_{90} = 4.717 - 0.207TT_{I-90}$$
$$U_{520} = \qquad -0.207TT_{SR-520} - 0.68Toll$$

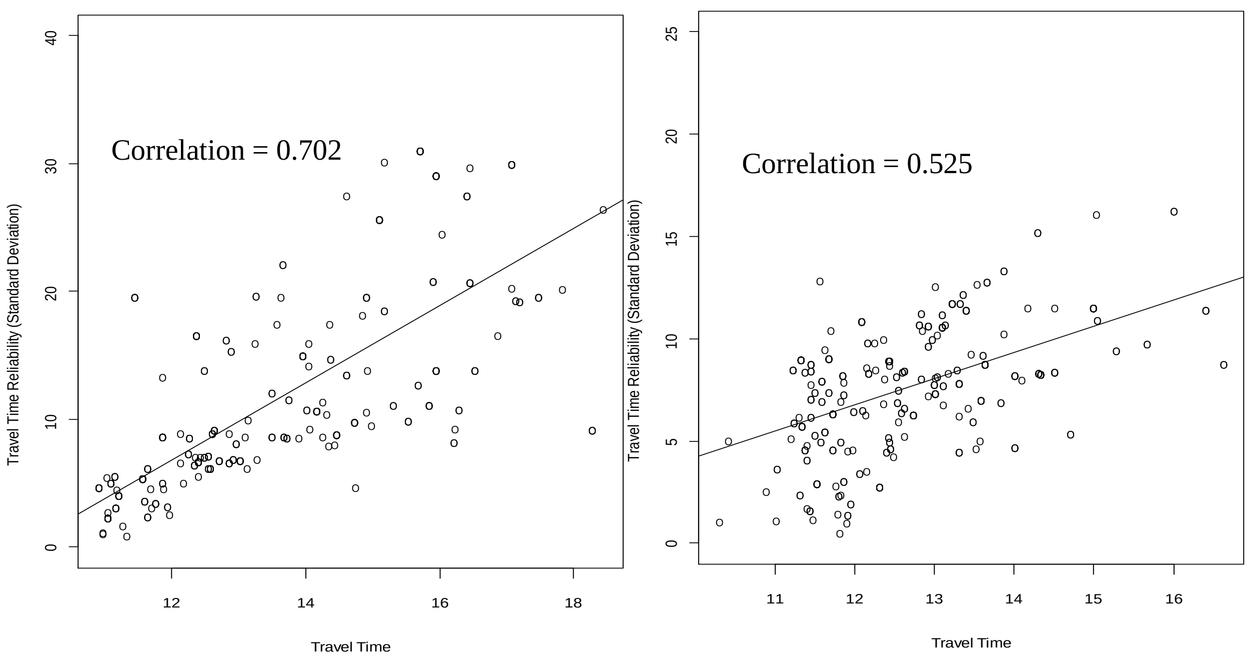
Off-peak period

$$U_{90} = 5.586 - 0.301TT_{I-90}$$
$$U_{520} = \qquad -0.301TT_{SR-520} - 0.703Toll$$

Table 5-2 Logit Model Results

Time period	Variables	Coefficient	z-Value	p-Value
Peak period (6 AM -9 AM and 3 PM to 6 PM)	Constant	4.717	4.278	<0.005
	Travel time	-0.207	-2.190	0.028
	Travel time reliability	0.001	0.022	0.983
	Toll rate	-0.680	-10.041	<0.005
Off peak period (9 AM to 3 PM)	Constant	5.586	3.563	<0.005
	Travel time	-0.301	-2.134	0.033
	Travel time reliability	-0.044	-0.826	0.408
	Toll rate	-0.703	-9.997	<0.005

Travel time reliability is not observed to be a significant variable in route choice during both time periods. This is due to the correlation between travel time and travel time reliability. Travel time reliability is low when travel times are high. Travel times are at their lowest when drivers can travel at free flow speed. If this is the case, travel times are very reliable as there is no congestion. Figure 4 presents the scatter plot of travel time versus travel time reliability during peak period and off peak period. Linear relationships between the two variables during both periods are observed, and the correlations are 0.702 and 0.525 respectively. To eliminate the influence of correlation, travel time was removed from both utility functions, and only travel time reliability and toll rate were considered. The updated model results are shown in Table 5-3. Both travel time reliability and toll rate are statistically significant, and the signs are both negative. The results indicate that travel time reliability is an influential factor determining truck route choice as well, and trucking industry is willing to pay for a toll for a reliable route.



(a) (b)

Figure 5-4 (a) Scatter plot of travel time versus travel time reliability during peak period (6 AM to 9 AM and 3 PM to 6 PM), and (b) scatter plot of travel time versus travel time reliability during off peak period (9 AM to 3 PM)

Table 5-3 Logit Model Results (without Travel Time)

Time period	Variables	Coefficient	z-Value	p-Value
Peak period	Constant	2.529	5.902	<0.005
(6 AM to 9 AM	Travel time	-0.046	-2.183	0.029
and 3 PM to 6	reliability	-0.707	-10.642	<0.005
PM)	Toll rate			
Off peak period	Constant	2.385	5.803	<0.005
(9 AM to 3 PM)	Travel time	-0.112	-2.528	0.012
	reliability	-0.760	-11.415	<0.005
	Toll rate			

Value of Truck Travel Time

The coefficients of travel time and toll rate reflect the sensitivity of commercial trips to changes in travel time and cost. Their ratio is able to capture the trade-off between travel time and toll rate, as shown in Equation 5-3. According to Equation 3 and the model results presented in Table 2, the value of truck travel time during peak period and off-peak period can be calculated, and are 18.26 $/hr and 25.69 $/hr respectively.

$$\beta_1 = \frac{\partial U}{\partial TT}$$
$$\beta_3 = \frac{\partial U}{\partial Toll} \qquad \text{Equation (5-3)}$$
$$Value\ of\ \ Truck\ \ Travel\ \ Time = \frac{\beta_1}{\beta_3}$$

It should be noted that if this study has bias, it will underestimate the value of truck travel time. First, all trucks on SR-520 were assumed to use the most economical payment method. It is possible that other payment methods are used, which generate higher toll costs, and consequently yield higher value of truck travel time. Second, there are vehicle weight and size restrictions on the SR-520 bridge, and therefore we may under-sample heavy trucks as compared to other facilities. Third, trucks may have less flexibility in rescheduling and changing routes.

It is also noted that the value of truck travel time during off-peak period is greater than the value during peak period. This is explainable within the commercial trips context. The peak period value of travel time for passenger vehicles is expected to be greater than the off-peak period value since most trips occurring during peak period, e.g. working trips, have fixed destinations and less flexibility in changing departure time or route. However, many commercial trips with strict delivery window and customer requirements do not necessarily occur during the peak period defined for passenger trips. To verify this assumption, we used the same data but redefined the truck specific peak period as 9 AM to 5 PM, and off peak period as 6 AM to 9 AM and 5 PM to 6 PM. The updated model results are shown in Table 4. According to the model results, the value of truck travel time during truck specific peak period and truck specific off peak period are 25.15 $/hr and 19.44 $/hr respectively.

Table 5-4 Logit Model Results (Truck Specific Peak and Off Peak Periods)

Time period	Variables	Coefficient	z-Value	p-Value
Truck Peak period (9 AM to 5 PM)	Constant	5.169	3.972	<0.005
	Travel time	-0.285	-2.433	0.015
	Travel time reliability	-0.002	-0.052	0.959
	Toll rate	-0.680	-11.338	<0.005
Truck off peak period (6 AM to 9 AM and 5 PM to 6 PM)	Constant	4.794	3.783	<0.005
	Travel time	-0.219	-1.988	0.047
	Travel time reliability	0.003	0.102	0.919
	Toll rate	-0.676	-8.064	<0.005

5.6 Conclusions

This research investigates how truck GPS data can be used to quantify the impacts of tolling on truck speed and routing, using SR-520 as a case study. It is found that roadway tolls affect truck speed on both the toll route and the alternative toll-free route. Tolling may alleviate the congestion on toll road during both peak and off-peak periods, and increase the congestion on the alternate free route during peak period.

A logit model was developed to understand the effects of toll rate, travel time and travel time reliability on truck routing. Two traffic phases were examined: the peak period (6 AM to 9 AM and 3 PM to 6 PM) and off-peak period (9 AM to 3 PM). It is found that travel time and toll rate are both significant factors during both phases. The travel time reliability is not significant in the combined model results due to the correlation between travel time and travel time reliability. However, it is significant when travel time is eliminated from the model, which demonstrates that travel time reliability is an influential factor determining truck route choice as well.

The value of truck travel time varies with the definition of peak and off-peak periods. The values during general traffic peak period (6 AM – 9 AM and 3 PM to 6 PM) and off-peak (9 AM to 3 PM) are 18.26$/hr and 25.69 $/hr respectively. If we look at the truck specific peak period (9 AM to 5 PM) and truck specific off peak period (6 AM to 9 AM and 5 PM to 6 PM), the value of truck travel time are 25.15 $/hr and 19.44 $/hr. The values are comparable with the estimates discussed in several literatures, in which the value of truck travel time ranges from 20 $/hr to 50 $/hr (in 2012 dollars) (Kawamura 2000). The results can be used to inform tolling rates, and to better forecast the impact of tolling on truck route choice. The fleet information is unknown as the data was provided on condition of anonymity. Any bias presented by differences between the fleet represented in the GPS dataset, and the truck population at large, are not known.

Chapter 6 CONCLUSIONS

Despite the fact that a considerable number of studies have examined passenger vehicle mobility performance and how passenger vehicles response to roadway tolls, limited truck-specific research has been done. To a large extent, this is due to the deficiency of truck-specific movement data. The increasing availability of truck GPS data provides reliable truck movement data, and therefore it initiates opportunities for truck-specific research. This dissertation begins to fill the gap by employing truck GPS data to freight planning applications. Three key truck-specific research questions are addressed, including measuring truck travel time reliability, forecasting truck travel time and travel time reliability, and analyzing tolling project impacts on truck performance and routing. This final chapter presents conclusions and findings of each research question. The studies are expected to be applied by planners and engineers to support freight planning. Thus this chapter also discusses the potential obstacles and challenges to implement the proposed methodologies.

6.1 Measuring Truck Travel Time Reliability

Travel time reliability is a critical factor for evaluating freight-highway system performance and determining truck route choice. Numerous quantitative approaches have been proposed to measure travel time reliability. The author classified existing reliability measurements into two categories according to the data on which these approaches are based: travel time based reliability measure and GPS spot speed based measure. The GPS spot speed based approach is superior than travel time based approach due to the fact that the GPS data employed in this research is low sample sized and provides spot speed data instead of travel time data. We improved the GPS spot speed based approach by calculating the speed distribution coefficient of variation. This improved approach provides numerical values, and therefore allows more quantitative reliability evaluation.

We ranked reliability of the same segment during different times-of-day and days-of-week using a number of reliability metrics, and found that the ranking results varied depending on the measures used. To further explore the relationship among these reliability measures, the correlations among each measure were computed. The results reveal that there exist large deviations among reliability measures, especially during peak period compared to off-peak period. Both ranking and correlation analyses suggest that different measures provide different conclusions for the same underlying data and traffic condition. We summarized the advantages and disadvantages of each measure, and provided recommendations of the appropriate measure to use under different situations.

When raw GPS data is readily available, the improved spot speed distribution approach is recommended as it does not require additional efforts to retrieve travel time information from raw data and allows small sample size. However, this approach has not been widely applied. Also, it is more complex to compute and explain to non-technical audience compared to other reliability measures, e.g. travel time standard deviation, percentile method and buffer time method. Nevertheless, it is an asset to avoid calculating travel time from raw GPS data as it requires resources and provides opportunities for introducing error.

6.2 Forecasting Truck Travel Time and Travel Time Reliability

This research initiates from one of our previous studies to develop a framework to determine highway-truck benefits and economic impacts associated with highway investments. In that project, we realized that there is no truck-specific benefits evaluation application, which is primarily due to the lack of truck data and models to predict truck performance associated with transportation investments. In light of this, this dissertation proposes approaches to develop such models to forecast truck travel time and travel time reliability using truck GPS data.

Forecast Freeway Truck Travel Time

The objective of chapter 3 is to propose a truck travel time forecasting approach that can support long-term freight project prioritization and planning, not real-time operations. The idea of this approach is based on multi-regime relationships between truck speed and segment density. Cluster analysis was employed to segment traffic regimes based on the characteristics of truck speed and segment traffic density using the K-means algorithm. We fitted the data of each cluster using linear, logarithmic and exponential models to find the best fit of the observed data. The fitted models were compared with two traditional travel time forecasting models, the BPR function and the Akçelik function. The proposed model generates less deviation between travel time estimates and observations, and therefore performs better than the existing two approaches.

The k-means cluster algorithm requires user defined number of clusters, which is determined by both the distribution of observations in a dataset and the desired resolution of the user. Meanwhile, the number should not be too many for convenient use of the model. Thus for the two case studies presented in this dissertation, the authors conducted the cluster analysis with two clusters and three clusters respectively, and compare the results. It is found that for case study II, three-cluster did not work as the clusters cannot be fitted properly. For case study I, no substantial improvement regarding model forecasting accuracy was observed from 2 clusters to 3 clusters.

This proposed approach is expected to forecast more accurate truck travel time in response to segment density changes. Given this, and the simplicity of the approach, we recommend users to apply the clustering and best-fit modeling approach, and develop their own equations for different locations.

Forecast Freeway Truck Travel Time Reliability

Chapter 4 of this dissertation explores how truck travel time reliability can be forecasted using truck GPS data. This chapter is based on the previous research efforts to measure truck travel time reliability using COV of the spot speed distribution illustrated in Chapter 2. To forecast reliability, we first plotted the hourly truck spot speed distribution from 5 AM – 10 AM based on one month GPS observations. The figures suggest that speed distribution is affected considerably by traffic density. More specifically, truck speed distribution follows a unimodal distribution when traffic density is low and most of the trucks travel at a desired speed. The presence of the second distribution emerges with the increase of traffic density. During congestion phase, the distribution presents a bimodal distribution, which is consisted of two phases: the low speed phase and high speed phase. The probability of truck speed falling within the low speed phase increases with the growth of traffic density. The low speed distribution wanes when traffic density drops.

Such changes of speed distribution can be reflected by the corresponding COV. Thus we plotted the COV versus traffic density, and found that there exist strong correlations between the two datasets. For the case study discussed in this dissertation, we found travel time was always reliable when density was less than 15 veh/mi/ln. When density was between 15 veh/mi/ln and 28 veh/mi/ln, COV increased dramatically with the growth of density. COV did not change considerably when density was greater than 28 veh/mi/ln. The K-means algorithm was then employed to segment the dataset into three groups based on the characteristics of the data. For the second cluster, in which density was between 15 veh/mi/ln and 28 veh/mi/ln, we fitted the data using linear regression. The fitted equation is able to predict travel time reliability in response to different traffic density when future density is available or predictable.

The idea of using COV of truck speed distribution to represent travel time reliable is relatively new and complex to be explained to first-time users. However, this method overcomes the sample size constraints and does not require additional data processing efforts to retrieve travel time information form raw GPS data. These are the strengths of this approach. The proposed COV forecasting approach is easy to implement and can be extended for studies of different locations and/or with different GPS datasets.

6.3 Impacts of Tolling on Truck Performance and Routing

In the past decades, there have been multiple studies seeking to collect truck movement data to support freight planning, but none of them have been investigated how truck GPS data can be used to support the impact analysis of tolling on truck mobility performance and vehicle routing. Thus Chapter 5 conducts an analysis to quantify tolling impacts on truck speed and routing. Both truck spot speed and vehicle trajectory (route choice) information was retrieved from GPS data.

The 6.8-mile-long SR-520 bridge in city of Seattle, WA was selected as the case study. The alternative toll-free route is I-90 bridges. The speed comparison shows that the truck speed along SR-520 improved considerably following the toll. The lowest truck speed during morning peak-period was 30 mph after tolling and increased to 50 mph after tolling. In addition to the speed improvement, the reliability along SR-520 was improved as well. Changes in travel speed on I-

90 were much less pronounced than on SR-520. The changes in truck speed along both bridges were mainly resulted from the traffic diversion from SR-520 to I-90.

A logit model was developed to identify the factors affecting truck route choice based on the assumption that trucking industry behaves in an economical manner and always maximizes utility while choosing travel routes. Two traffic phases were modeled: the peak period (6 AM -9 AM and 3 PM -6 PM) and off-peak period (9 AM – 3 PM). The model results reveal that travel time, travel time reliability and toll rate are significant variables for both phases. The signs of travel time and toll rate are both negative, which means that the increase of travel time and toll rate of a specific route will reduce the utility of that route as well as the probability that it will be chosen. We further estimated the value of truck travel time and found that the value of truck travel time during truck specific peak period and truck specific off peak period are 18.26 $/hr and 25.69 $/hr respectively.

This study proves the capability of truck GPS data in supporting tolling impact analysis. The results can be used to inform tolling rates, and to better forecast the impact of tolling on truck route choice. The fleet information is unknown as the data was provided on condition of anonymity. Any bias presented by differences between the fleet represented in the GPS dataset, and the truck population at large, are not known.

By addressing the above three research questions, the capability of truck GPS data in supporting truck mobility measurement, forecasting and tolling impact analysis have been fully investigated. The dissertation proposes a systematic set of approaches to take the advantage of the emerging GPS data source to support freight planning. The reliability measurement approach is innovative compared to traditional travel time based reliability measures. Both truck travel time forecasting and travel time reliability forecast analyses demonstrate the impacts of traffic conditions on truck mobility performance. More specifically, the research quantifies to what extend traffic may influence truck travel time and reliability. The proposed approaches are adaptive for different locations and/or different GPS dataset as long as the GPS data formats are consistent.

REFERENCES

The American Association of State Highway and Transportation Officials (AASHTO). (2012). *SCOPM task force findings on national-level performance measures*. Retrieved from http://www.dot.ca.gov/hq/transprog/map21/Map-21%20Useful%20Links/SCOPM%20Task%20Force%20Findings%20on%20National%20Level%20Measures%20FINAL%20(11-9-2012).pdf

Akçelik, R. (1991). Travel Time Functions for Transport Planning Purposes: Davidson's Function, Its Time-Dependent Form and an Alternative Travel Time Function. *Australian Road Research* 21 (3), pp. 49–59.

Atlanta Regional Commission. (2011). *The Travel Forecasting Model Set For the Atlanta Region 2010 Documentation*.

American Trucking Association (ATA). (2008).*Trucking and the Economy 2007-2008 Edition*, Retrieved from http://www.trucking.org/Newsroom/Trucks%20Are/Trucking%20and%20the%20Economy.pdf

American Transportation Research Institute (ATRI) and Federal Highway Administration (FHWA). (2005). *Measuring travel time in freight-significant corridors*. *Rep. FHWA-HOP-05-036*, Washington, D.C.

American Trucking Research Institute (ATRI). (2010). *2009 Bottleneck Analysis of 100 Freight Significant Highway Locations*. Retrieved from http://atri-online.org/2010/05/08/944/

ATRI. (2011). *Freight Performance Measure Congestion Monitoring at 250 Freight Significant Highway Locations*. Retrieved from http://atri-online.org/2011/10/01/fpm-congestion-monitoring-at-250-freight-significant-highway-locations/

ATRI. (2013). *Freight Performance Measure Analysis of Freight-Significant Highway Locations 2013*. Retrieved from http://atri-online.org/2013/07/08/atri-100-freight-locations/

Battelle. (1999). *Heavy-Duty Truck Activity Data Collection and Analysis Using Global Positioning Systems*. Prepared for USDOT, FHWA, Office of Highway Information Management. Retrieved from http://www.fhwa.dot.gov/ohim/tvtw/final.pdf

Battelle. (2011) *FAF3 Freight Traffic Analysis*. Retrieved from http://faf.ornl.gov/fafweb/Data/Freight_Traffic_Analysis/faf_fta.pdf

Beagan, D., Fischer, M., and Kappam, A. (2007). *Quick Response Freight Manual II*, prepared for USDOT, FHWA, Office of Freight Management and Operations, Washington, D.C.

Bogers, E. A. I. and van Zuylen, H. J. (2004). The importance of reliability in route choice in freight transport for various actors on various levels. *Proc., European Transport Conference,* Strasbourg, France.

Cambridge Systematics, Inc. and Jack Faucett Associates. (2001). National Cooperative Highway Research Program 298, Truck Trip Generation Data, Washington, D.C.

Cambridge Systematics, Inc. (2007). *PSRC Travel Model Documentation (for version 1.0), Updated for Congestion Relief Analysis, Final Report*. Retrieved from http://www.psrc.org/assets/1511/model_doc_final_.pdf

Cambridge Systematics, Inc. (2013). *Analytical procedures for determining the impacts of reliability mitigation strategies*. Prepared for Strategic Highway Research Program 2 (SHRP 2) Reliability Project L03. Rep. SHRP 2 Report S2-L03-RR-1, Washington, D.C.

Chien, S. I. J., & Kuchipudi, C. M. (2003). Dynamic travel time prediction with real-time and historic data. *Journal of transportation engineering*, *129*(6), 608-616.

Drake, J. S., J. L. Schofer, and A. D. May. (1967). A Statistical Analysis of Speed Density Hypotheses. *Third International Symposium on the Theory of Traffic Flow*. Elsevier North-Holland, Inc., New York

D'Angelo, M. P., Al-Deek, H. M., & Wang, M. C. (1999). Travel-time prediction for freeway corridors. *Transportation Research Record: Journal of the Transportation Research Board*, *1676*(1), 184-191.

Dowling, R. G., Singh, R., & Cheng, W. W. K. (1998). Accuracy and Performance of Improved Speed-Flow Curves. *Transportation Research Record: Journal of the Transportation Research Board*, 1646(1), 9-17

Dowling Associates, Inc. (2000). *Washington State Department of Transportation Mobility Project Prioritization Process Benefit-Cost Software Users' Guide*. Retrieved from http://www.wsdot.wa.gov/mapsdata/travel/pdf/Mobility_Users_Guide_2001.PDF

Edie, L. C. (1961). Car Following and Steady-State Theory for Non-Congested Traffic. *Operations Research*, Vol. 9, pp. 66–76.

Federal Highway Administration (FHWA). (2012a). *Freight Facts and Figures 2012*, FHWA-HOP-13-001, Retrieved from http://ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/docs/12factsfigures/index.htm

Federal Highway Administration (FHWA). (2012b). Moving Ahead for Progress in the 21st Century Act (MAP 21) (2012 b). Retrieved form:http://www.fhwa.dot.gov/map21/docs/map21_summary_hgwy_provisions.pdf

Federal Highway Administration (FHWA). (2013). *National Performance Management Research Data Set*. Retrieved from http://www.ops.fhwa.dot.gov/freight/freight_analysis/perform_meas/vpds/npmrdsfaqs.htm

Figliozzi, M., Wheeler, N., Albright, E., Walker, L., Sarkar, S., and Rice, D. (2011). Algorithms for Studying the Impact of Travel Time Reliability Along Multi-segment Trucking Freight Corridors, *Transportation Research Record: Journal of the Transportation Research Board,* 2224, 26 - 34.

Florida Department of Transportation (FDOT). (2012). *Truck GPS Data for Freight Performance Measurement and Planning: Applications, Issues, and Opportunities*. Retrieved from: http://www.fsutmsonline.net/images/uploads/mtffiles/Truck_GPS_Data__Abdul_Pinjari.pdf

Greaves, S., and Figliozzi, M. (2008). Collecting Commercial Vehicle Tour Data with Passive Global Positioning System Technology: Issues and Potential Applications. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2049,158-166.

Hartigan, J.A. (1975). Clustering algorithms. John Wiley & Sons, Inc.

Highway Capacity Manual (2000)

Jones, C., Murray, D., and Short, J. (2005). Methods of Travel Time Measurement in Freight-Significant Corridors, Presented at the *Transportation Research Board Annual Meeting*, Washington, D.C.

Kerner, B. S., & Rehborn, H. (1996). Experimental properties of complexity in traffic flow. *Physical Review Series*, 53, R4275-R4278.

Liao, C. (2009). *Using Achieved Truck GPS Data for Freight Performance Analysis on I-94/I-90 from the Twin Cities to Chicago.* Retrieved from http://www.dot.state.mn.us/ofrw/PDF/I94GPSFinalReport.pdf

Liao, C. (2014). *Using Truck GPS Data for Freight Performance Analysis in the Twin Cities Metro Area*. Report no. MnDOT 2014-14. Retrieved from http://www.dot.state.mn.us/ofrw/PDF/GPS/task3.pdf

Lomax, T., Schrank, D., Turner, S., and Margiotta, R. (2003). Selecting Travel Reliability Measures. Retrieved from http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/TTI-2003-3.pdf

Ma, X., McCormack, E., and Wang, Y. (2011). Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measure Program. *Transportation Research Record: Journal of Transportation Research Board*, No. 2246, 92-100.

McCormack, E., and Hallenbeck, M. (2005). *Options for Benchmarking Performance Improvements Achieved from Construction of Freight Mobility Projects*, final research report, contract T2695, task 52.

McCormack, E., and Hallenbeck, M. (2006). ITS Devices Used to Collect Truck Data for Performance Benchmarks. *Transportation Research Record: Journal of Transportation Research Board*, No. 1957, 43-50.

McCormack, E., Ma, X., Klocow,C., Currarei, A. and Wright, D. (2010). *Developing a GPS-Based Truck Freight Performance Measure Platform*. WA-RD 748.1, Retrieved from http://www.wsdot.wa.gov/research/reports/fullreports/748.1.pdf

McCormack, E., Scharnhorst, E., and Zhao, W. (2011). *Using GPS truck data to identify and rank bottlenecks in Washington State*. Report prepared for Washington State Department of Transportation, Olympia, WA.

McFarland, W. F., J. Memmott, and M. Chui. (1993). *Microcomputer Evaluation of Highway User Benefits* - Final Report for NCHRP 7-12. Texas Transportation Institute, Texas A&M University. College Station, TX.

Golias, M. (2013). *Evaluating the Hours-of-Service Rule via GPS/GIS Truck Trip Data*. Retrieved from https://wwwnew.memphis.edu/ifti/pdfs/cifts_evaluating_hours_of_service.pdf

National Cooperative Highway Research Program (NCHRP). (2008). NCHRP Report 618*: Cost-Effective Performance Measure for Travel Time Delay, Variation, and Reliability*. Transportation Research Board, Washington, D.C.

National Cooperative Freight Research Program (NCFRP). (2011). *NCFRP Report 10-Performance Measures for Freight Transportation*. Transportation Research Board, Washington, D.C.

Pu, W. (2010). Analytic Rrelationships between Travel Time Reliability Measures. Paper presented at *the 90th Transportation Research Board Annual Meeting*, Washington, D.C.

Puget Sound Regional Council. (2009). *Travel Model Improvements for the Congestion Management Process and Long Range Transportation Plan Update*, Retrieved from http://www.psrc.org/assets/3129/Final_Report_PSRC_Travel_Model_Freight.pdf

R Software Package “Cluster”. Retrieved from http://cran.r-project.org/web/packages/cluster/cluster.pdf

Sharman, B. W., & Roorda, M. J. (2011). Analysis of Freight Global Positioning System Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2246(1), 83-91.

Schofield, M. and Harrison, R. (2007). *Developing Appropriate Freight Performance Measures for Emerging Users*. Rep. 473700-00073-1, Austin, Texas.

Sun, L. and Zhou, J. (2005). Development of Multiregime Speed-Density Relationship by Cluster Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1934, 64-71.

Texas Transportation Institute (TTI) and Cambridge Systematics, Inc. (2006). *Travel Time Reliability: Making it There on Time, All the Time*. Rep. prepared for Federal Highway Administration.

The National Commission。 (2008), The National Surface Transportation Policy and Revenue Study Commission Transportation for Tomorrow Executive Summary, Retrieved from http://www.mtc.ca.gov/news/NSTPRSC/nstprsc_exec_summ.pdf

U.S. Bureau of Public Roads (1964) Traffic Assignment Manual, Washington, DC

U.S. Department of Transportation (USDOT). (2006a). *Travel Time Reliability: Making It There On Time, All The Time*. Retrieved from http://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm

USDOT. (2006b). *Freight Performance Measurement: Travel Time in Freight-Significant Corridors*. Report FHWA-HOP-07-071, Washington, D.C.

USDOT. (2010). *Measuring Cross-Border Travel Times for Freight: Otay Mesa International Border Crossing Final Report*. Report FHWA-HOP-10-051, Washington, D.C.

USDOT Office of the secretary of transportation. (2012), *Interim Guidance on State Freight Plans and State Freight Advisory Committee*, Federal Register, 77 (199). Retrieved from http://www.fhwa.dot.gov/map21/docs/guidance_freight_101512.pdf

van Lint, J.W.C. and van Zuylen, H. J. (2005a). Monitoring and predicting freeway travel time reliability using width and skew of day-to-day travel time distribution. *Transportation Research Record: Journal of the Transportation Research Board 1917*, Transportation Research Board, Washington, D.C., 54-62.

van Lint, J. W. C., Hoogendoorn, S. P., & van Zuylen, H. J. (2005b). Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*, *13*(5), 347-369.

van Lint, J.W.C., van Zuylen, H.J., AND Tu, H. (2008). Travel time unreliability on freeways: why measures based on variance tell only half the story. *Transportation Research Part A 42*, 258-277.

Wang, Z., Goodchild, A.V. (2014). GPS Data Analysis of Impact of Tolling on Truck Speed and Routing: Case Study in Seattle, Washington. *Transportation Research Record: Journal of the Transportation Research Board*.

Washington State Department of Transportation (WSDOT). (2010). The 2010 congestion report. *Rep. Prepared for the Washington State Dept. of Transportation*, Olympia, WA.

Washington State Department of Transportation (WSDOT). (2011a). *Using GPS Truck Data to Identify and Rank Bottlenecks in Washington State*

Yeon, J., Elefteriadou, L., & Lawphongpanich, S. (2008). Travel Time Estimation on a Freeway using Discrete Time Markov Chains. *Transportation Research Part B: Methodological*, *42*(4), 325-338.

You, S. I. (2012). *Methodology for tour-based truck demand modeling*. University of California, Irvine.

Zhao, W., Goodchild, A., and McCormack, E. (2011). Evaluating the Accuracy of Spot Speed Data from Global Positioning Systems for Estimating Truck Travel Speed, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2246, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 101–110.

Zhao, W., McCormack, E., Dailey, D. J., & Scharnhorst, E. (2012). Using truck probe GPS data to identify and rank roadway bottlenecks. *Journal of Transportation Engineering*, *139*(1), 1-7.