Using Sparse GPS Data to Estimate Link Travel Time for Truck Transport

National Center for Freight & Infrastructure Research & Education
Department of Civil and Environmental Engineering
College of Engineering
University of Wisconsin–Madison

Authors:
Wen Wang, Chun-Wei Lin, Kai Yin, Qi Gong, Teresa M. Adams, and Xiubin Bruce Wang

Principal Investigator:
Teresa M. Adams
National Center for Freight & Infrastructure Research & Education
University of Wisconsin–Madison

Technical Report Documentation

<table>
<tr><td colspan="2">1. Report No.
CFIRE 07-07</td><td colspan="2">2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.
CFDA 20.701</td></tr>
<tr><td colspan="4" rowspan="2">4. Title and Subtitle
Using Sparse GPS Data to Estimate Link Travel Time for Truck Transport</td><td colspan="2">5. Report Date October 2013</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="4">7. Author/s
Wen Wang, Chun-Wei Lin, Kai Yin, Qi Gong, Teresa M. Adams, and Xiubin Bruce Wang</td><td colspan="2">8. Performing Organization Report No.
CFIRE 07-07</td></tr>
<tr><td colspan="4" rowspan="2">9. Performing Organization Name and Address
National Center for Freight and Infrastructure Research and Education (CFIRE)
University of Wisconsin-Madison
1415 Engineering Drive, 2205 EH
Madison, WI 53706</td><td colspan="2">10. Work Unit No. (TRAIS)</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTRT06-G-0020</td></tr>
<tr><td colspan="4" rowspan="2">12. Sponsoring Organization Name and Address
US Department of Transportation
The Office of the Secretary of Transportation – Research
1200 New Jersey Avenue, SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period Covered
Final Report [November 1, 2012 – October 31, 2013]</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
DTRT06-G-0020</td></tr>
<tr><td colspan="6">15. Supplementary Notes
Project completed by CFIRE and Texas A&M University</td></tr>
<tr><td colspan="6">16. Abstract
Network link travel time estimation for commercial vehicles is essential to freight operations and planning. This project studies link travel time estimation using sparse truck GPS data. First, the research team presents a naïve method for truck speed analysis on individual links. The naïve method can compute the average travel speed and variation on each link individually as input for further travel time analysis. Second, to address the issue of ignoring those truck trips with large intervals covering multiple links, the research team presents a network mapping method that maps the link performances using the itinerary travel times. A case study is conducted using the San Antonio corridor network, a major freight corridor. In addition, the research team further tests the proposed methods with simulation data partially on I-94 and I-894, a highway corridor in the Milwaukee, Wisconsin area. The test results indicate that the proposed network mapping method appears practical for application to the link performance analysis on freight corridors for truck transport.</td></tr>
<tr><td colspan="2">17. Key Words
Travel time estimation , link performance, truck transport, GPS data</td><td colspan="4">18. Distribution Statement
No restrictions. This report is available through the Transportation Research Information Services of the National Transportation Library.</td></tr>
<tr><td colspan="2">19. Security Classification (of this report)
Unclassified</td><td>20. Security Classification (of this page)
Unclassified</td><td colspan="2">21. No. of Pages
49</td><td>22. Price
-0-</td></tr>
</table>

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

DISCLAIMER

Table of Contents

Tables of Figures

Executive Summary

Network link travel time estimation for commercial vehicles is essential to freight operations and planning. Time-dependent link travel time may help carriers as they plan routes and schedule fleet dispatching to avoid traffic congestion, especially recurrent urban congestion. However, travel time estimation for commercial vehicle movement is challenging due to limited data and high variability arising from traffic congestion. Most sensor or loop detector data available do not cover the entire network over a sufficiently long time for quality analysis of network link performance. This project estimates link travel times on a transportation network based on network connectivity and trips' entry/exit locations and time stamps. The entry/exit time stamp of each trip tells the time duration of entire itinerary of a trip. The research team utilizes this limited information to infer the entire network performance by developing appropriate method. The proposed method is expected to be used as an effective network travel time analysis tool for the purpose of improving traffic operations and planning. It may also be further used to assess performance of networks of other modes such as subway or bus systems with passengers' entry and exit information tracked at check in/out points. Major freight corridors where trucks are tracked with GPS periodically with a certain time interval (e.g. 5 or 15 minutes) may also be considered a special case of a network where itineraries overlap with each other on the road sections.

This study uses the sparse truck GPS data in particular for the link travel time estimation. First, we present a naïve method for truck speed analysis on individual links. The naïve method can compute the average travel speed and variation on each link individually as input for further travel time analysis. Second, to address the issue of ignoring those truck trips with large intervals covering multiple links, we present a network mapping method that maps the link performances using the itinerary travel times. We conduct a case study using the San Antonio corridor network, a major freight corridor. In addition, we further test the methods with simulation data partially on I-94 and I-894, a highway corridor in the Milwaukee, Wisconsin area. The test results indicate that the proposed network mapping method appears practical for application to the link performance analysis on freight corridors for truck transport. The research team recommends the mapping method for the future use of network link travel time estimation.

Introduction

Background

Travel time is usually the most important factor when transporting freight from an origin to a destination. Accurate travel time estimation on a transportation network can provide supportive information for decision making in both the private sector (e.g., tracking fleets for trucking companies) and public agencies (e.g., congestion mitigation for urban transportation planners). In particular, link travel time estimation for commercial vehicle movement is essential to both planning and operations in terms of both efficiency and reliability. For example, time of day-dependent link travel time may help carriers as they plan routes and schedule fleet dispatching to avoid traffic congestion, especially recurrent urban congestion.

However, travel time estimation for commercial vehicle movement is always challenging due to limited data and high variability due to recurrent and non-recurrent congestion. Most available sensor or loop detector data do not cover the entire network over a sufficiently long time for quality analysis of network link performance. In addition, these data are typically collected and stored in an aggregate format, not suitable for detailed travel time analysis (McCormack & Hallenbeck, 2006). In the last two decades, the Global Positioning Systems (GPS) data for traffic monitoring and planning has become widely available. Probe vehicles equipped with tracking devices (e.g. GPS or mobile phones) act as mobile traffic sensors. They are capable of tracing a vehicle's travel times on the real-time basis or collecting instantaneous speeds at any network location without need of roadside equipment. These probe vehicles can provide researchers with large amounts of disaggregated traffic data for evaluating link performance and detecting congested roadway locations. The freight industry is aware of the importance of GPS tracking data for freight planning, and many companies have replaced the traditional sheet-based truck trip logs with GPS recordings of the disaggregated trip information (Greaves & Figliozzi, 2008). However, most trucking companies are reluctant to share their data for planning or research purposes due to customer privacy and strategic concerns. Truck tracking GPS data is limited and only available for specific regions (Morgul et al., 2013). Although highway performance data is available from contracted trucking firms and at shorter time intervals than before, the fundamental scientific question still remains: what is the utility of sparse GPS data for evaluating network link performance? How much data is considered sufficient for network link performance analysis?

This project aims to address those issues, but also explore the research problem to a larger context. Since nowadays the wide use of blue tooth and Automated Vehicle Identification (AVI) technologies has created a situation in which vehicles' entry and exit information on a network is known, how to efficiently utilize this information is a meaningful question. Furthermore, efficient utilization of this data may give rise to methodologies about how to most economically obtain necessary data. The entry/exit time stamp of a recorded trip tells the time duration of the entire itinerary. The research group proposes to develop methods to utilize this trip information to infer the link travel times on the entire network. Prior studies on travel time estimation mostly focus on a line structure network, such as arterial roadway with signalized intersections, or a highway corridor comprising of segments with varying traffic conditions. This project proposes to extend the existing travel time estimation techniques to a general network. Note that the travel time estimation methods may also be used to assess the performance of networks of other modes such as subway or bus systems with passengers' entry and exit information tracked at check in/out points. Major freight corridors where trucks are tracked with GPS periodically with a certain time interval may also be considered a special case of a network where itineraries overlap with each other on the road sections.

Study Objectives

This project provides a methodology that uses the limited but available truck trip information, motivated by the importance of estimating network-wide link travel times for freight transport. The project report provides a review of existing methods to estimate link travel times for traffic monitoring or planning purposes. The study uses the empirical truck GPS data that track truck movement in certain time intervals such as 5 or 15 minutes. Each tracked time interval is considered as a trip in this research. The trips overlap with each other on links over a network. The study aims to utilize those overlapping trips to estimate the link travel times along the freight corridors.

This research problem also arises in a larger context. For example, on a toll road, entry and exit information is recorded for each vehicle. In that problem, a large number of itineraries are recorded overlapping with each other, much like subway transit passenger data, for example. Navigable waterway systems share similarities and may be able to apply the method for maritime transport performance analysis. In summary, the research problem in this project arises out of a trucking performance application, but is applicable to many more transportation problems.

Task Overview

The research team completes the following tasks, as described in the original scope of work. These efforts are summarized and discussed in greater detail throughout this report.

1. Review the previous study on roadway travel time estimation.
2. Present the methodology.
    a. Naïve method using speed analysis.
    b. Mapping method using trip travel times.
3. Conduct the case study of San Antonio corridor network.
4. Develop a simulation framework to validate and evaluate the proposed method.
5. Conclude the study and propose future research directions.

Literature Review

This chapter reviews the relevant previous work on roadway travel time estimation. There are several systems commercially available that are capable of estimating roadway travel times on the real-time basis. They can be broadly classified into three categories: spot speed measurement systems, spatial travel time systems, and probe vehicle technologies (Dion and Rakha, 2006). The relevant research based on these systems is introduced as follows.

Spot Speed Measurement Systems

During the last two decades, spot speed measurement systems, specifically inductance loop detectors, have been the main source of real-time traffic information. Substantial research efforts have focused on this indirect estimation of roadway travel times using each vehicle's speed observed at discrete points along the roadway. Other technologies for measuring spot speeds have also evolved, such as infrared and radar technologies. Regardless of the technology, the spot measurement approaches only measure traffic stream speeds over a short roadway segment at fixed locations along a roadway. These spot speed measurements are used to compute spatial travel times over an entire trip using space-mean-speed estimates.

The prominence of this spot speed measurement approach results from the large number of available traffic data provided by inductance loop detectors. Additional efforts have also been made in improving the accuracy of spot speed estimations from single loop detectors (Coifman, 2001; Dailey, 1999; Pushkar et al., 1994). However, even with accurate spot speed estimations obtained (as in the case of using double loop detectors), travel time estimates could still be flawed due to extrapolating spot measurements to a roadway section, with the possibility of different traffic conditions along its length. It is noted that this issue particularly arises on roadways with a low density of detection sites.

Hopkin et al. (2001) suggest that one detector site every half kilometer (i.e. nearly 0.3 mile) of highway is desirable to provide accurate travel time estimates. Several approaches have been developed to overcome the issue and avoid the enormous cost of intensive loop surveillance, such as the identification of vehicle trajectories between loop detectors (Coifman, 2002; Li et al., 2006), and the sensor deployment methods for reliable travel time estimation (Hu et al., 2009; Li and Ouyang, 2011). In addition, it shall also be taken into account that the loop speed estimates in the case of stop and go traffic situations do not adequately represent the space mean speed of the traffic stream. Therefore, the indirect estimation of roadway travel times using spot speed measurement systems has limitations to generate accurate travel time estimates.

Spatial Travel Time Systems

Different from indirect estimation using spot speed measurement, study on travel time estimation using loop detector data has also focused on direct measurement, consisting of measuring the time interval that a particular vehicle takes to travel from one point to another. Many researchers have proposed a smart use of loop detector data, by matching the particular vehicles in consecutive loop detectors based on their characteristic lengths (Coifman and Cassidy, 2002; Coifman and Ergueta, 2003; Coifman and Krishnamurthya, 2007), or particular inductive signature on the detectors (Abdulhai and Tabib, 2003; Sun et al., 1999). However, these techniques would require the upgraded hardware and/or software loop configurations, thus they have not been widely put into practice for operating highway agencies.

Other than the loop detector data, deployment of Intelligent Transportation Systems (ITS) during the last decade has brought the opportunity of using more suitable traffic data to directly measure travel times (Turner et al., 1998). The merging spatial travel time measurement systems use fixed location equipment to automatically identify and track a subset of vehicles in the traffic stream. By matching the unique vehicle identifications at different reader locations, spatial estimates of travel times can be computed. This is the case with Automated Vehicle Identification (AVI) data obtained from the readings of vehicle toll tags or from video license plate recognition. By matching the vehicle ID at different locations on highway, link travel times can be directly obtained if the clocks at each location are properly synchronized.

The concept of using AVI data from toll collection systems to directly measure highway travel times was first proposed by Davies et al. (1989). A large number of literatures primarily deal with the usage of Electronic Toll Collection (ETC) data to measure travel times. These systems identify the vehicles by means of on-vehicle electronic tags and roadside antennas located on the main highway sections. However, the basic problems of this configuration include the level of market penetration of the electronic toll tags, and how to deal with time periods when only small samples are available in order to obtain a continuous measurement of travel times (Dion and Rakha, 2006).

Some additional research has also been done on the travel time measurement using the typical configuration of a closed toll system, which is widely extended in Europe and Asia (Ohba et al., 1999). The concept of a closed toll system refers to the fact that the toll a particular driver pays varies depending on the origin and destination of his trip and is approximately proportional to the distance traveled on the highway. Soriguera et al. (2010) presented a new approach for measuring travel times on closed toll highways using the existing surveillance infrastructure. Since the toll plazas were located on the on/off-ramps and each vehicle was charged a particular fee depending on its origin and destination, the data from toll collection system were filtered and fused in a statistical way in order to extract the relevant itinerary travel time information. Their proposed method allows estimating the travel times on single sections of highway using itineraries covering different origin-destinations.

Probe Vehicle Technologies

Another approach for directly measuring travel times is to use probe vehicle technologies, which are capable of tracking a sample of probe vehicles as they travel within a transportation network. The emerging technologies include cellular geo-location, global positioning systems (GPS), and automatic vehicle location (AVL) systems. Those probe vehicles act as mobile traffic sensors equipped with tracking devices (e.g. GPS or mobile phones). They are being used to collect network-wide traffic information such as instantaneous speeds and travel times at any network location without need of roadside equipment.

Prior research using GPS data or probe vehicle data have addressed the route inference in map-matching processes (Yokota & Tamagawa, 2012; Rahmani & Koutsopoulos, 2013), and the number of probe vehicles needed to reflect realistic traffic conditions (Srinivasan & Jovanis, 1996; Chen & Chien, 2000). A significant body of literature focuses on model-based and data-driven methods for estimating travel time with probe vehicle data for traffic monitoring or planning purposes (Zheng & Van Zuylen, 2013).

A mathematical model by Jula *et al.* (2008) estimates link travel times and arrival times at nodes on a real-time, stochastic network. Hellinga *et al.* (2008) proposed an analytical model to decompose partial link or route travel time from probe vehicle into individual link travel times along urban arterials utilizing the real traffic conditions on arterial network. Their evaluation suggested that the proposed method outperformed the benchmark (deterministic) method. Based on data collected from highly frequent GPS readings, Westgate *et al.* (2013) proposed a Bayesian model to estimate the distribution of ambulance travel times on road segments in

Toronto. Their methods can provide additional support information such as the probability an ambulance will reach its destination within a certain time threshold.

Different from the conventional loop detector data, probe vehicle data does not provide direct information about flow, density, and average speeds that are usually the inputs for analytical models. Instead, people use data-driven methods for travel time estimation. The existing data-driven methods include regression models (Chan et al., 2009) and neural network based models. Zheng and Van Zuylen (2013) proposed a three-layer neural network model to estimate link travel times for individual probe vehicles. The results with simulated data suggest their model can outperform the analytical model. However, the many required parameters associated with those models limit their applicability in practice.

Instead of exclusively utilizing probe vehicle data, Bhaskar *et al.* (2009) proposed a model to incorporate probe vehicle data into traditional cumulative plots in order to estimate the average travel times on signalized urban network. Zhang and Rice used data from both probe vehicles and double loop detectors to develop a linear model for travel time prediction on freeways (Zhang & Rice, 2003). To address the issue of sparse truck GPS data available, Morgul *et al.* (2013) presented an empirical method for truck travel time estimation by using taxi GPS data to supplement the limited truck GPS data on the Manhattan network. Their results indicated that the taxi GPS data supplemented the sparse truck data well. A summary of other relevant literature using GPS-based probe vehicle data for traffic monitoring or transportation planning purposes is also provided by Morgul et al*.,* as shown in Table 1.

Table 1: Literature using GPS-Based Probe Vehicle Data (Source: Morgul et al., 2013)

Authors	**Year**	**Data Size**
Traffic Monitoring		
Zou et al.	2005	100 taxis
Brockfield et al.	2007	700 taxis
Lahrmann	2007	1 month taxi-GPS data, 10 taxis
Liu et al. *(14)*	2008	4000 taxis
Sananmongkhonchai et al.	2008	1 month taxi-GPS data, 4000 taxis
Hunter et al.	2009	2 months taxi-GPS data, 50 taxis
Li et al.	2009	57 days taxi-GPS data, 7000 taxis
Liu et al.	2009	6 months taxi-GPs data, 249 taxis
Herring et al.	2010	3 months taxi-GPS data, 500 taxis
Yuan et al.	2010	3 months taxi-GPS data, 33000 taxis
Ehmke et al.	2012	2 years taxi-GPS data
Miwa et al.	2012	6 months taxi-GPS data, 1500 taxis
Yokota and Tamagawa	2012	1 month truck-GPS, 300 trucks
Transportation Planning		
Li et al.	2009	57 days taxi-GPS data, 7000 taxis
Uno et al.	2009	10 day bus-GPS data, 100 buses
Kinuta et al.	2010	6 months taxi-GPS, 1 month truck-GPS, 70 taxis, 130 trucks
Munehiro et al.	2012	1 year taxi-GPS data, 115 taxis

Methodology

In this chapter, two methods are presented to estimate the link travel times on a network using truck tracking GPS data: one is the naïve method for truck speed analysis on individual links; the other is the network method that maps the recorded trip travel times onto a link time matrix using the least squares method (LSM). The motivation of application here is for link performance analysis on freight corridors using truck GPS data.

Naïve Method Using Speed Analysis

This method computes the travel time on each link individually, as summarized below.

1. Identify the GPS records falling onto a single link along each truck's travel trajectory.
2. Calculate the truck speed between two consecutive GPS records by dividing the traveled link length by the recorded time interval.
3. Group all resulting speeds traversing each link for each time period to get the mean speed using $t_{a,k} = \frac{l_a}{\bar{v}_{a,k}}$, where $t_{a,k}$ denotes the estimated travel time on link *a* at time period *k*, l_a denotes the roadway length for link *a*, and $\bar{v}_{a,k}$ denotes the statistical speed estimate (e.g. mean speed) on link *a* at time period *k*.

Note that this way may also give rise to the median and standard deviation of the speeds for a particular link. However, this method is rough because two consecutive GPS records may cover a portion of two or more links. In the naïve method, we choose not to use the time intervals that correspond to three or more road links for which the travel time is estimated.

This naïve method can easily identify the extreme speed records (due to noises or technical errors) compared with the posted link speed limits. Besides, the temporal speed variation on each link can be used for travel time reliability analysis. There are several disadvantages of the naïve method. First, the sparse truck GPS data may lead to only a small sample of intra-link trip records, thus affect the statistical estimate of truck speed/travel time on a single link. Second, by ignoring the records of long time intervals, a chance for improving the time estimation is missed. In contrast, the second method, network mapping method, which we are presenting next, uses all the trip records including those of large time intervals and covering three or more links. The numerical tests also confirm that the network mapping method provides better estimates of link travel time.

Mapping Method Using Trip Travel Times

Considering the recorded travel time for a particular trip as the sum of the travel times on the links comprising this trip, a mapping relationship is presented as follows.

$$c_i = \sum\nolimits_a \delta_{i,a} t_a$$

where c_i denotes the recorded travel time of truck trip *i*, t_a denotes the travel time on link *a*, $\delta_{i,a}$ is the fraction of the total length of link *a* traversed in trip *i*. $0 \leq \delta_{i,a} \leq 1$, $\delta_{i,a} = 1$ if link *a* is fully on trip *i*, and $\delta_{i,a} = 0$ if link *a* is not traversed by trip *i*.

Here, we use the term trip to mean sequential GPS records denoting location and time stamps for start and end points. If the GPS record interval is 15 minutes, for example, then each 15 minute interval is a trip in this study. A trip must include both start and end stamps. We only consider 'elementary' trips for use in this study; for example, three consecutive GPS points form two elementary trips. However, if the consecutive elementary trips are along one link then for convenience, we create a single trip using the first and the last records of the sequence.

Let *P* be the incidence matrix with elements $\delta_{i,a}$. Typically, the number of rows in this incidence matrix equals the number of trips, and the number of columns equals the number of links on the study network. Using vector operation, the mapping can be written in matrix notation as:

$$\boldsymbol{c} = \boldsymbol{P}\boldsymbol{t}$$

However, as a practical challenge, the positions of trucks within the beginning and ending links are randomly distributed in the GPS data. Travel times collected by GPS may partially cover the beginning and ending links. Technical preprocessing is needed to deal with this 'end' effect. We use $0 < \delta_{i,a} < 1$ to record partial link coverage.

For example, Figure 1 shows three trips reported from truck GPS device along three links. For trip (a), the first GPS point corresponds to a distance equal to 1/4 of Link 1, the second GPS point corresponds to the 3/4 distance on Link 1. For trip (b), the first GPS point is at 1/3 length of Link 1, the second GPS point is at 3/4 length of Link 2. We will explain trip (c) later in this example.

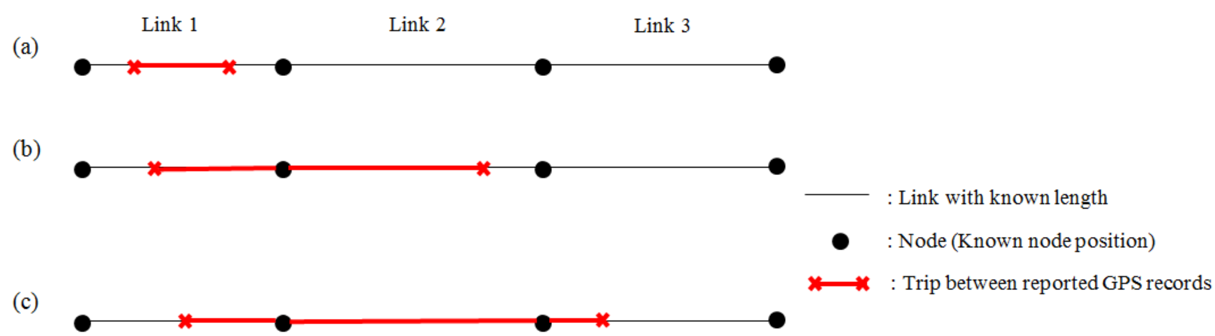


Figure 1: Illustration of Recorded Truck Trips

In the preprocessing, we assume the travel time on an individual link is proportional to the covered distance on that link. For example, since trip (a) in Figure 1 covers half the length of Link 1, the trip time is half the required travel time to traverse Link 1 and nothing else. We can use GIS tools to measure the proportional distances along cartographic representations of roadway links. When there are many trips, the amount of work for this preprocessing is non-trivial.

We can generate the following incidence matrix, *P*, to map the relationship between the trips and links in Figure 1:

$$P = \begin{bmatrix} 1/2 & 0 & 0 \\ 2/3 & 3/4 & 0 \\ 1/2 & 1 & 1/4 \end{bmatrix}$$

where the columns correspond to Links 1, 2 and 3, and the rows correspond to the trips (a), (b), and (c). The third row indicates that for trip (c), the first GPS point occurs at the 1/2 point on Link 1 and the second GPS point occurs at 1/4 point on Link 3. The 1 in element (3, 2) indicates the truck traveled the entire length of Link 2.

Then, the model formulation for estimating of link travel times using the above mapping relationship is proposed as follows.

Objective function: To minimize the total difference between the recorded trip times and the estimated trip times for all trips.

Assumption: The truck travel time on an individual link is proportional to the covered traveled distance on that link. This assumption also implies that each link has consistent traffic condition and sufficient truck GPS records.

Notations

Set

- A: link set
- I: trip set

Parameters

- t_a^0 free flow travel time to traverse link *a*
- c_i observed travel time on trip *i*
- $\delta_{i,a}$ elements comprising the incidence matrix *P*

Variables

- t_a estimated travel time on link *a*
- r_i estimated travel time for trip *i*

Model formulation:

Minimize

$$\sum_{i \in I} \|r_i - c_i\|^2$$

Subject to

$$\sum_{a \in A} \delta_{i,a} t_a = r_i \qquad i \in I \tag{1}$$

$$t_a \geq t_a^0 \qquad a \in A \tag{2}$$

The constraint (1) represents the mapping relationship between trip times and link travel times, and the constraint (2) represents the minimum free flow travel time restrictions. With sufficient trip records across the study network, the link travel times can be estimated simultaneously using least squares method (LSM) to minimize the sum of squared errors between the given itinerary travel times and estimated itinerary times. Considering link travel time must be no less than the free flow travel time, we solve the LSM with bound constraints by coding in MATLAB.

Case Study of San Antonio Corridor Network

Truck Data

The truck tracking GPS data used in this study is from the American Transportation Research Institute (ATRI) data collected in March 2009 along highway I-35 from Laredo to San Marcos, Texas. It covers a major freight corridor in the Southwest Texas, as illustrated in Figure 2. This corridor carries a significant volume of freight for trade with Mexico, and also connects several major metropolitan areas such as Austin and San Antonio. The raw data include unique truck ID, the longitude and latitude position, and the time stamp when the location of truck is recorded. Participating trucks record their GPS information (both location and time) at different time intervals (e.g., 1, 3, 5 minutes, etc.) after their engines are turned on until the travel stops. The overall temporal distribution of collected truck GPS data during March 2009 are displayed in Figure 3 and Figure 4.

Our pilot study finds that the highway corridor close to San Antonio metro area has varying speeds/travel times with time of day, and truck travel time varies significantly along I-35 and the alternative route, highway loop I-410 across the San Antonio metropolitan area. Here, we focus on the San Antonio corridor network as shown in Figure 5. The San Antonio network is chosen also because of the overlapping itineraries recorded via GPS that make it an ideal application of the mapping method that we propose. Those periodical truck GPS records can be used to estimate the link travel times across the San Antonio network.

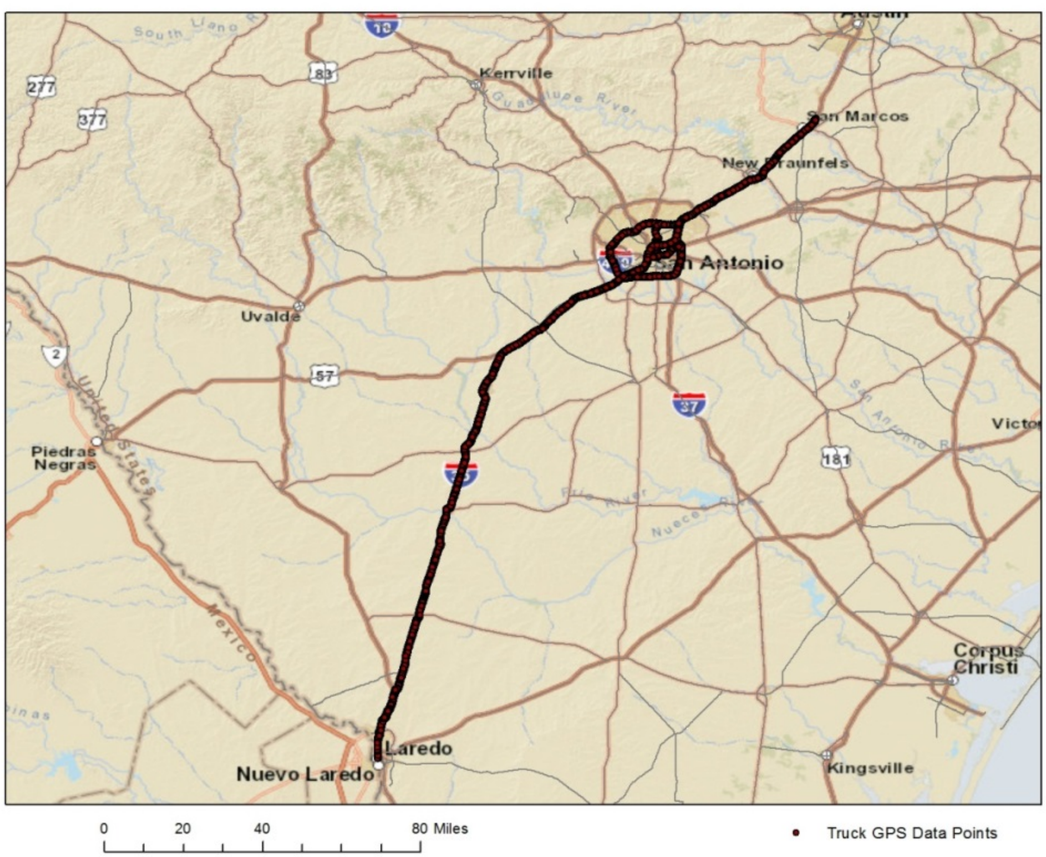


Figure 2: Scope of Freight Corridor in the Southwest Texas

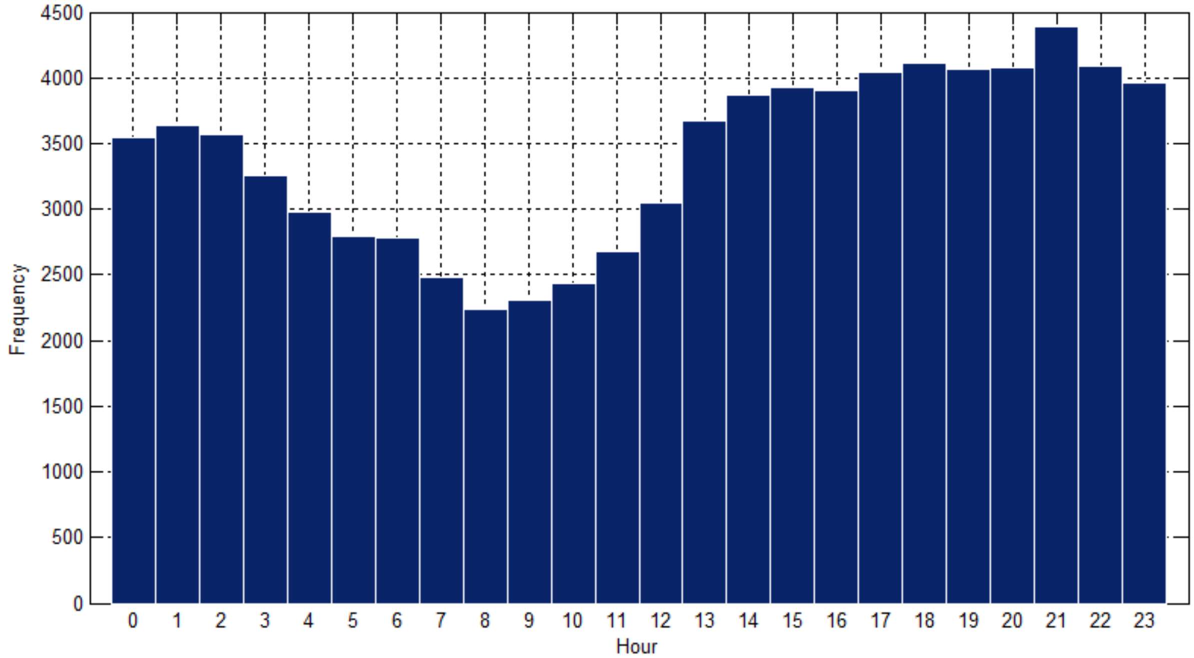


Figure 3: Distribution of Texas GPS data during March 2009 by hour

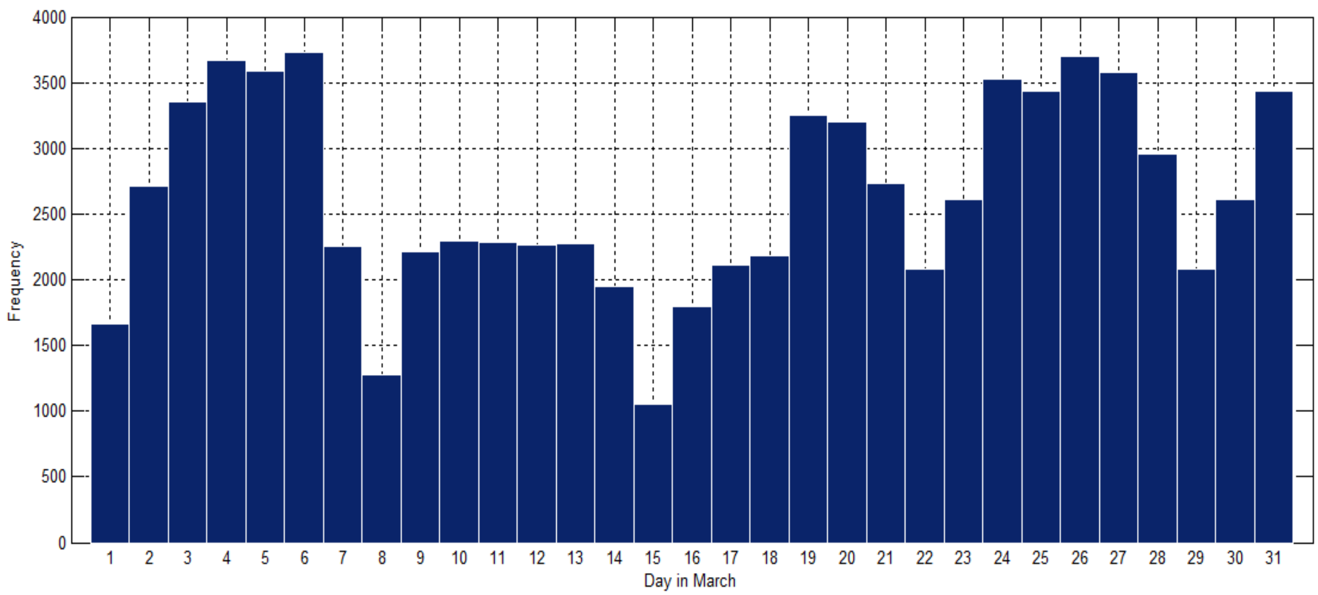


Figure 4: Distribution of Texas GPS data during March 2009 by day

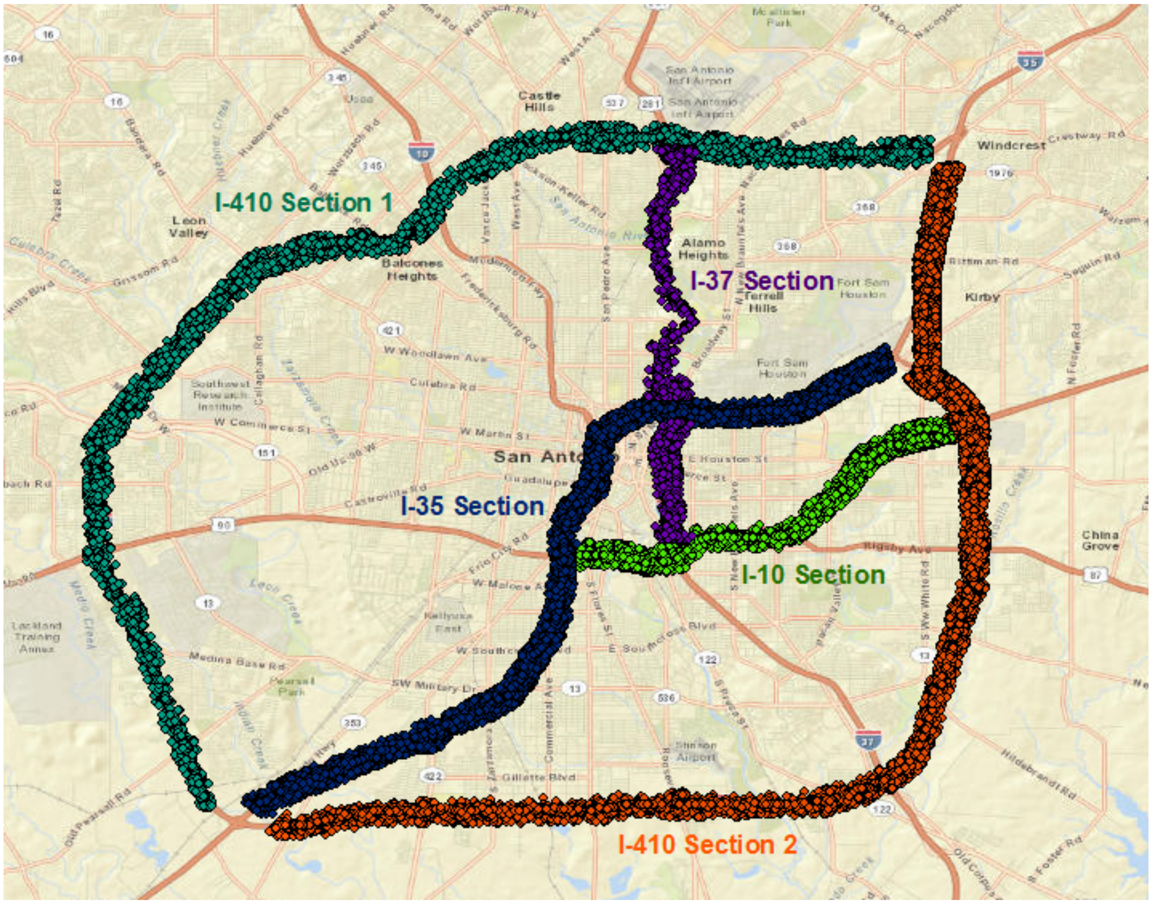


Figure 5: Study Corridors Highlighted on the San Antonio Network

Data Preprocessing

The raw GPS data contained records of time and location for fueling, local deliveries and pickups and related activities off the highway corridor. Data preprocessing removed the records for positions falling outside the mainline highway. The filter procedure is as follows.

1. Group truck data by their ID and recording time.
2. Use ArcGIS geo-processing tool to identify data points not falling onto highway mainline. Remove the records for travel not on the highway with ArcGIS.
3. Detect the outlier GPS records using Chauvenet's criterion (Barnett & Lewis, 1994). The criterion was applied for speed measurements along a single link during one time period.

Chauvenet's criterion is a statistical means widely used for detecting outliers in measurement data. Given a sample of N data points, the criterion states that data points can be considered for rejection as outlier only if the probability of obtaining their deviation from the mean is less than 1/(2N). By following this procedure, the raw GPS data can be filtered for use in the link travel time estimation.

Link Travel Time Estimation with Two Methods

In the case study on San Antonio corridor network, we choose to only focus on the GPS records traveling northbound. Sections of I-35, I-37, I-10 and I-410 on the study network were divided into links based on their conjunction with other major roadways. Accordingly, I-35 comprises five links (351 to 355), I-410 section 1 comprises four links (411 to 414), I-410 section 2 comprises six links (421 to 426), I-10 comprises three links (101 to 103), and I-37 comprises two links (371 and 372). The link numbering sequence along each highway section increases in the northbound direction. A summary of the link lengths and free-flow travel times is presented in Table 2.

To compare the travel time estimates using the two methods, we used the truck trip records during the afternoon peak hours from 5pm to 7pm. There are 231 trucks (unique truck IDs) with nearly 2,000 filtered GPS data points during this period. For the naïve method, we code in ArcGIS to aggregate all the speed results obtained from data points falling onto a single link. In this way, consecutive GPS records covering multiple links are ignored. Then we computed a statistical summary of the truck speed on each link (i.e., mean, median, standard deviation, and percentile).

For the mapping method using the least squares method, we developed the incidence matrix as explained earlier by coding in ArcGIS and MATLAB. This mapping method uses more valid trip data (i.e., inter-link trip data) compared with the naïve method. Table 3 compares the numbers of valid data points used in the two methods.

We compare the link travel time estimates obtained from the two methods in Table 2. As shown, the naïve estimates are found by dividing the link lengths on the mean speeds. The mapping estimates consider both intra- and inter-link trips.

It can be seen that generally the estimates from the two methods are very close. However the estimated travel time from the naïve method for link 425 deviates from the mapping estimate by nearly 20 percent. This is due to a very small sample of available intra-link truck trips (only 31 trip records) falling on link 425 during this period. Therefore, the resulting travel time estimate for link 425 with naïve method would be inaccurate in this case, which also implies the limitation of naïve method to deal with sparse GPS data.

Both methods are capable of generating reasonable travel time estimates for links 355, 426, 103 and 371 with recurrent heavy traffic during the afternoon peak. For these links the free flow travel time is significantly less than the estimated travel times using observed data. Those links conjunct with another highway section or connect the dense roadways in downtown San Antonio. The detailed truck speed distribution obtained from the naïve method along those links during four aggregated daily periods (AM Peak: 7:00-9:00, Noon: 11:00-13:00, PM Peak: 17:00-19:00, Midnight: 22:00-24:00) are also displayed in Figures 6 and 7.

Table 2: Link Travel Time Estimates with Two Methods for Trips during 5-7 PM

	Link No.	Length (miles)	Free-flow Travel Time (sec)	Naïve Estimate (sec)	Mapping Estimate (sec)
I-35 Section	351	4.2	228	265	271
	352	3.8	210	290	284
	353	2.5	138	156	151
	354	1.8	90	96	109

	355	4.0	216	388	407
I-410 Section 1	411	7.3	402	416	422
	412	9.3	510	550	515
	413	4.5	246	335	331
	414	4.9	270	319	307
I-410 Section 2	421	3.2	174	201	188
	422	4.6	252	317	326
	423	3.7	204	287	304
	424	4.9	270	319	325
	425	3.2	174	275	340
	426	2.7	150	349	351
I-10 Section	101	2.0	108	184	189
	102	2.3	126	214	215
	103	3.6	198	334	343
I-37 Section	371	3.1	174	378	403
	372	5.8	324	391	381

Table 3: Summary of Valid Data Points in Two Methods

	Trip Records Used
Naïve Method	1361
Mapping Method	1854

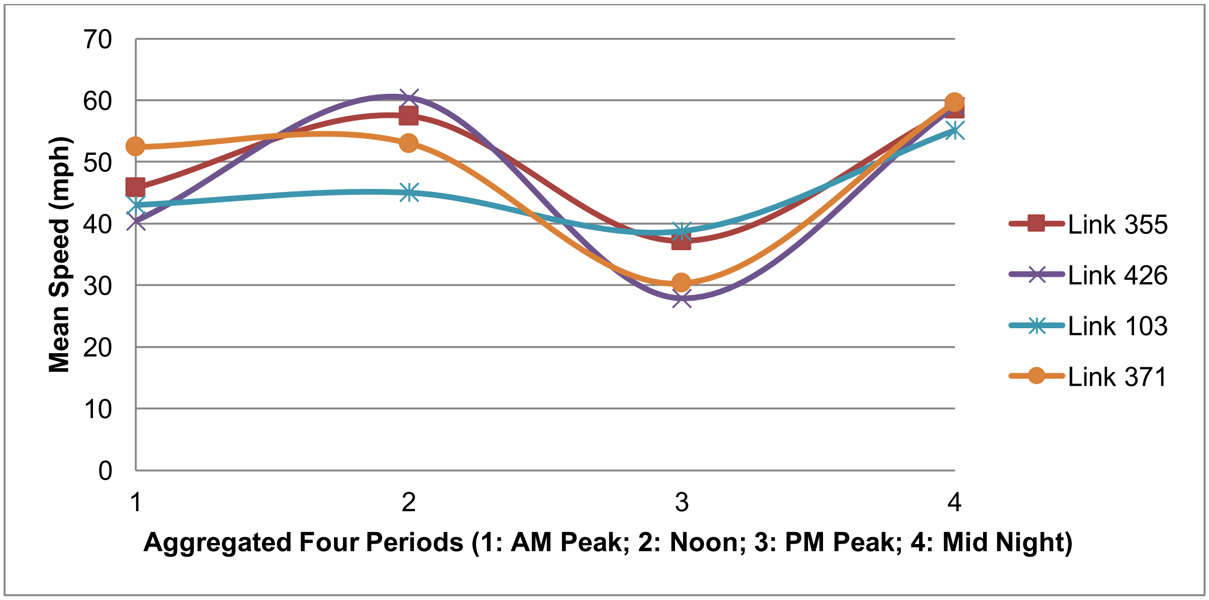


Figure 6: Mean Speed Distribution on the Links with Recurrent Heavy Traffic

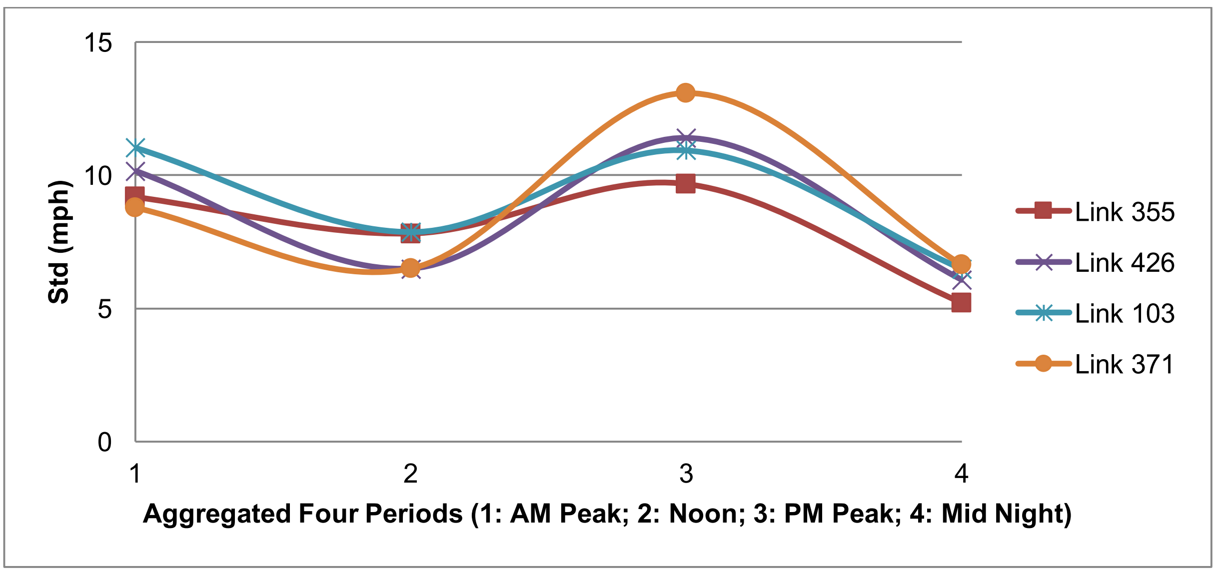


Figure 7: Standard Deviation of Speed on the Links with Recurrent Heavy Traffic

The specific statistical summary of speed distribution during four aggregated periods on these four links with recurrent heavy traffic is listed in Table 4.

Table 4: Statistical Summary on Four Links with Recurrent Heavy Traffic

	Period	Mean (mph)	Median (mph)	Std Dev (mph)
Link 355	1	45.78	49.76	9.18
	2	57.42	55.27	7.81
	3	37.15	33.41	9.67
	4	58.60	62.45	5.20
Link 426	1	40.41	46.43	10.16
	2	60.37	58.60	6.49
	3	27.84	35.01	11.40
	4	58.96	60.36	6.06
Link 103	1	43.03	46.74	11.04
	2	45.03	50.28	7.87
	3	38.78	42.80	10.94
	4	55.17	59.72	6.48
Link 371	1	52.40	45.14	8.79
	2	53.01	54.54	6.51
	3	30.34	37.14	13.09
	4	59.55	56.39	6.64

We further test the applicability of the mapping method with simulated data from the calibrated highway corridor in the Milwaukee, Wisconsin area in next chapter.

Simulation Test

To validate and evaluate the proposed estimation methods, a simulation framework for a highway corridor in the Milwaukee, Wisconsin area was developed using VISSIM. As shown in Figure 8, the Milwaukee corridor network in the simulation framework includes sections of I-94 and I-894. This corridor starts from County Highway BB to West Loomis Road. It is calibrated with the two-hour loop detector data on the eastbound sections of the highway and ramps during the peak-hour period from 4 to 6 PM on March 6, 2009, which contain the traffic volume and average travel speed on each detector point in 15-minute intervals. The locations of the data detectors are indicated at the round green points in Figure 8. The three study links for the test are selected and shown in Figure 8 with red stars indicating the start and end points. The link lengths and free flow travel times are listed in Table 5. This simulation framework generates the eastbound vehicle travel trips for the two-hour period. We tested the methods using data for the second-hour trips.

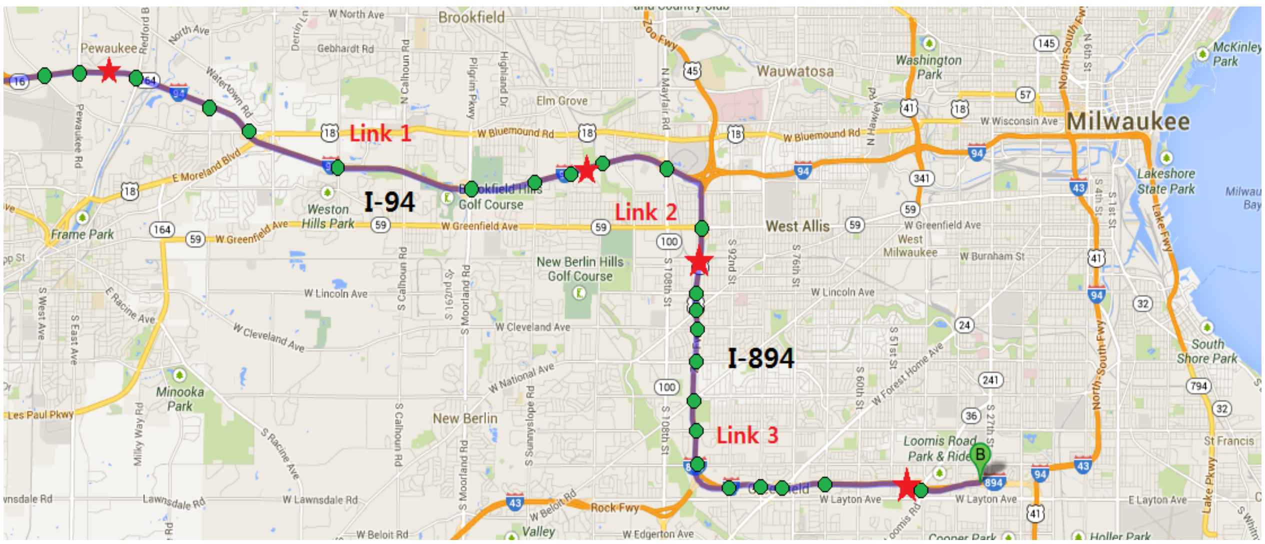


Figure 8: Scope of Highway Corridor and Illustration of Study Links in Milwaukee, WI

Table 5: Link Information along Test Corridor

	Length (miles)	Free-Flow Time (sec)	Ground Truth (sec)
Link 1	7.5	387	639
Link 2	3.5	178	901
Link 3	4.8	245	472

The travel time distribution for the vehicles passing through each of the three study links are shown in Figures 9-11. Notice that although Link 2 is the shortest link of all three links, it has the longest average travel time to pass through.

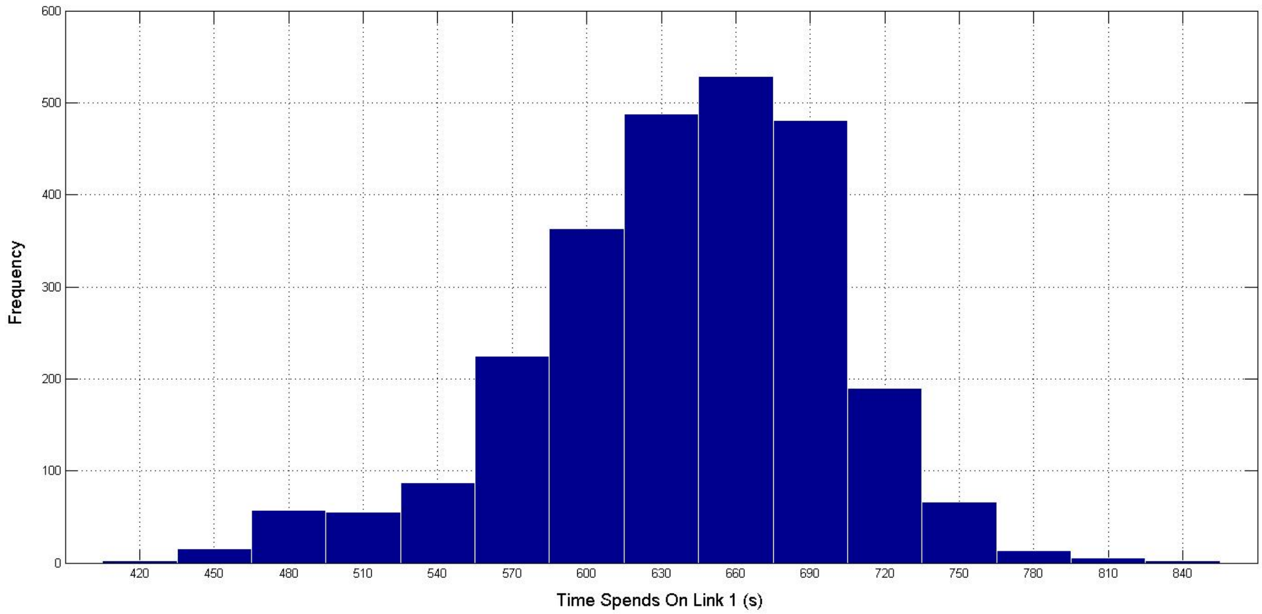


Figure 9: Travel Time Distribution for the Vehicles Passing Through Link 1

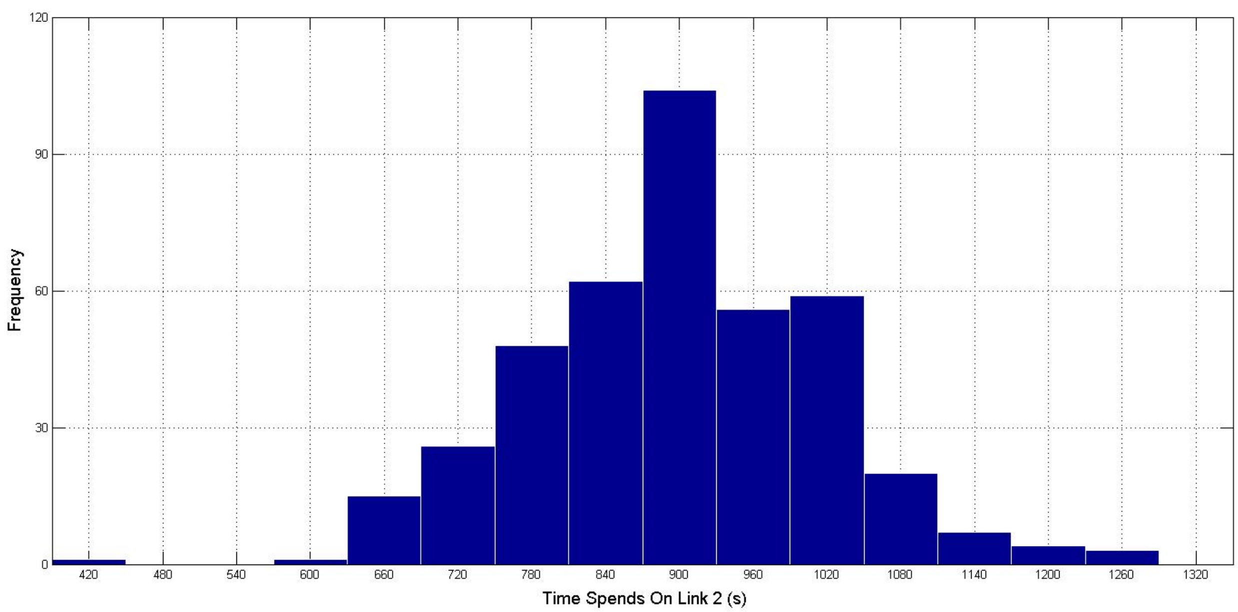


Figure 10: Travel Time Distribution for the Vehicles Passing Through Link 2

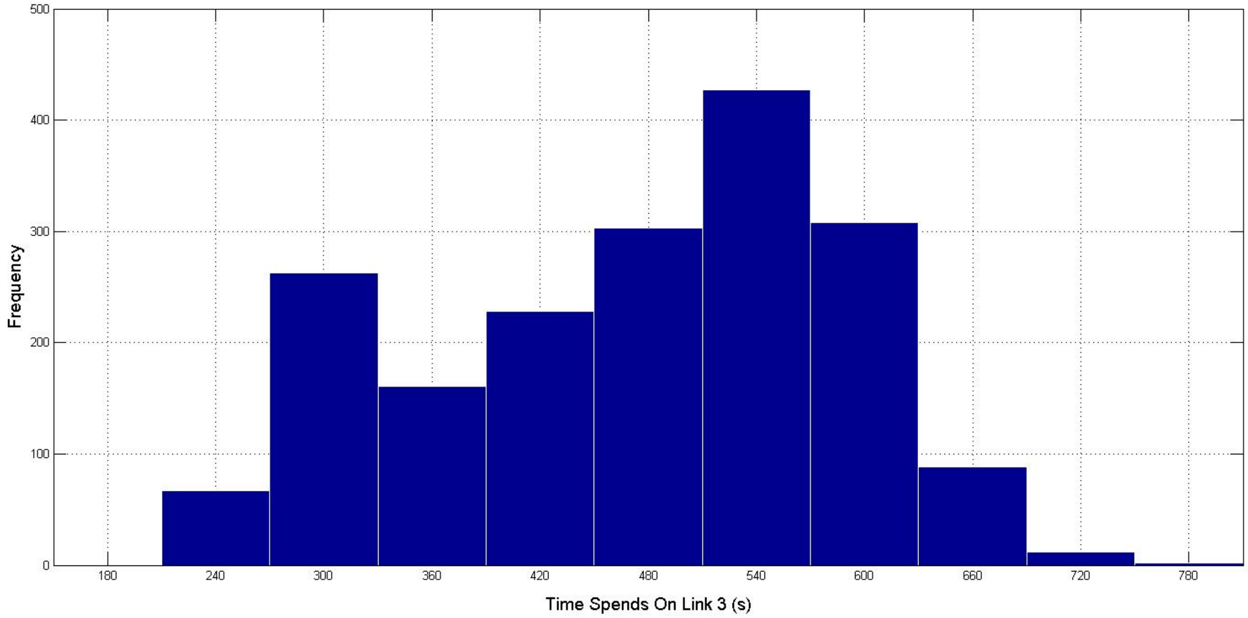


Figure 11: Travel Time Distribution for the Vehicles Passing Through Link 3

For each test, we extract a sample of the vehicle trips along the corridor, and randomly select from the position and time records along the study corridor for each vehicle so that we have sparse and non-uniform position and time records for the vehicle trips, such as the example shown in Figure 12. It is worth noting the significant decrease in speeds between points 7 and 12 due to the traffic merging from the ramps. Hence, the density of the recording points on this part of highway is set to be higher than others. The tested data simulates the realistic situation where the probe vehicles report travel information that is randomly distributed.

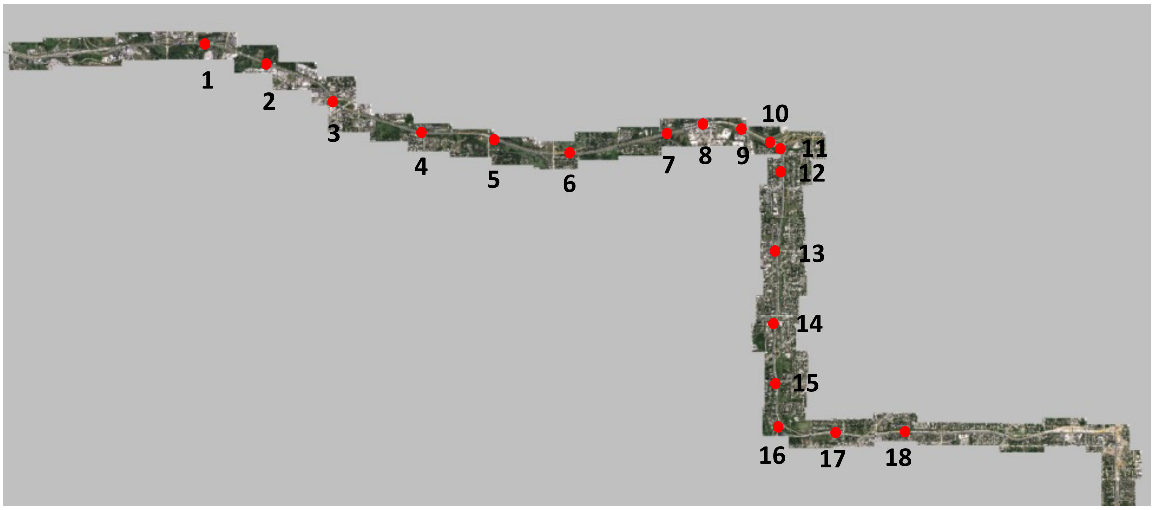


Figure 12: Locations of Those Recording Points along the Test Corridor

The distribution of the number of records collected from each of the data collection point is shown in Figure 13. The number of records collected from data collection point 11 is significantly less than all the other data collection points since it is on the collector ramp that connects from eastbound I-94 to southbound I-894.

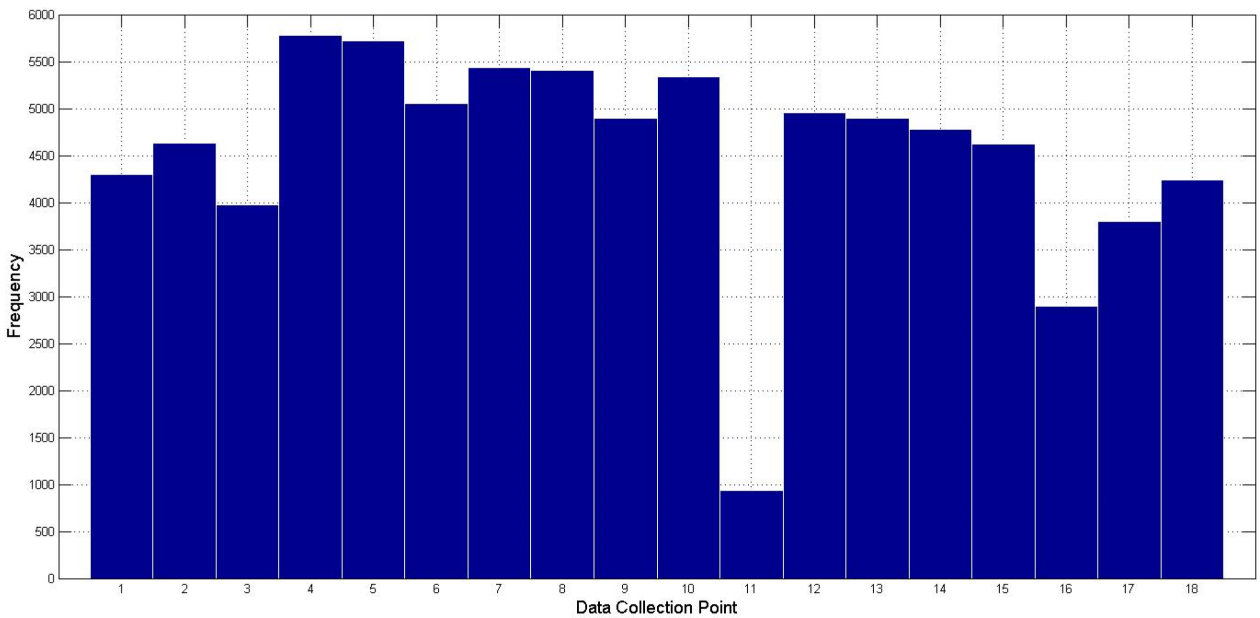


Figure 13: Distribution of Records in the Second Hour Collected by Each Data Collection Point

We applied both the naïve and mapping methods to the sampled vehicle trips to estimate the link travel times. To compare the estimation accuracy of the methods, we tested several cases with varying sample sizes. The ground truth values of link travel times were obtained from the population of nearly 4500 vehicle trips in total, and we tested samples representing 70, 40, and 10 percent of the trips. For each case, we computed the relative gap between the estimated link travel time and the ground truth value as a measure of accuracy.

Results of the accuracy tests are shown in Table 6. The estimated travel times from the mapping method are more accurate (with smaller gaps) than the estimates from naïve method. The naïve estimates deviate from the ground truth values by more than 20 percent when the sample accounts for only 10 percent of the population, while the mapping method estimates are within 10 percent. These results indicated the proposed network mapping method is promising for future practical applications.

Table 6: Link Travel Time Estimates with Two Methods for Simulation Trips

Test Sample (%)	Link No.	Naïve Method		Mapping Method	
		Estimate (sec)	Gap	Estimate (sec)	Gap
70	1	589	7.8%	615	3.7%

	2	827	8.2%	846	6.1%
	3	429	9.1%	500	5.9%
40	1	570	10.8%	591	7.5%
	2	1001	11.2%	937	4.0%
	3	373	21.0%	440	6.8%
10	1	523	18.1%	558	12.6%
	2	1086	20.6%	989	9.8%
	3	332	29.7%	412	12.7%

Conclusions

The link travel time estimation for commercial vehicle movement is challenging for two reasons: limited available data and high variability of travel time due to recurrent and non-recurrent roadway congestion. In this project, the research team presents a method for estimating link travel time using limited truck GPS data. The method has potential practical application for analyzing the link level performance on freight corridors.

A simple and basic method, referred to as the naïve method in this report, computes the average travel speed and variation on each link individually. The naïve method cannot make use of sparse trip data having intervals covering multiple links. The proposed mapping method solves for link travel times based on trip itinerary information using the least squares method.

The research team conducts a case study of San Antonio corridor network, and applies the methods to estimate link travel times along the freight corridors. The results show that the estimates from the two methods are generally close, and both are capable of generating reasonable travel time estimates even for links with recurrent heavy traffic. The naïve estimate method uses less of the available data in a sparse data set than the mapping method. As a result the naïve method may yield estimates that are less accurate than the mapping method. The authors further test the methods with simulation data from a calibrated highway corridor in the Milwaukee, Wisconsin area. The test results show that for varying sample sizes the estimates with mapping method are more accurate than those with naïve method.

While the network mapping method appears to be a promising tool for future practical application on freight corridors, much remains to be addressed. For example, can we use small sample data to ensure reliable travel time estimate? What is the minimum size of data for reliable estimates? Can we estimate link travel times that vary over the time of study period? The authors will make efforts to address those issues as future research directions.

References

Abdulhai, B., Tabib, S.M. (2003). *Spatio-Temporal Inductance-Pattern Recognition for Vehicle Reidentification*. Transportation Research Part C 11 (3–4), 223–239.

Bar-Gera, H. (2007). *Evaluation of a Cellular Phone-Based System for Measurements of Traffic Speeds and Travel Times: a Case Study from Israel*. Transportation Research Part C, 15 (6), 380-391.

Barnett, V., and T. Lewis (1994). *Outliers in Statistical Data*. 3rd edition. Chichester: J. Wiley and Sons.

Bertini, R., S. Tantiyanugulchai (2004). *Transit Buses as Traffic Probes: Empirical Evaluation Using Geo-Location Data.* Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 1870, pp. 35–45.

Bhaskar, A., E. Chung, and A. G. Dumont (2009). *Travel Time Estimation on Urban Networks with Mid-Link Sources and Sinks.* Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 2121, pp. 41–54.

Brockfield, E., Passfeld, B., Wagner, P. (2007). *Validating Travel Times Calculated on the Basis of Taxi Floating Car Data with Test Drives.* 14th ITS Conference, October 9-13, 2007, Beijing, China.

Chan, K.S., W. H. K. Lam, and M. L. Tam (2009). *Real-Time Estimation of Arterial Travel Times with Spatial Travel Time Covariance Relationships.* Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 2121, pp. 102–109.

Chen, M., and S. I. J. Chien (2000). *Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation.* Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 1719, pp. 61-68.

Coifman, B. (2001). *Improved Velocity Estimation Using Single Loop Detectors.* Transportation Research Part A 35 (10), 863–880.

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

Coifman, B., Cassidy, M. (2002). *Vehicle Reidentification and Travel Time Measurement on Congested Freeways.* Transportation Research Part A 36 (10), 899–917.

Coifman, B., Ergueta, E. (2003). *Improved Vehicle Reidentification and Travel Time Measurement on Congested Freeways.* ASCE Journal of Transportation Engineering 129 (5), 475–483.

Coifman, B., Krishnamurthya, S. (2007). *Vehicle Reidentification and Travel Time Measurement across Freeway Junctions Using the Existing Detector Infrastructure*. Transportation Research Part C 15 (3), 135–153.

Dailey, D.A. (1999). *Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements.* Transportation Research Part B 33 (5), 313–322.

Davies, P., Hill, C., Emmott, N. (1989). *Automatic Vehicle Identification to Support Driver Information Systems.* Proceedings of IEEE Vehicle Navigation and Information Systems

Conference, 11–13 September, pp. A31–A35.

Dion, F., Rakha, H. (2006). *Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data for Low Sampling Rates*. Transportation Research Part B 40 (9), 745–766.

Ehmke, J. F., Meisel, S., Mattfeld, D. C. (2012). *Floating Car Based Travel Times for City Logistics.* Transportation Research Part C, Vol. 21, pp. 338-352.

Greaves, S. P., and M. A. Figliozzi (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, Transportation Research Board of the National Academies, Washington, D.C., No. 2049, pp. 158-166.

Hellinga, B., P. Izadpanah, H. Takada, and L. Fu (2008). *Decomposing Travel Times Measured by Probe Based Traffic Monitoring Systems to Individual Road Segments.* Transportation Research Part C, Vol. 16, No. 6, pp. 768-782.

Herring, R., Hofleitner, A., Abbeel, P., Bayen, A. (2010). *Estimating Arterial Traffic Conditions Using Sparse Probe Data.* 13th International IEEE Conference on Intelligent Transportation Systems, Madeira Island, Portugal, pp. 929-936.

Hopkin, J., Crawford, D., Catling, I. (2001). *Travel Time Estimation.* Summary of the European Workshop Organized by the SERTI Project, Avignon.

Hu, S.-R., Peeta, S., Chu, C.-H. (2009). *Identification of Vehicle Sensor Locations for Link-Based Network Traffic Applications*. Transportation Research Part B 43 (8-9), 873–894.

Hunter, T., Herring, R., Abbeel, P., Bayen, A. (2009). *Path and Travel Time Interference from GPS Probe Vehicle Data.* Neural Information Processing Systems Foundation (NIPS), Vancouver, Canada.

Jula, H., M. Dessouky, and P. A. Ioannou (2008). *Real-time Estimation of Travel Times along the Arcs and Arrival Times at the Nodes of Dynamic Stochastic Networks.* IEEE Transactions on Intelligent Transportation Systems, Vol. 9, pp. 97-110.

Kinuta, Y., Kitamura, S., Nakamura, T., Makimura, K., Takahashi, M., Morokawa, T. (2010). *Examination of the Applicability of Probe Car Data to Assessment of the Effects of Road-Safety Projects.* International Journal of Intelligent Transportation Systems Research, Vol. 8, pp. 67-76.

Lahrmann, H. (2007). *Floating Car Data for Traffic Monitoring*. Towards Intelligent Future; Easy Way; i2TERN Conference, 20-21 June 2007, Aalborg, Denmark.

Li, M., Zhang, Y., Wang, W. (2009b). *Analysis of Congestion Points based on Probe Car Data*. Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St.Louis, MO, USA, October 3-7, 2009.

Li, R., Rose, G., Sarvi, M. (2006*). Evaluation of Speed-Based Travel Time Estimation Models*. ASCE Journal of Transportation Engineering 132 (7), 540–547.

Li, X., Shu, W., Li, M., Huang, H., Luo, P., Wu, M. (2009a). *Performance Evaluation of Vehicle-Based Mobile Sensor Networks for Traffic Monitoring.* IEEE Transactions on Vehicular Technology, Vol.58, No.4.

Li, X., Ouyang, Y. (2011). *Reliable Sensor Deployment for Network Traffic Surveillance.* Transportation Research Part B 45 (1), 218–231.

Liu, C., Meng, X., Fan, Y. (2008). *Determination of Routing Velocity with GPS Floating Car Data and WebGIS-Based Instantaneous Traffic Information Dissemination*. The Journal of Navigation, Vol. 61, pp. 337-353.

Liu, K., Yamamoto, T., Morikawa, T. (2009). *Feasibility of Using Taxi Dispatch System as Probes for Collecting Traffic Information.* Journal of Intelligent Transportation Systems, Vol.13, pp. 16-27, 2009.

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

Miwa, T., Kiuchi, D., Yamamoto, T., Morikawa, T. (2012). *Development of Map Matching Algorithm for Low Frequency Probe Data*. Transportation Research Part C, Vol. 22, pp. 132-145.

Morgul, E. F., K. Ozbay, S. Iyer, and J. Holguín-Veras (2013). *Commercial Vehicle Travel Time Estimation in Urban Networks using GPS Data from Multiple Sources.* Presented at 92nd Annual Meeting of the Transportation Research Board, Washington D.C.

Munehiro, K., Takahashi, N., Watanabe, M., Asano, M. (2012). *Winter Road Traffic Evaluation Using Taxi Probe Data in Sapporo City of Japan.* Presented at 91st Annual Meeting of the Transportation Research Board, Washington D.C..

Ohba, Y., Ueno, H., Kuwahara, M. (1999). *Travel Time Calculation Method for Expressway Using Toll Collection System Data.* Proceedings of the 2nd IEEE Intelligent Transportation Systems Conference, 471–475.

Pushkar, A., Hall, F., Acha-Daza, J. (1994). *Estimation of speeds from single-loop freeway flow and occupancy data using cusp catastrophe theory model.* Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 1457, pp. 149–157.

Rahmani, M., and H. N. Koutsopoulos (2013). *Path Inference from Sparse Floating Car Data for Urban Networks*. Transportation Research Part C, Vol. 30, pp. 41–54.

Sananmongkhonchai, S., Tangamchit, P., Pongpaibool, P. (2008). *Road Traffic Estimation from Multiple GPS Data Using Incremental Weighted Update*. 8th International Conference on ITS Telecommunications, pp. 62-66.

Soriguera, F., Rosas, D., Robusté, F. (2010). *Travel Time Measurement in Closed Toll Highways.* Transportation Research Part B 44 (10), 1242–1267.

Srinivasan, K. K., and P. P. Jovanis (1996). *Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network*. Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., No. 1537, pp. 15-22.

Sun, C., Ritchie, S.G., Tsai, K., Jayakrishnan, R. (1999). *Use of Vehicle Signature Analysis and Lexicographic Optimization for Vehicle Reidentification on Freeways*. Transportation Research Part C 7 (4), 167–185.

Turner, S.M., Eisele, W.L., Benz, R.J., Holdener, D.J. (1998). *Travel Time Data Collection Hand Book.* Research Report FHWA-PL-98-035. Federal Highway Administration, Office of Highway Information Management, Washington, DC.

Uno, N., Kurauchi, F., Tamura, H., Iida, Y. (2009). *Using Bus Probe Data for Analysis of Travel Time Variability.* Journal of Intelligent Transportation Systems, Vol. 13, pp. 2-15.

Westgate, B. S., D. B. Woodard, D. S. Matteson, and S. G. Henderson (2013). *Travel Time Estimation for Ambulances Using Bayesian Data Augmentation*. Annals of Applied Statistics, Vol. 7, No. 2, pp. 1139-1161.

Ygnace, J.-L., Drane, C., Yim, Y., de Lacvivier, R. (2000). *Travel Time Estimation on the San Francisco Bay Area Network using Cellular Phones as Probes*. California PATH Working Paper UCB-ITS-PWP-2000-18, Institute of Transportation Studies, University of California, Berkeley, CA.

Yokota, T., and D. Tamagawa (2012). *Route Identification of Freight Vehicle's Tour Using GPS Probe Data and its Application to Evaluation of on and off Ramp Usage of Expressways.* Procedia-Social and Behavioral Sciences, Vol. 39, pp. 255-266.

Yuan, J., Zheng, Y., Zhang, C., Xie,W., Xie, X., Sun, G., Huang, Y. (2010). *T-Drive: Driving Directions Based on Taxi Trajectories.* Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographinc Information Systems, pp. 99-108.

Zhang, X., and J. A. Rice (2003). *Short-Term Travel Time Prediction*. Transportation Research Part C*,* Vol. 11, No. 3, pp. 187-210.

Zheng, F., and H. Van Zuylen (2013). *Urban Link Travel Time Estimation Based on Sparse Probe Data.* Transportation Research Part C, Vol. 31, pp. 145-157.

Zou, L., Xu, J., Zhu, L. (2005). *Arterial Speed Studies with Taxi Equipped with Global Positioning Receivers as Probe Vehicles*. Wireless Communications, Networking and Mobile Computing, Vol. 2, pp. 1343-1347.

Appendix 1: Supplemental Data Analysis Results on San Antonio Corridor Network

This section provides the supplemental data analysis results based on the truck GPS data across the San Antonio corridor network. The research team summarizes the daily temporal speed distributions, the frequency of speed results, and also the percentile distributions along various highway sections.

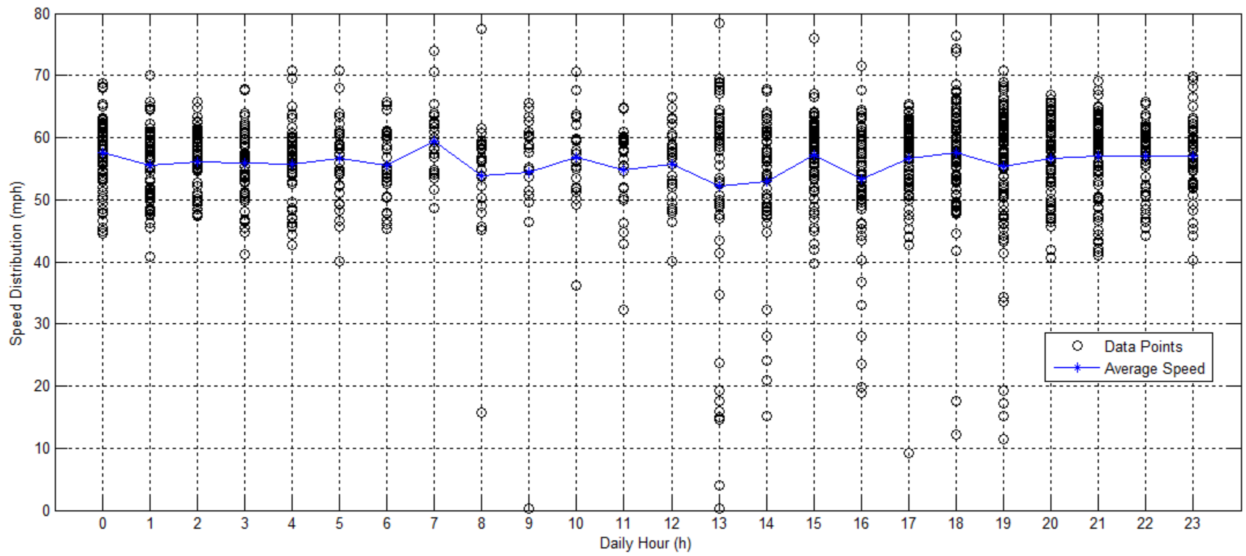


Figure 14: Temporal Speed Distribution along I-35 Section

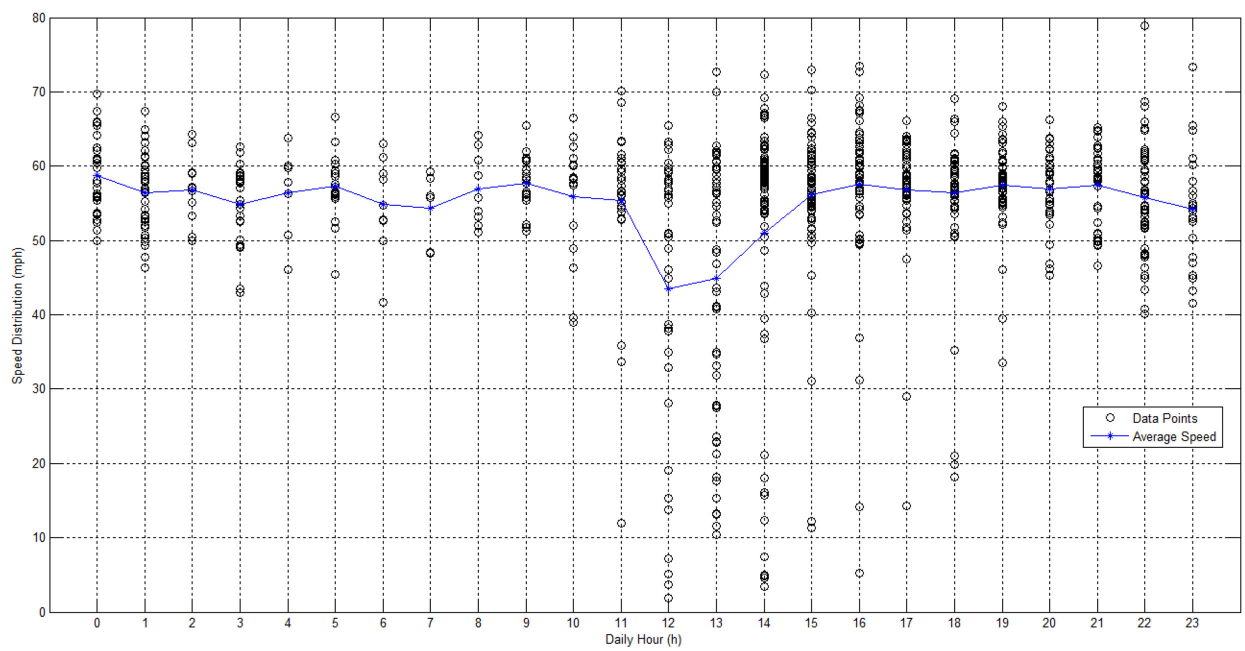


Figure 15: Temporal Speed Distribution along I-410 Section 1

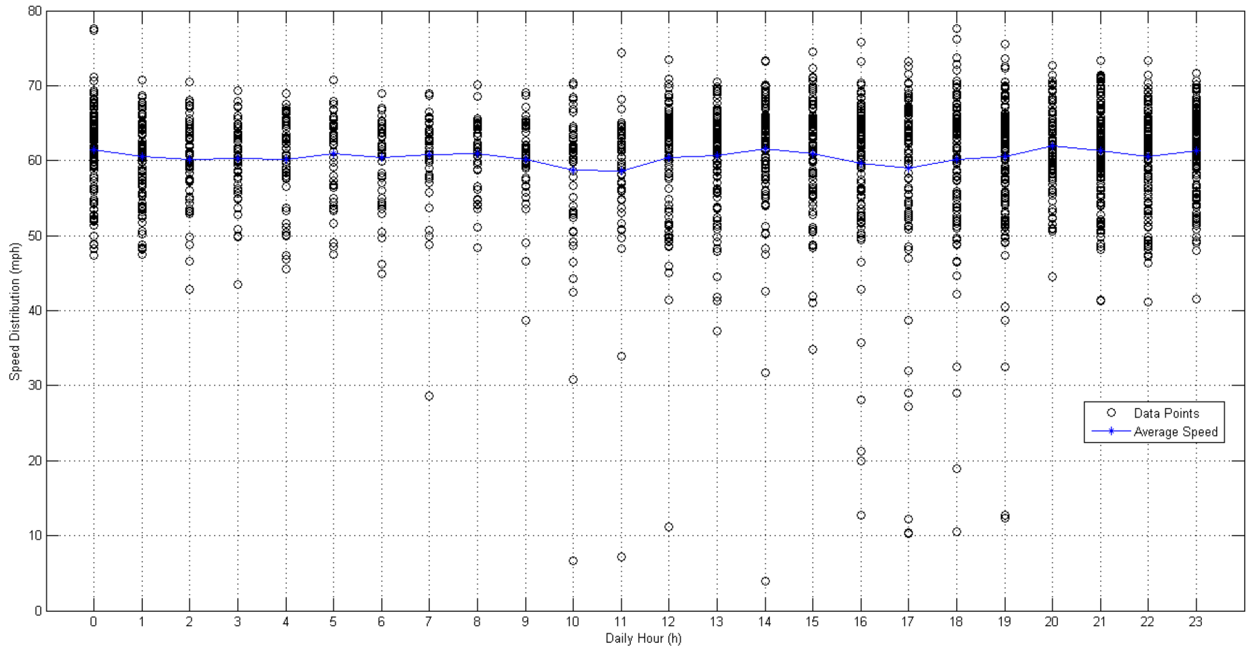


Figure 16: Temporal Speed Distribution along I-410 Section 2

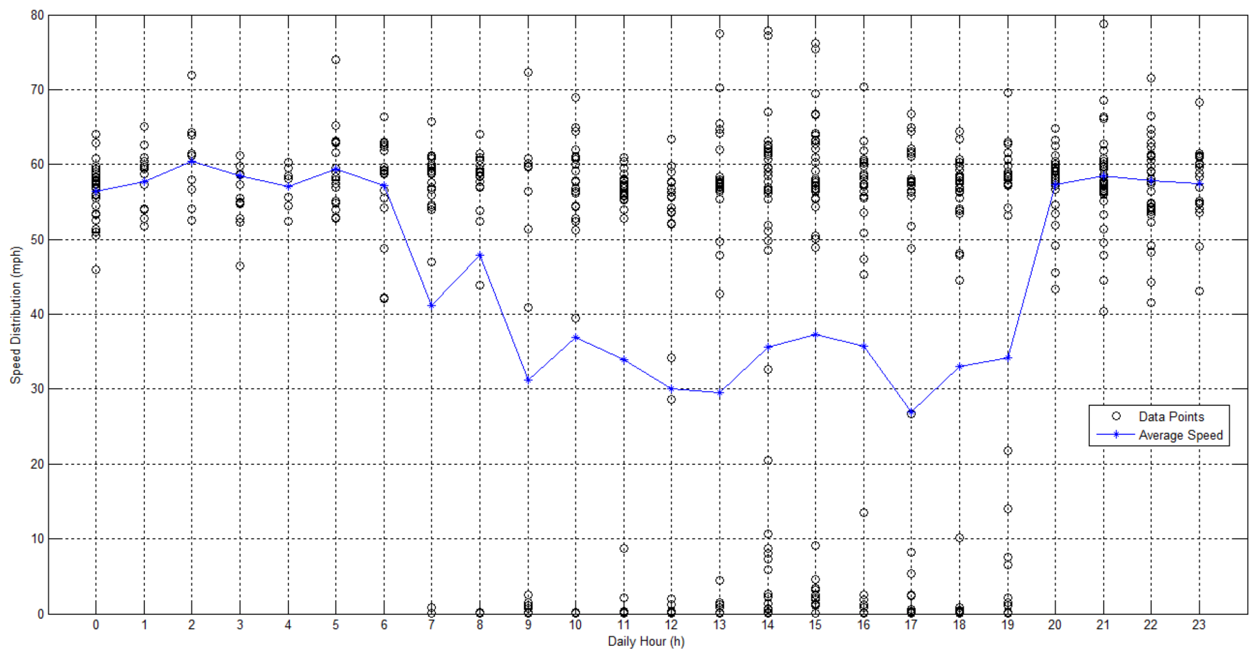
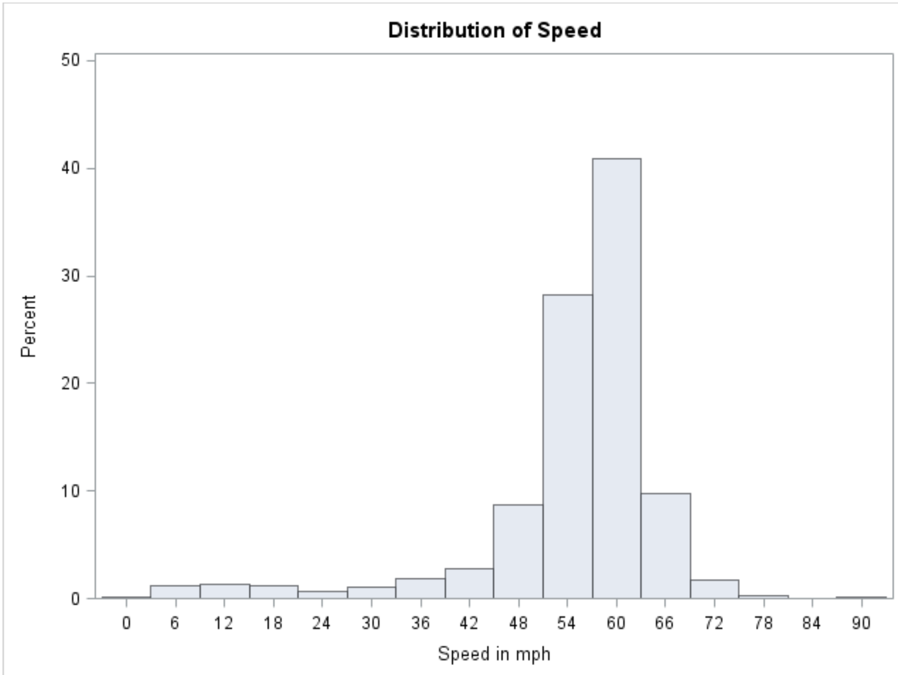


Figure 17: Temporal Speed Distribution along I-10 Section

We also analyze the variance and other statistical features of truck speed along major highway corridors across the San Antonio metro area, in order to illustrate the freight corridor travel time reliability as well as to assist the travel decisions for trucks.



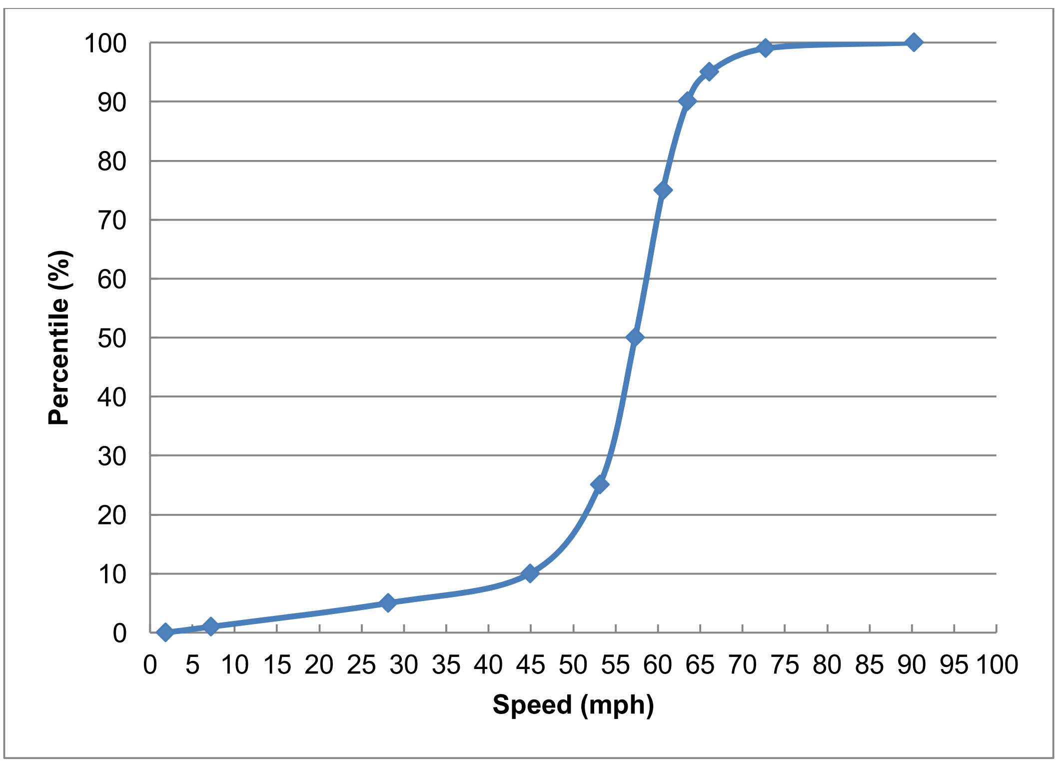
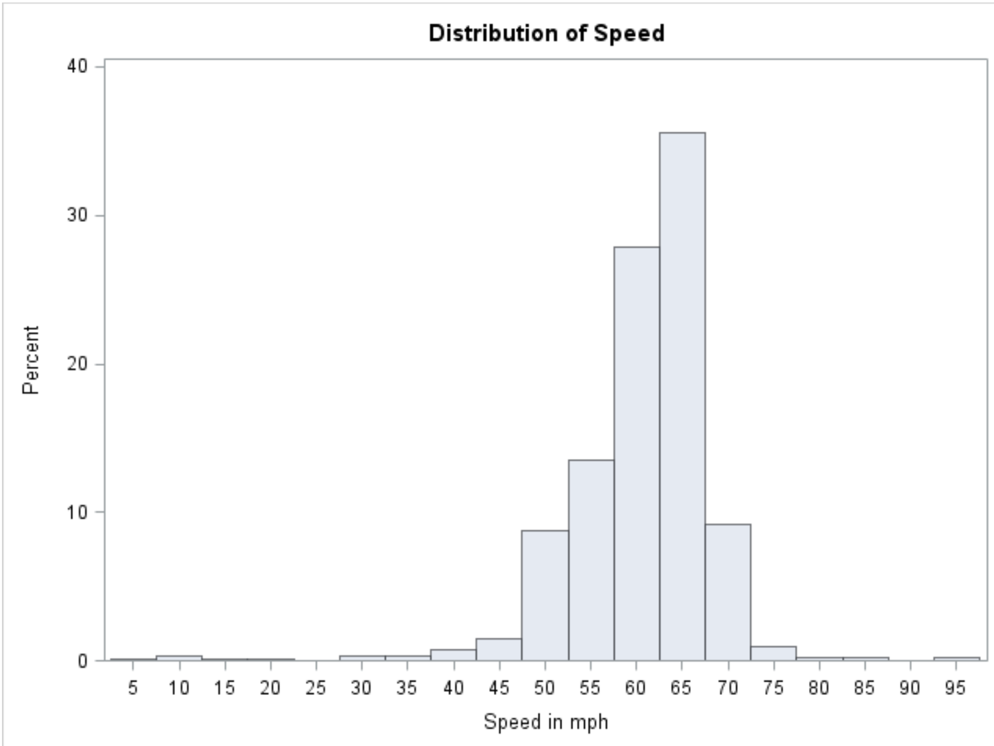


Figure 18: Frequencies and Percentiles Distribution along I-410 Section 1



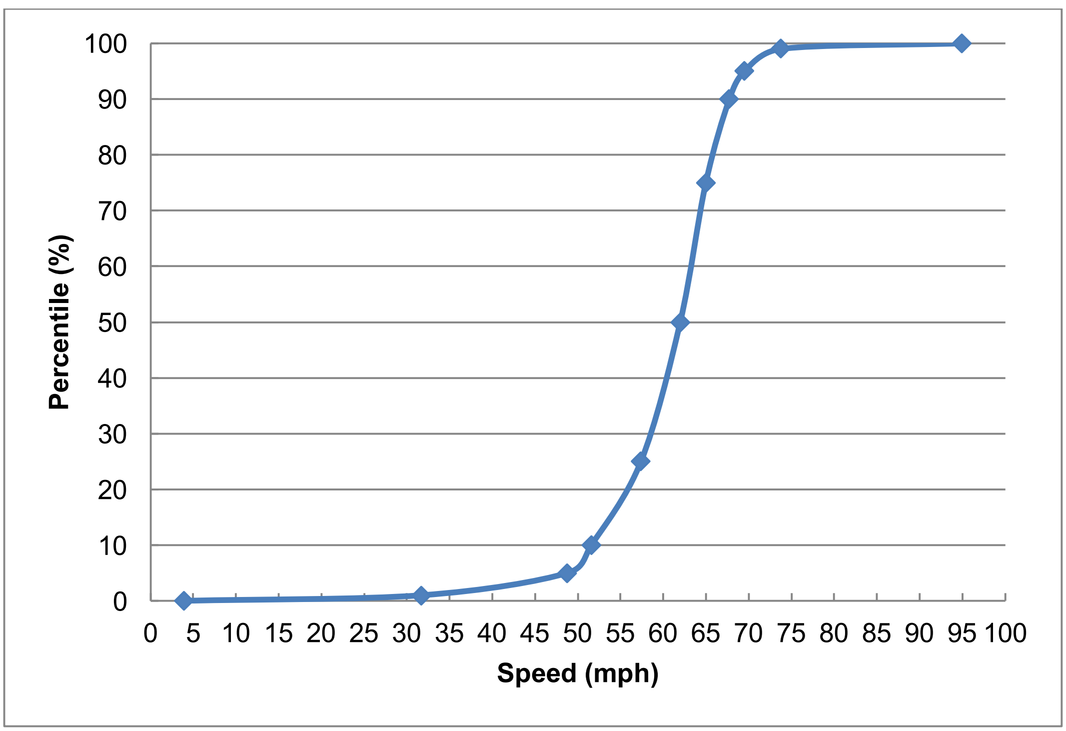
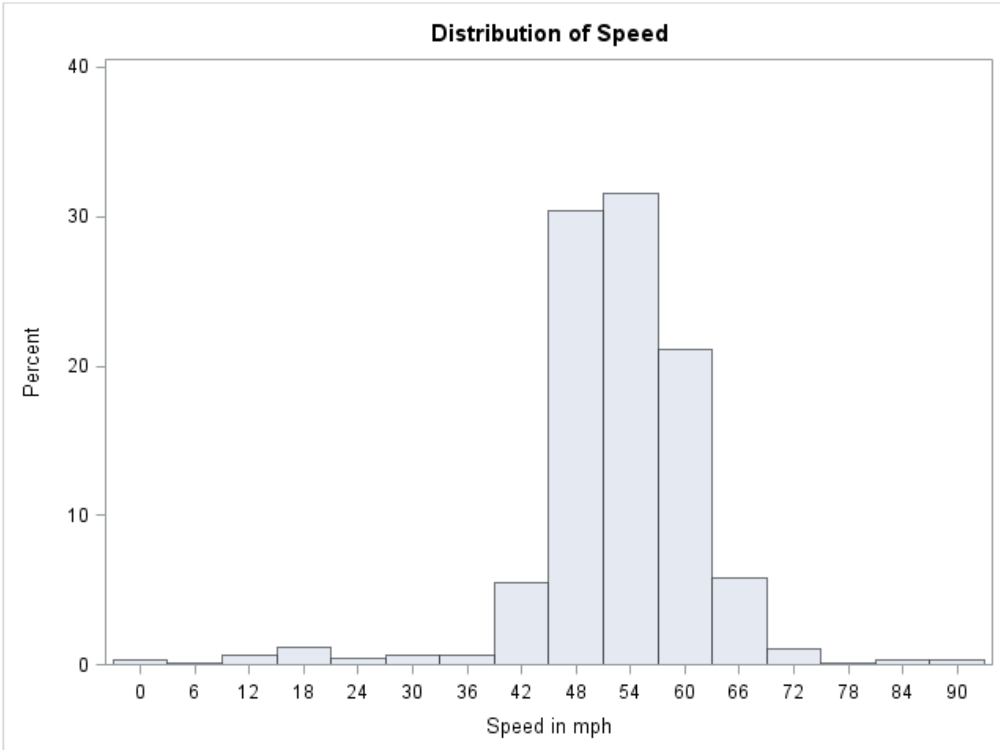


Figure 19: Frequencies and Percentiles Distribution along I-410 Section 2



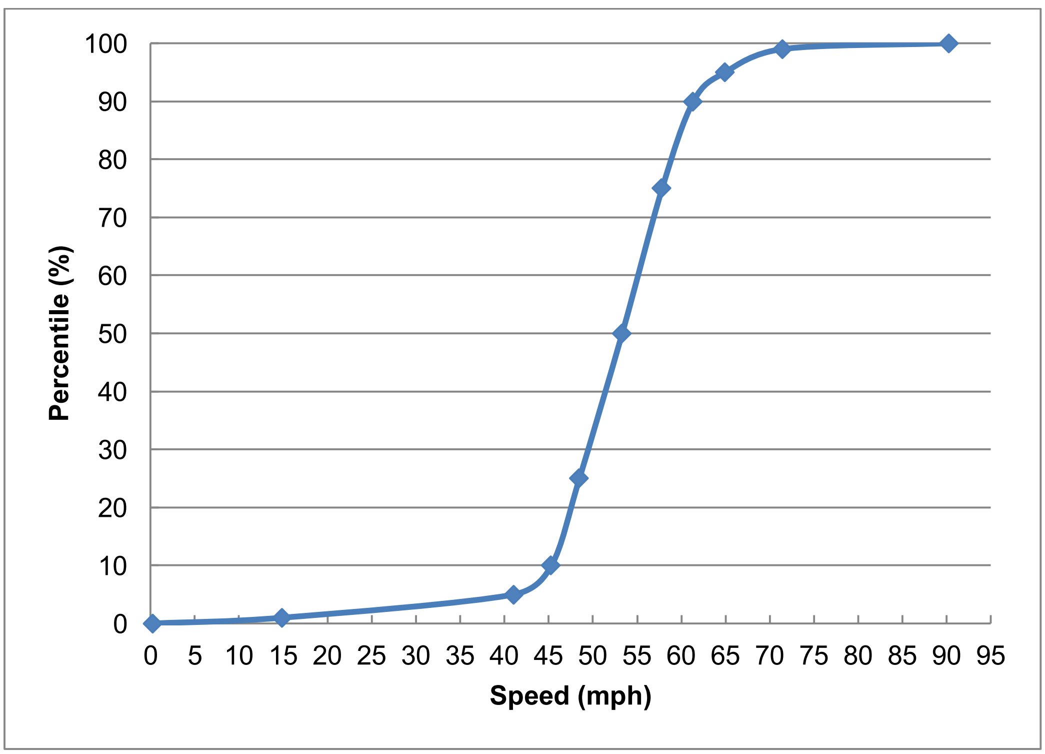


Figure 20: Frequencies and Percentiles Distribution along I-35 Section

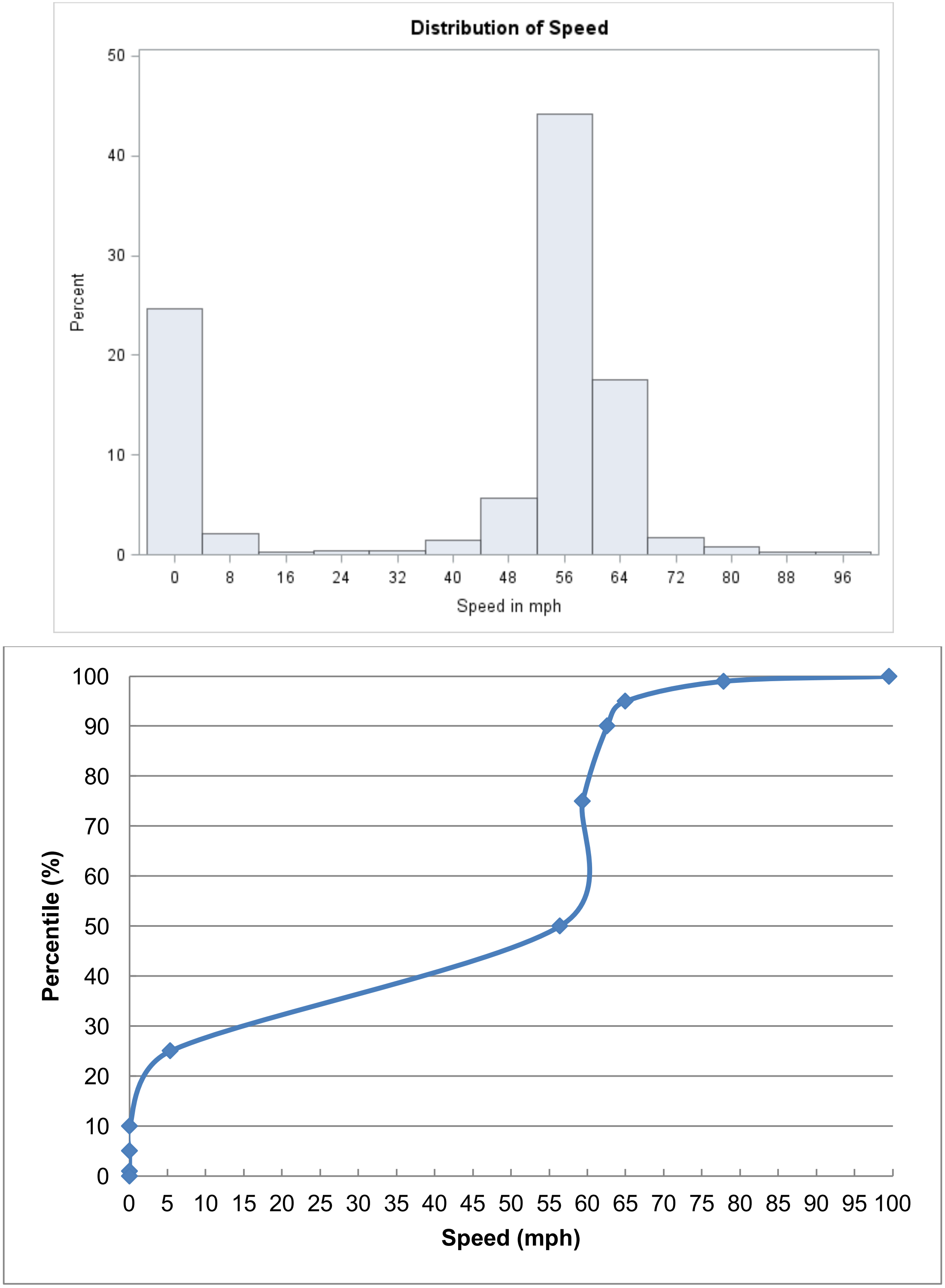
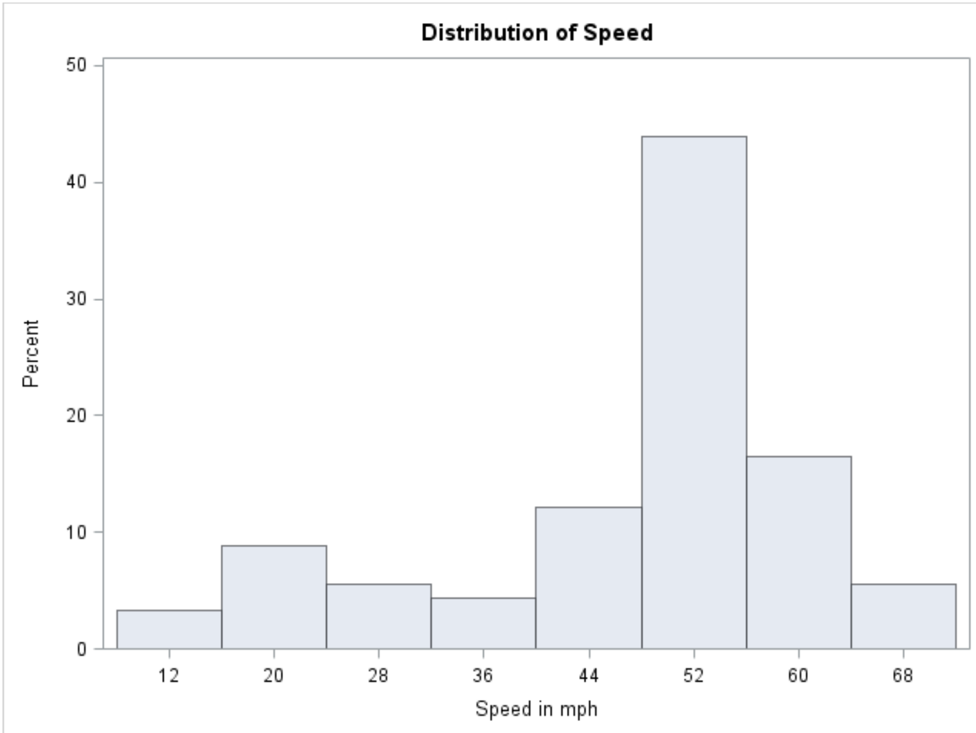


Figure 21: Frequencies and Percentiles Distribution along I-10 Section



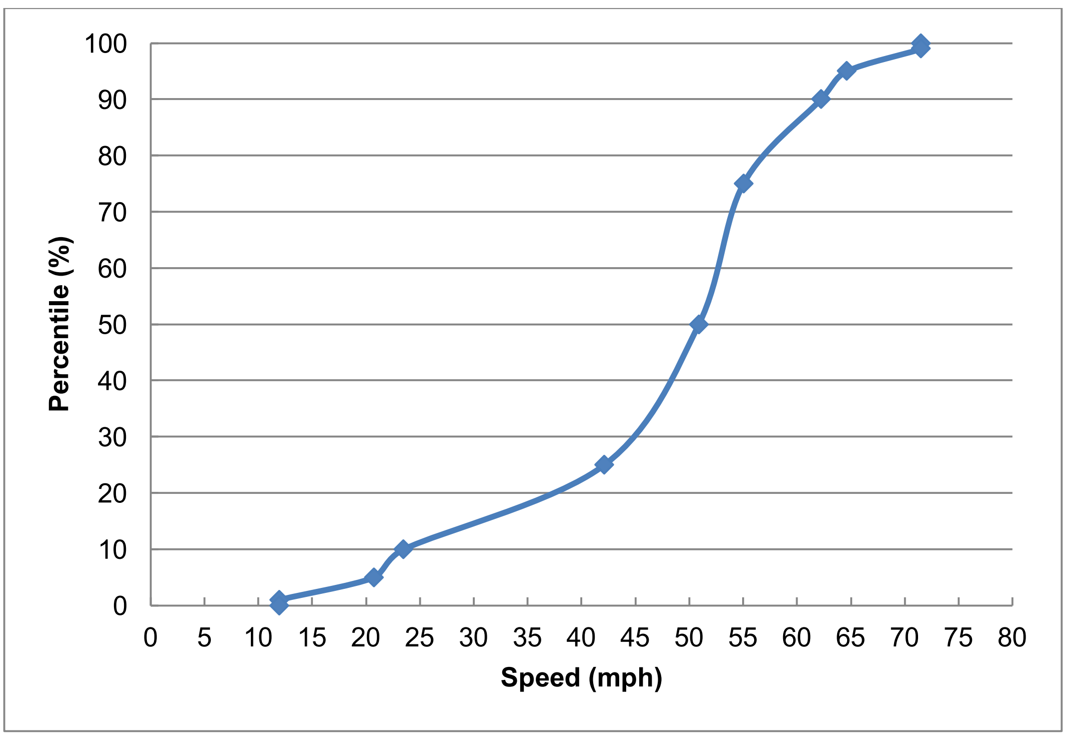


Figure 22: Frequencies and Percentiles Distribution along I-37 Section

Appendix 2: Data Analysis on Hudson-Beloit Corridor in WI

The objective of this section is to process and analyze the truck GPS data along the corridor from Hudson to Beloit, Wisconsin. This American Transportation Research Institute (ATRI) data is also collected in March 2009. The majority of this corridor carries over 8,500 trucks per day. Similar to the processing of Texas truck GPS data, we analyze the statistical summary of collected data, and detect the outlier points based on travel speed results. The research team demonstrates that it is necessary to apply the outlier detection to filter the truck GPS data for further use of travel time estimation.

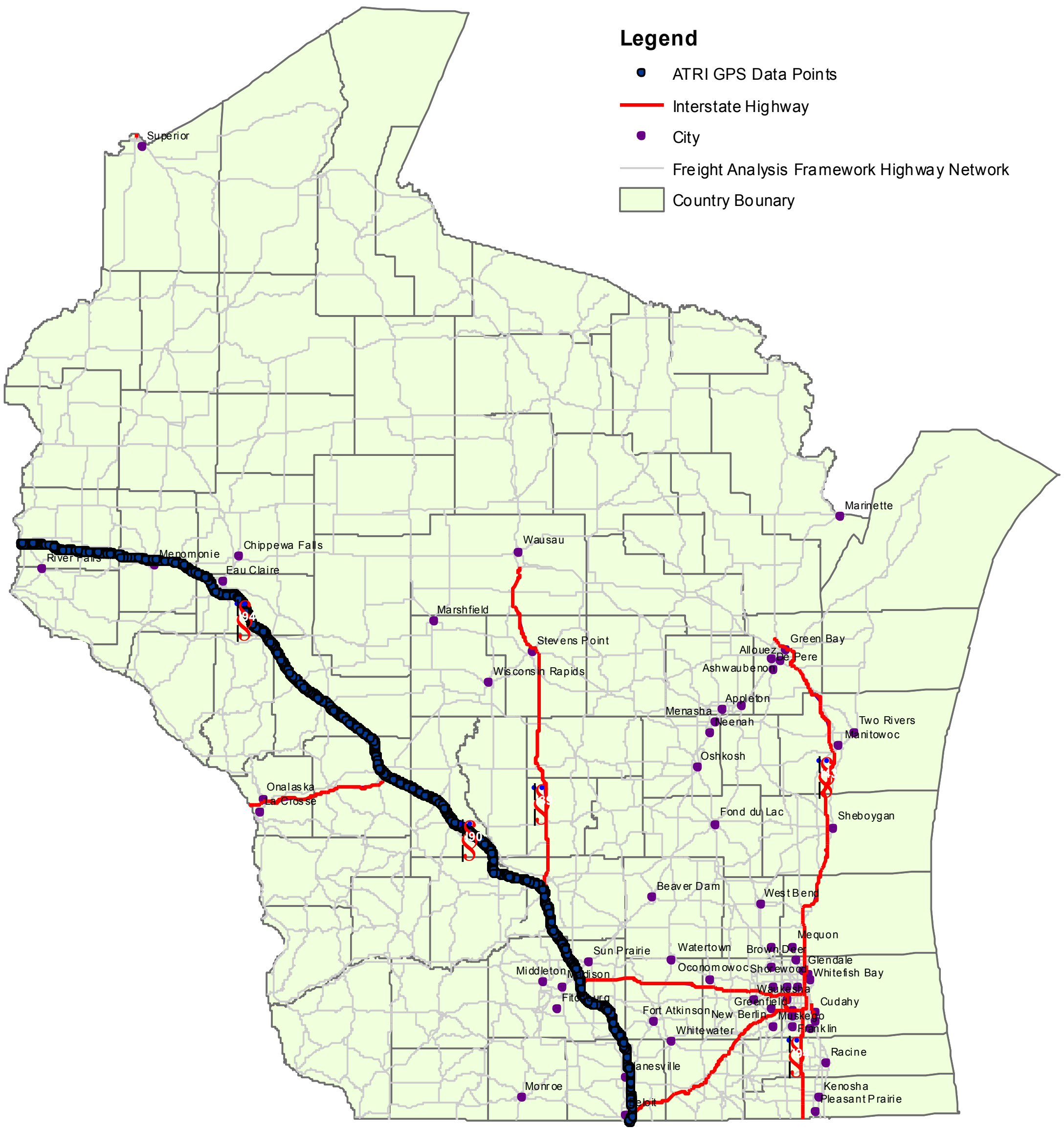


Figure 23: Scope of Corridor from Hudson to Beloit, WI

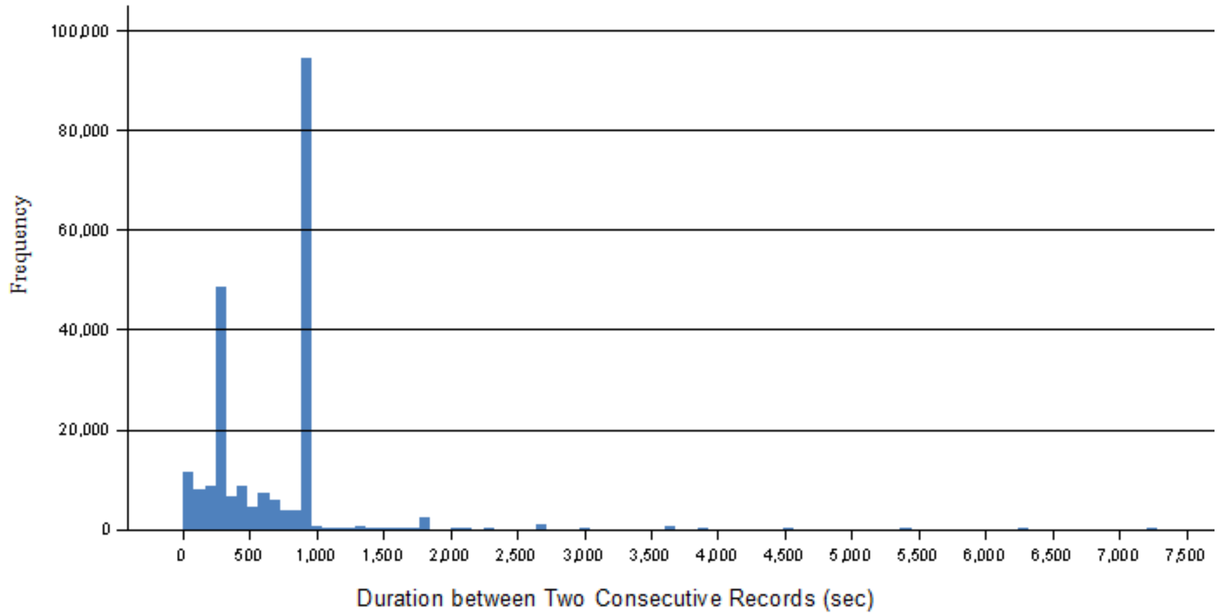


Figure 24: Distribution of Time Interval between Consecutive Records

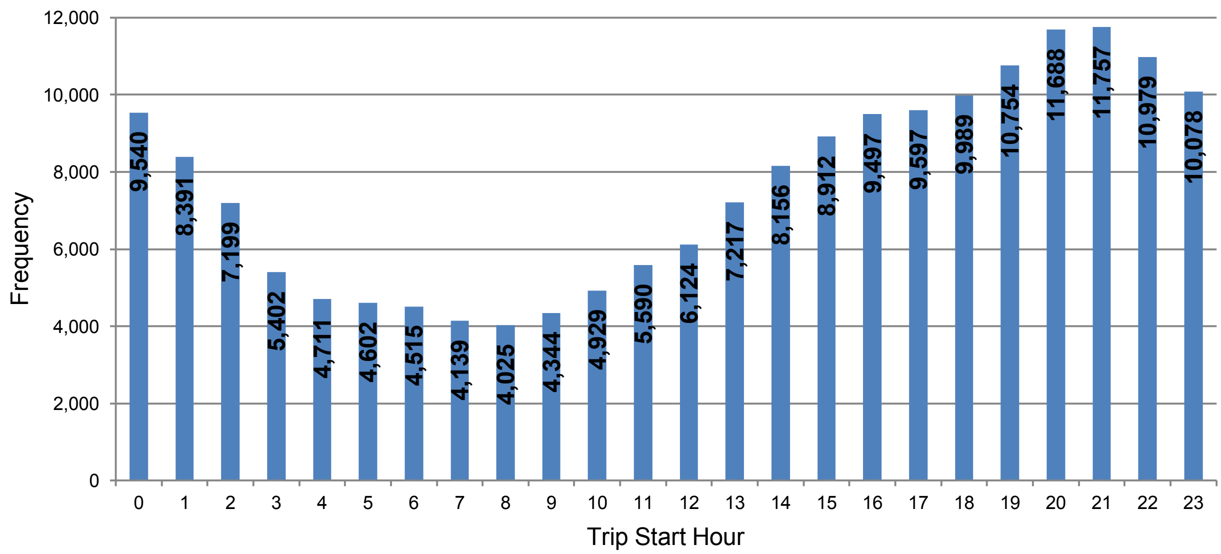


Figure 25: Distribution of Truck Trips by Hour

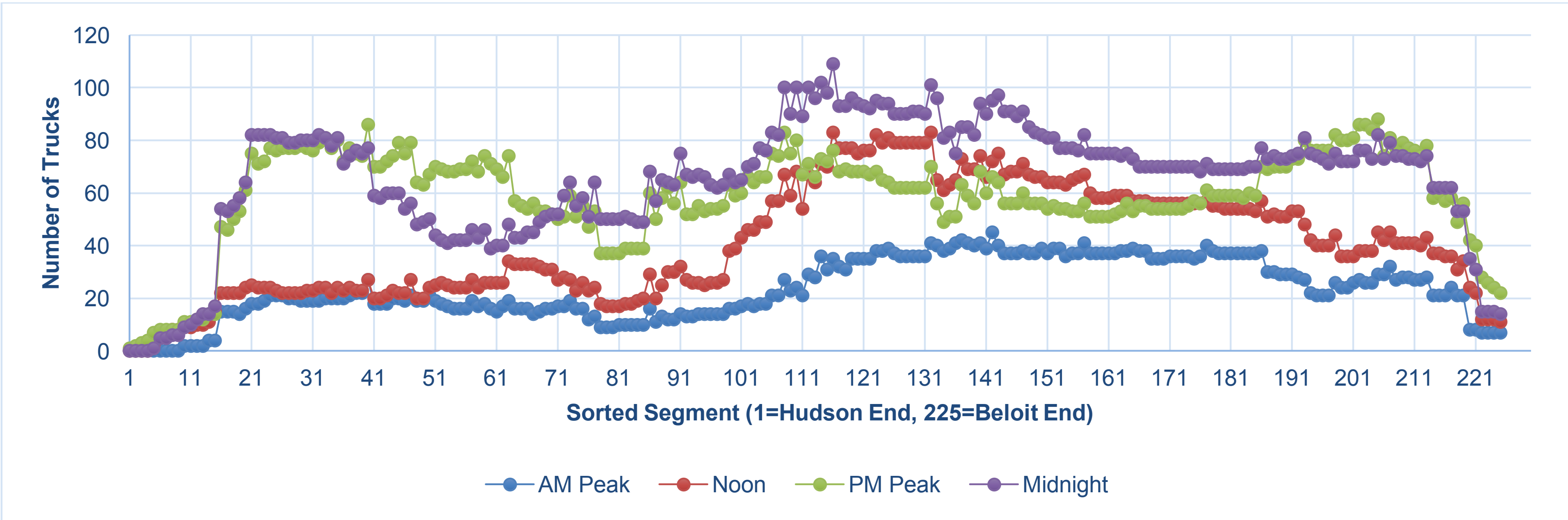


Figure 26: Numbers of Trucks Traversing Each Segment

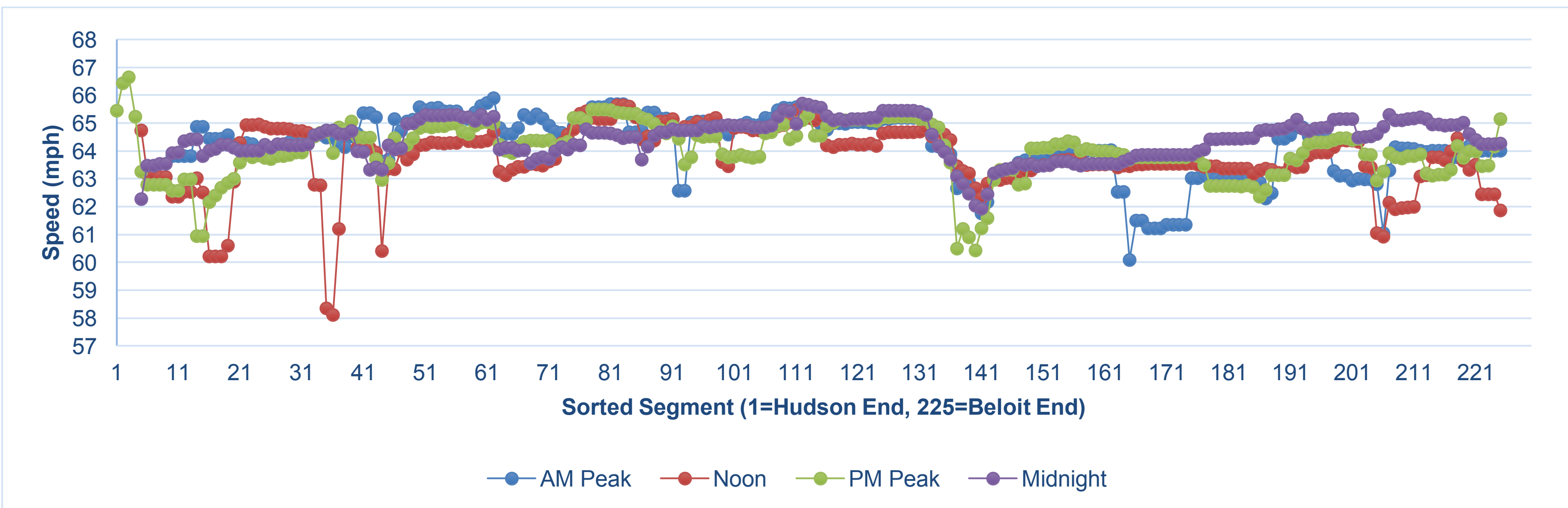


Figure 27: Truck Mean Speed on Each Segment

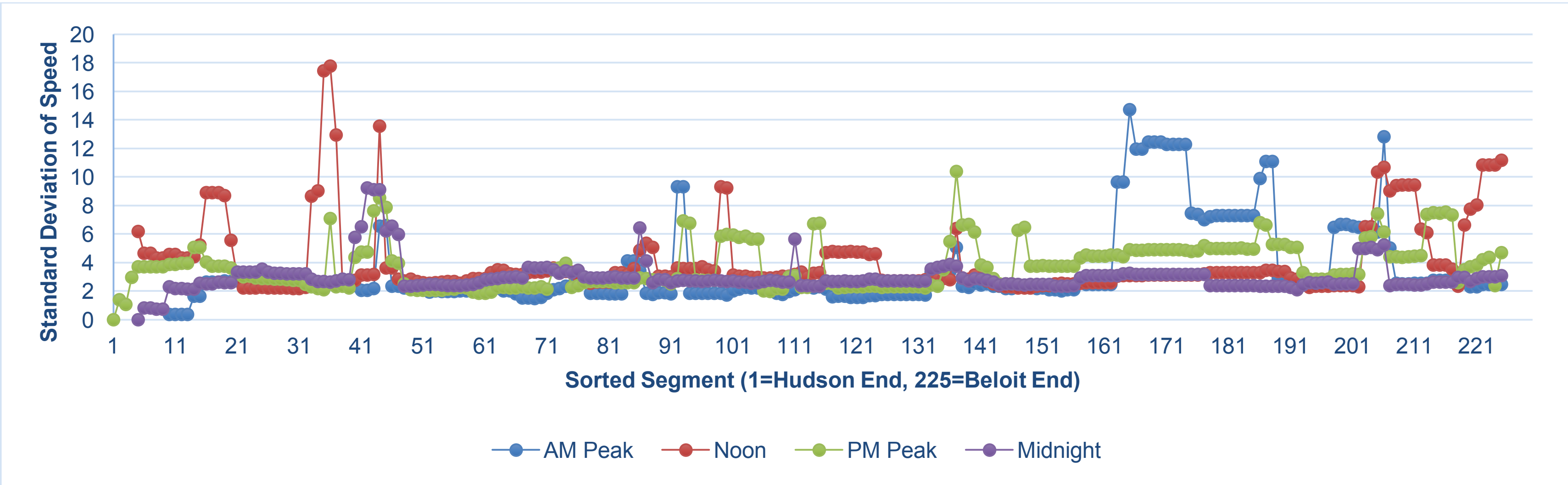


Figure 28: Standard Deviation of Speed on Each Segment

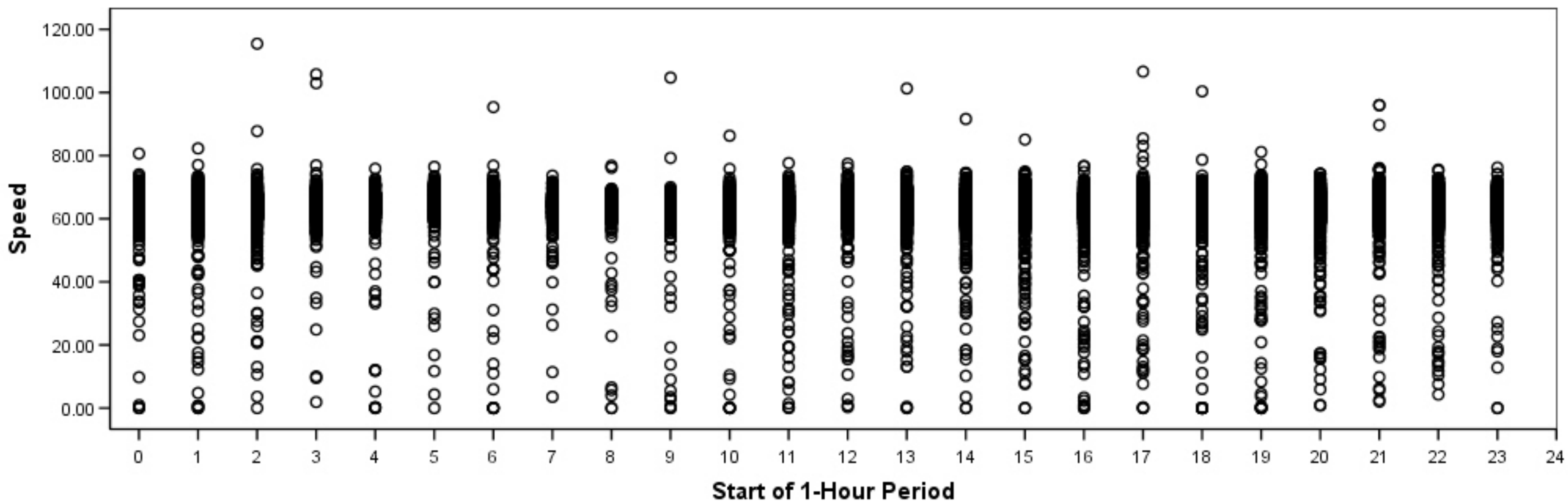


Figure 29: Distribution of Speed in Each Hour

Considering a truck trip may experience non-traffic related stops, such as fueling on a trip, the estimation of travel speed will be biased without identifying and removing such trips. After the outlier detection by comparing the travel speed of a trip with other trips on the same segment and within same period, we summarize the detected outlier distribution in Tables 7 and 8. The according travel speed distribution over the corridor after removing outliers are also displayed in Figures 30-33.

Table 7: Distribution of Outliers by Hour

Hour	Number of Trips	Number of Outliers	Percentage
0	4815	156	3.24%
1	4018	129	3.21%
2	3429	122	3.56%
3	2574	97	3.77%
4	2180	91	4.17%
5	2114	88	4.16%
6	2059	105	5.10%
7	1805	76	4.21%
8	1661	80	4.82%
9	1865	91	4.88%
10	2165	111	5.13%
11	2393	126	5.27%
12	2741	112	4.09%
13	3226	134	4.15%
14	3722	160	4.30%
15	4233	187	4.42%
16	4844	187	3.86%
17	4927	173	3.51%
18	5372	190	3.54%
19	5916	203	3.43%

20	6445	170	2.64%
21	6391	185	2.89%
22	5737	149	2.60%
23	5067	170	3.36%

Table 8: Distribution of Outliers by Day

Day	Number of Trips	Number of Outliers	Percentage
1	2501	116	4.64%
2	5767	214	3.71%
3	6057	228	3.76%
4	6056	182	3.01%
5	5391	198	3.67%
6	4976	176	3.54%
7	1539	63	4.09%
8	10	3	30.00%
9	20	1	5.00%
10	15	2	13.33%
11	14	0	0.00%
12	15	0	0.00%
13	20	0	0.00%
14	9	0	0.00%
15	29	0	0.00%
16	36	1	2.78%
17	32	3	9.38%
18	1726	43	2.49%
19	4710	172	3.65%

20	4488	177	3.94%
21	2891	107	3.70%
22	2145	101	4.71%
23	5178	187	3.61%
24	5665	210	3.71%
25	5078	181	3.56%
26	5345	171	3.20%
27	4564	184	4.03%
28	2969	111	3.74%
29	2182	97	4.45%
30	4784	185	3.87%
31	5487	179	3.26%

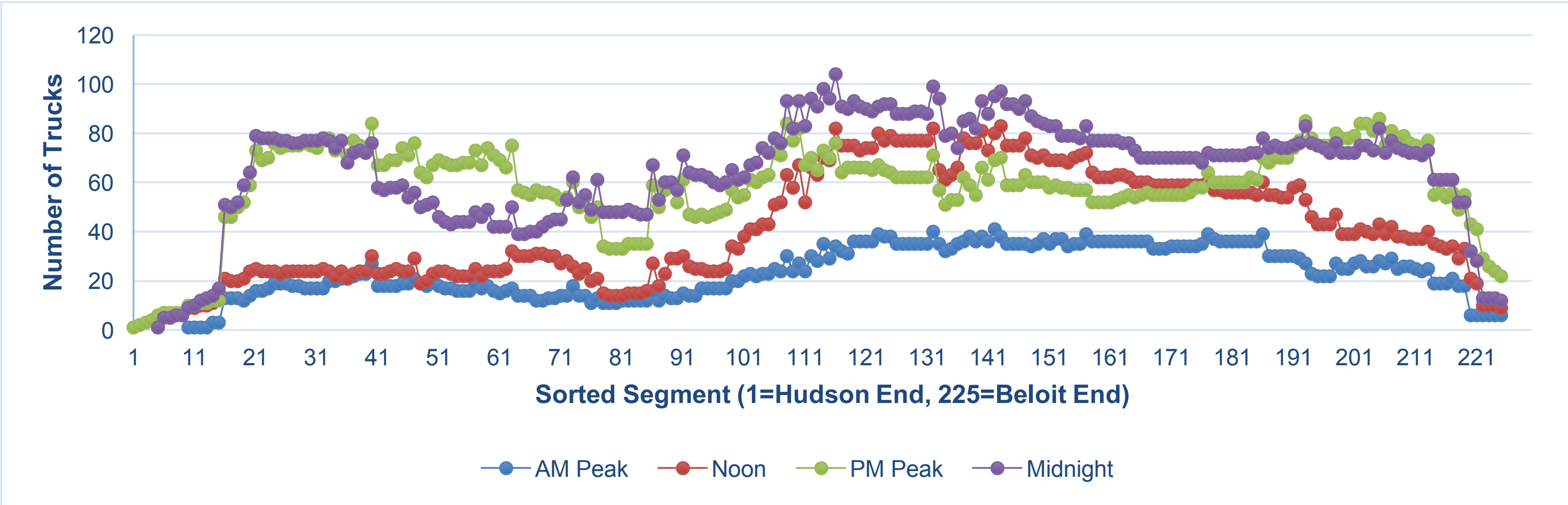


Figure 30: Numbers of Trucks Traversing Each Segment after Removing Outliers

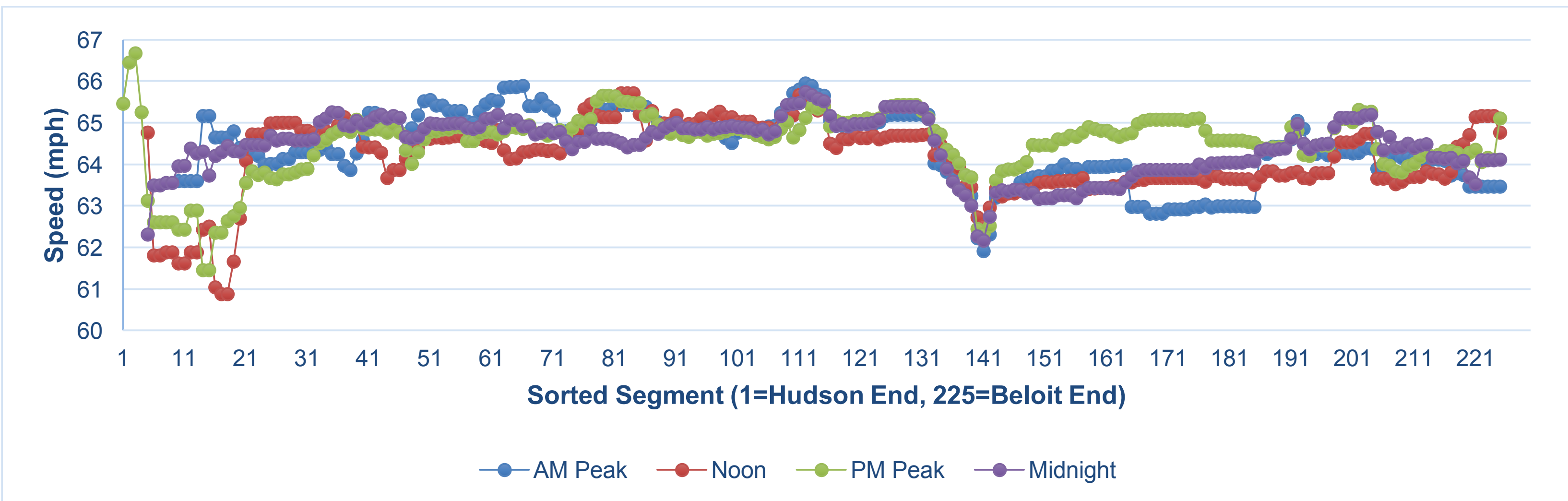


Figure 31: Truck Mean Speed on Each Segment after Removing Outliers

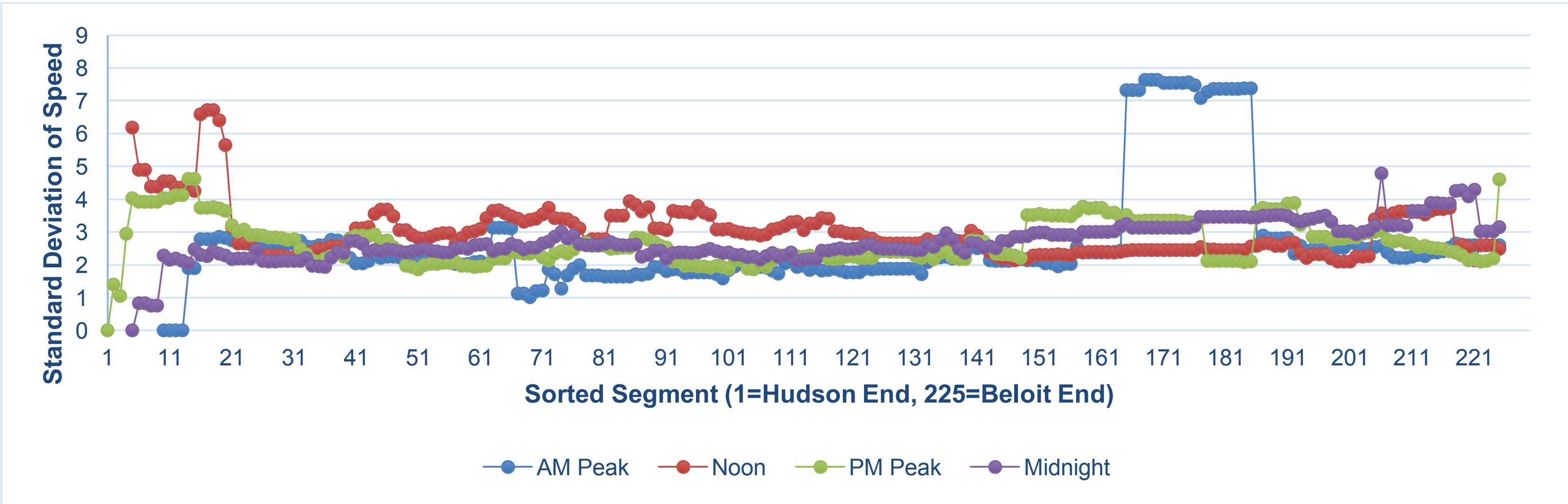


Figure 32: Standard Deviation of Speed by Segment after Removing Outliers

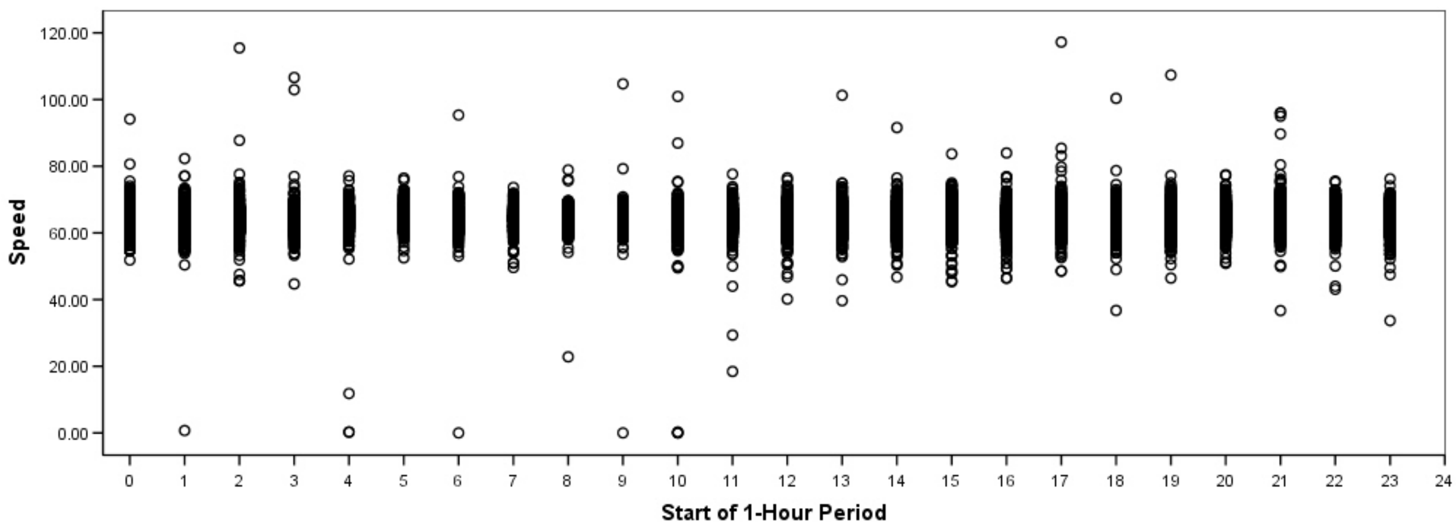


Figure 33: Distribution of Speed in Each Hour after Removing Outliers