Final Research Report
Agreement T4118, Task 31
Truck Performance
TNW 2010-02
TransNow Budget No. 61-6327

DEVELOPING A GPS-BASED TRUCK FREIGHT PERFORMANCE MEASURE PLATFORM

by

Edward D. McCormack
Research Assistant Professor,

Xiaolei Ma
Research Assistant

Charles Klocow
Research Assistant

Anthony Curreri
IT Specialist

Duane Wright
Systems Analyst Programmer

Department of Civil and Environmental Engineering
University of Washington, Box 352700
Seattle, Washington 98195-2700

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
1107 NE 45th Street
Seattle, Washington 98105-4631

Washington State Department of Transportation
Technical Monitor
Barbara Ivanov, Freight Systems Division Co-Director

Sponsored by

Transportation Northwest (TransNow)
University of Washington
129 More Hall, Box 352700
Seattle, Washington 98195-2700

The State of Washington
Department of Transportation
Paula J. Hammond, Secretary

and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

April 2010

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 748.1 TNW2010-02		

4. TITLE AND SUBTITLE	5. REPORT DATE
DEVELOPING A GPS-BASED TRUCK FREIGHT PERFORMANCE MEASURES PLATFORM	April 2010
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Edward D. McCormack, Xiaolei Ma, Charles Klocow, Anthony Currarei, Duane Wright	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Transportation Northwest Regional Center X (TransNow) Box 352700, 129 More Hall University of Washington Seattle, WA 98195-2700 Washington State Transportation Center (TRAC) University of Washington 1107 NE 45th St, Suite 535 Seattle, WA 98105	11. CONTCT OR GRANT NO. Agreement T4118, Task 31
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Washington State Dept of Transp. Transportation Building, MS 47372 Olympia, Washington 98504-7372 Doug Brodin, 360-705-7972 U.S. Department of Transportation Office of the Secretary of Transp. 400 7th St. SW Washington, DC 20590	Final Research Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the University of Washington and the US Department of Transportation.

16. ABSTRACT

Although trucks move the largest volume and value of goods in urban areas, relatively little is known about their travel patterns and how the roadway network performs for trucks. The Washington State Department of Transportation (WSDOT), Transportation Northwest (TransNow) at the University of Washington, and the Washington Trucking Associations have partnered on a research effort to collect and analyze global positioning system (GPS) truck data from commercial, in-vehicle, truck fleet management systems used in the central Puget Sound region. The research project is collecting commercially available GPS data and evaluating their feasibility to support a state truck freight network performance monitoring program. WSDOT is interested in using this program to monitor truck travel times and system reliability, and to guide freight investment decisions.

The researchers reviewed truck freight performance measures that could be extracted from the data and that focused on travel times and speeds, which, analyzed over time, determine a roadway system's reliability. The utility of spot speeds and the GPS data in general was evaluated in a case study of a three-week construction project on the Interstate-90 bridge. The researchers also explored methods for capturing regional truck travel performance.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Truck data, GPS data, freight performance measures, travel reliability	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None		

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. This document is disseminated through the Transportation Northwest (TransNow) Regional Center under the sponsorship of the U.S. Department of Transportation UTC Grant Program and through the Washington State Department of Transportation. The U.S. government assumes no liability for the contents or use thereof. Sponsorship for the local match portion of this research project was provided by the Washington State Department of Transportation. The contents do not necessarily reflect the official views or policies of the U.S. Department of Transportation or Washington State Department of Transportation. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

OVERVIEW

Although trucks move the largest volume and value of goods in urban areas, relatively little is known about their travel patterns and how the roadway network performs for trucks. Global positioning systems (GPS) used by trucking companies to manage their equipment and staff and meet shippers' needs capture truck data that are now available to the public sector for analysis. The Washington State Department of Transportation (WSDOT), Transportation Northwest (TransNow) at the University of Washington (UW), and the Washington Trucking Associations (WTA) have partnered on a research effort to collect and analyze GPS truck data from commercial, in-vehicle, truck fleet management systems used in the central Puget Sound region. The research project is collecting commercially available GPS data and evaluating their feasibility to support a state truck freight network performance monitoring program. WSDOT is interested in using this program to monitor truck travel times and system reliability, and to guide freight investment decisions.

The success of the truck freight performance measurement program will depend on developing the capability to

- efficiently collect and process GPS devices' output
- extract useful truck travel time and speed, roadway location, and stop location information and
- protect the identity of the truckers and their travel information so that business sensitive information is not released.

While earlier studies have evaluated commercial vehicles' travel characteristics by using GPS devices, these researchers did not have access to commercial fleet data and had to estimate corridor travel speeds from a limited number of portable GPS units capable of making frequent (1- to 60-second) location reads (Quiroga and Bullock 1998, Greaves and Figliozzi 2008, Due and Aultman-Hall 2007). This read frequency permitted a fine-grained analysis of truck movements on specific segments of the road network but did not provide enough data points to reliably track regional or corridor network performance.

This research project is taking a different approach. The data analyzed in this project are drawn from GPS devices installed to meet the trucking sector's fleet management needs. So the truck locations are collected less frequently (typically every 5 to 15 minutes) but are gathered from a much larger number of trucks over a long period of time. The researchers are collecting data from 2,000 to 3,000 trucks per day for one year in the central Puget Sound region.

This report discusses the steps taken to build, clean, and test the data collection and analytic foundation from which the UW and WSDOT will extract network-based truck performance statistics. One of the most important steps of the project has been to obtain fleet management GPS data from the trucking industry. Trucking companies approached by WSDOT and the UW at the beginning of the study readily agreed to share their GPS data, but a lack of technical support from the firms made data collection difficult. The researchers overcame that obstacle by successfully negotiating contracts with GPS and telecom vendors to obtain GPS truck reads in the study region. The next challenge was to gather and format the large quantities of data (millions of points) from different vendors' systems so that they could be manipulated and evaluated by the project team. Handling the large quantity of data meant that data processing steps had to be automated, which required the development and validation of rule-based logic that could be used to develop algorithms.

Because a truck performance measures program will ultimately monitor travel generated by trucks as they respond to shippers' business needs, picking up goods at origins (O) and dropping them off at destinations (D), the team developed algorithms to extract individual truck's O/D information from the GPS data. The researchers mapped (geocoded) each truck's location (as expressed by a GPS latitude and longitude) to its actual location on the Puget Sound region's roadway network and to traffic analysis zones (TAZs) used for transportation modeling and planning.

The researchers reviewed truck freight performance measures that could be extracted from the data and that focused on travel times and speeds, which, analyzed over time, determine a roadway system's reliability. Because the fleet management GPS data from individual trucks typically consist of infrequent location reads, making any one truck an unreliable probe vehicle, the researchers explored whether data from a larger quantity of trucks could compensate for infrequent location reads. To do this, the project had to evaluate whether the spot (instantaneous) speeds recorded by one truck's GPS device could be used in combination with spot speeds from other trucks on the same portion of the roadway network.

The utility of spot speeds and the GPS data in general was evaluated in a case study of a three-week construction project on the Interstate-90 (I-90) bridge. The accuracy of the spot speeds was then validated by comparing the results with speed data from WSDOT's freeway management loop system (FLOW).

The researchers also explored methods for capturing regional truck travel performance. The approach identified zones that were important in terms of the number of truck trips that were generated. Trucks' travel performance as they traveled between these economic zones could then be monitored over time and across different times of day.

LITERATURE REVIEW

Several other recent research efforts have used GPS data to measure truck performance. An ongoing project by the Federal Highway Administration (FHWA) and the American Transportation Research Institute (ATRI) concluded that GPS data from trucks can be processed in a confidential manner to provide average travel rates along major long-distance U.S. highways (ATRI 2009, Short, Picket and Christianson 2009). ATRI purchased data from GPS service vendors, and aggregated the spot speeds from GPS devices over time to identify truck bottlenecks.

In an earlier WSDOT–UW project, McCormack and Hallenbeck found that GPS devices can be used to measure truck travel times along specific roadway corridors. That study, which placed 25 portable GPS devices on board trucks traveling the Interstate-5 (I-5) corridor, concluded that GPS data can provide an indication of roadway performance but require many more data points for statistical reliability (2006). Greaves and Figliozzi discussed implementation and analysis issues associated with a truck GPS data collection effort and described the algorithms used to process the raw data to identify trip ends. Their study, which collected data from 30 trucks, also discussed the potential uses and limitations of GPS technology in urban freight modeling and planning (2008).

Czerniak recognized that GPS data can provide a powerful set of planning and programming tools for DOT's but also identified major processing issues associated with using these data (2002). A number of researchers have documented means to process GPS data, as well as retrieve trip information to detect origin-destination patterns and to calculate travel times. Quiroga and Bullock proposed a new methodology for performing travel time studies that used GPS and geographic information system (GIS) technologies (1998a). They documented the data collection, reduction, and reporting procedures used to produce a measure of effectiveness for travel times at various levels of resolution, including the roadway system, corridor, and local road levels. Du and Aultman-Hall developed an automatic trip end identification algorithm by using a combination of maximum and minimum dwell times, heading changes, and a check for distance between the GPS points (2007). This process was used to increase the accuracy of trip rate information. Schuessler and Axhausen described a post-processing procedure for cleaning and smoothing raw GPS data and automatically identified the trip activity and trip modes by using fuzzy logic (2008). Hunter et al. used GPS-instrumented test vehicles to calculate travel times on urban arterial streets by developing an algorithm that identified the traversal time between intersections for a GPS device mounted in a probe vehicle (2008).

All of these research efforts used GPS services that produced frequent reads (1 to 60 seconds) but small sample sizes of probe vehicles to identify trip ends and calculate travel times. While these studies typically used some level of automated processing to clean and organize the data, the post-processing step also included manual processes to fix or remove data with problems.

RESEARCH APPROACH

The project developed a data foundation for a Washington State truck performance measures program. This required the project research team address a series of research steps which include:

- working with the GPS vendors and developing a contracting mechanism to acquire usable data while protecting the privacy of the trucking companies
- developing a data feed framework that worked with different GPS vendors and efficiently retrieved and stored a large stream of data
- writing software that organized and reformatted the vendor's GPS data and that indentified erroneous and bad data
- processing the data within geographic information systems (GIS) software to map match (geocode) the truck's GPS output to the region's roadway network
- writing and validating a series of programs that located the origin and destination for each truck's trip
- identifying and testing freight performances measure on case studies
- determining what sample sizes of GPS data are statistically usable
- recommending data needs for the next phase of a statewide GPS based freight performance measures program.

The completion of these steps required a multi-disciplinary research team with skills in contracting, information technology, database management, software development, GIS analysis, statistics, and graphics.

DATA ACQUISITION

At the beginning of this research, the project team contacted 20 regional and national companies that operated trucking fleets in the Puget Sound region and asked them to share data from their fleets' GPS devices. With one exception, every company that used a fleet management GPS system (several did not) agreed to give the UW and WSDOT access to their trucks' GPS data. However, actually obtaining the data from these individual companies proved to be a challenge. Even if a company was willing to spend staff resources to support the project, it was difficult to work out the technical details of transferring the data with its company's in-house data staff (if they existed).

The project team realized that obtaining data directly from GPS service providers would be easier. This approach had a number of advantages:

- The vendors had the technical staff experienced in setting up and sending out routine GPS data reports.
- Each vendor was able to provide truck GPS data from multiple companies.
- The vendors were interested in our performance measures program because it represented a way to obtain additional value from their GPS data.
- A contract was drawn up with each vendor, creating a business relationship instead of voluntary participation.

Purchasing data directly from the vendors, however, also had some disadvantages:

- The data were collected to help trucking companies manage their fleets and might not be ideal for a public sector freight performance measure program.
- There was an ongoing cost for the data, whereas the individual companies were willing to provide the data for free.
- Because the vendors were not accustomed to selling data to a university or a DOT, there were no contracting models available and the team had to negotiate each agreement separately.
- The data typically had geographic limits as specified in the contract.

- Because protecting the privacy of the vendors' customers was critical, non-disclosure agreements were required, and legal review of those agreements slowed the project.

WSDOT and the UW signed one-year data acquisition contracts with three GPS vendors. Three different types of vendors were selected in order to evaluate different data acquisition methods. The project currently receives data daily for up to 2,500 trucks traveling in the Puget Sound region. The contract with vendor A involved setting up a near real-time raw data feed for all its client trucks when they travel in the central Puget Sound region. The contract with vendor B included paying for the installation of 25 GPS devices in the trucks of volunteer trucking companies and paying for the service for one year. The UW receives raw GPS data from these devices, and the trucking companies have access to fleet management services for the year. Vendor C receives data from cell phone-based GPS devices in trucks. The project paid for service and 60 phones that were placed in trucks belonging to Washington Trucking Associations member firms. An overview of the data acquired from each vendor is shown in Table 1. GPS data were also purchased through a fourth contract with ATRI, but to protect privacy, trip starts and ends were suppressed, limiting the data's utility. At the time of this report, the ATRI data provide GPS reads in intervals of 15 to 60 minutes, which is too infrequent for urban network analysis.

The GPS devices report, using cellular technology, both at preset intervals and when the trucks stop. Common to each data set are the reported longitude, latitude, truck ID (scrambled for privacy in the case of vendor A), and a date and time stamp. Other variables, in some of the data sets, include GPS signal strength, travel heading and the status of a trucks stop (parked with engine on or engine off). Appendix 1 contains detailed information about each vendor's output.

DATABASE DEVELOPMENT

Once contracts had been signed with each vendor, the next step was to set up mechanism to input, clean, store, and manipulate the GPS data from the vendors to create a database. The goal of the truck performance measures research is to quantify truck travel times in the state's largest urban center, the central Puget Sound region, and to demonstrate the feasibility of a statewide truck performance monitoring program. Because of the diversity and large volumes of GPS data, building a program that can support these long-term data needs requires a robust server, automated processing, and ongoing database support. The truck freight database system architecture is shown in Figure 1.

Table 1. GPS Data Overview by Vendor

GPS Vendors	Average Total Daily Records	Total Trucks	Frequency of reads (minutes)	Data type
Vendor A	94,000	Approx 2,500 per day	5-15	Near real-time In-vehicle GPS with a cellular connection
Vendor B	12,000	25	0.5	Real-time In-vehicle GPS with a cellular connection
Vendor C	3,000	60	1-5	GPS cell Phone

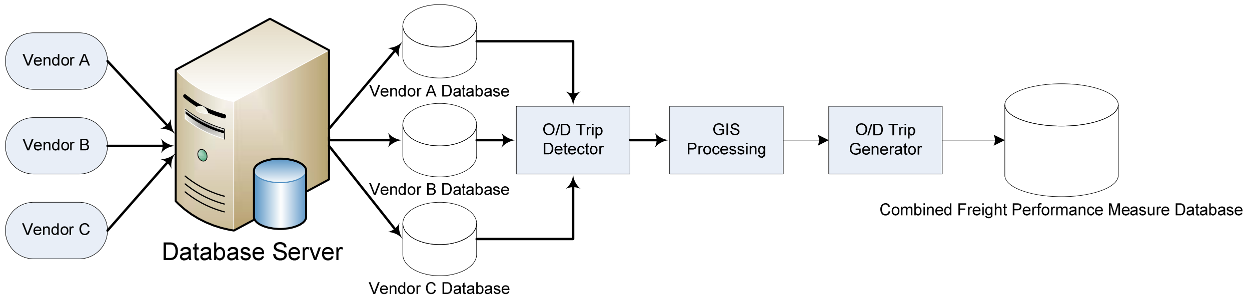

Figure 1: Database System Architecture

Database Organization

The project's database system manipulates the real-time and archived data from the three GPS vendors. Each data set is gathered into a database server that was set up with an automatic program to retrieve the data from each vendor. Because the data sets provided by the three GPS vendors differ from each other, the research team developed a custom interface in cooperation with each vendor's technical staff for each database feed and retrieval program. Technical details about the GPS data feeds is found in Appendix 2.

The three data sets, as received, were initially stored into three databases. However, because thousands of location reads were fed into the database each day, the researchers became concerned that querying and analyzing such a large and growing database as part of a performance measures program could be extremely cumbersome. They decided to optimize the database to enhance performance and wrote a program (in PHP Hypertext Preprocessor script language) to automatically split the three vendors' databases into three monthly data sets. They also created a separate working core data set that contained only the fields necessary for a freight performance measures development program. Common information from each vendor's data set was formatted into standard columns. Additional information unique to each vendor, such as spot speeds, mileage, and data descriptions, was added in separate columns in the combined database. After several programming iterations, querying nearly 10 million rows of data within a second became possible.

Data Processing

Once the project database had been built, the first data processing task was to confirm the trucks' GPS locations (on the basis of the latitude and longitude reported by the GPS) by map matching (geocoding) them to the road network using a geographic information system (GIS). Geocoding is a critical part of a freight performance measures program, as truck performance characteristics must be assigned to the appropriate roadways. However, the use of GIS software to geocode location data to a road network typically creates errors resulting from spatial mismatches between the base network and the latitude and longitude points, as well as other problems, such as assignment confusion regarding overpasses and frontage roads. These issues typically require post-processing and error checking (Czerniak 2002). Because of the size of the data set, the researchers decided that the geocoding process had to be able to automatically identify suspect points so that they could be discarded. The post-processing step was designed to use a GIS scripting (i.e., programming) language to assign each truck's GPS reads to roadway segments. First, the geocoding process had to first identify the roadway corridor of interest and then assign the truck trips to that roadway. Next, each latitude and longitude read was assigned (snapped) to the nearest roadway by using a 100-foot buffer around the road. Heading data for the trucks were then checked to associate the GPS travel bearing with the road segment's bearing.

A test of the geocoding rules on a case study on Interstate-90 indicated that about 60 percent of the points could be retained by using the geocoding process. The points that were eliminated were on cross-streets, or their locations were ambiguous. While a portion of these points could have been located by using manual techniques, the size of the database precluded using such resource intensive methods. Nevertheless, this was considered a successful indication that the fleet management data can be effectively assigned to pre-selected roadways. Appendix 3 covers the technical details and limitations of the geocoding process for the I-90 case study.

IDENTIFYING ORIGINS AND DESTINATIONS

Algorithm Development

Developing a performance measures program requires an understanding of the route choices and travel patterns of trucks. The program needs to monitor travel generated by trucks as they respond to shippers' business needs, picking up goods at origins (O) and dropping off goods at destinations (D). Such understanding requires identifying each origin and destination where a truck stops to complete the productive transaction that defines the purpose of each truck's trip. Because GPS devices record all stops, the team developed a methodology to differentiate between traffic-based stops at intersections or in congestion and stops at origins and destinations. Several studies that have analyzed GPS truck data have attempted to separate traffic stops from O/D stops on the basis of the stop duration (i.e., dwell time) (Greaves and Figliozzi 2008, Du and Aultman-Hall 2007, Schuseeler and Axhausen 2008, Hunter, Wu and Kim 2008). In these studies the O/Ds identified by stop duration were also manually examined to identify any unusual situation or problem.

For this effort, the project team also developed a stop duration tool in an attempt to identify trucks' origins and destinations. Because of the large size of the GPS data sets involved, an automatic and efficient trip end identification algorithm was required. The resulting algorithm was based on a dwell time plus a distance threshold to detect the trip's origin and destination. The algorithm was written in Java to handle the complexity of the data.

Determining the dwell time threshold setting was critical, as too large or too small a time could either miss or incorrectly identify trips. The selection of the dwell time is a function of traffic conditions in a given city or area (Du and Aultman-Hall 2007). McCormack and Hallenbeck's truck GPS data research completed earlier in Washington state determined that 3 minutes is a reasonable dwell threshold (2006). The 3-minute period filters out most trucks' non-O/D stops for traffic signals, since most signals have a shorter cycle length. In addition, traffic congestion in which a truck does not move for more than 3 minutes is unusual in the central Puget Sound.

In the process of developing the algorithm, occasional GPS signal blockage occurred when overhead obstructions such as tall buildings, canopies over loading docks, or tunnels prevented GPS devices from communicating effectively with the GPS satellites (Greaves and Figliozzi 2008, Czerniak 2002). This resulted in a loss of O/D data in a few cases. Fortunately, vendors of GPS devices compensate for short-term signal loss by simply waiting until an adequate number of satellites are available before reporting a position. During a short period without a GPS signal, one can assume that a truck continues to travel at a constant speed. By using the GPS points recorded before and after the signal loss to calculate the average speed, it was possible to set up a threshold speed limit (5 mph was selected). If average travel time is below this threshold, the program determines that a trip end has occurred in an area of signal loss.

In many cases, when a series of the GPS points for a truck was mapped in GIS software, the points were found to fluctuate around a position when a truck idled (an occurrence known as jiggle or wandering), creating a false report of movement (Czerniak 2002). This was due to GPS signal inaccuracy. To address this issue, the distance between consecutive GPS points was used to refine the O/D algorithm. If the difference of longitude and latitude for two consecutive GPS points is less than 0.000051 degrees (around 65 feet), a flag is tagged to the record. When the average speed for this trip is calculated, this delay time due to the fluctuations is subtracted to achieve a more accurate result.

The algorithm was programmed for a second round of screening to detect abnormal trips. These errors are typically caused by the following:

- trucks leaving the Puget Sound study area
- travel in areas with tall buildings that results in poor GPS signal reception
- GPS multipath sign interference, in which the GPS signals are reflected off of the surface of objects located between the GPS satellites and the GPS devices in trucks (Czerniak 2002).

These trips are flagged. This secondary process detects the following:

- extremely short trips between GPS reads for an individual truck
- false trips in which the elapsed truck travel time is zero
- trips with extremely high speeds
- trips in which a truck's O or D is external to the study area and cannot be captured by the O/D algorithm
- external trips in which a truck crosses the study area boundaries and then returns.

These flagged trips can be removed from any freight performance program calculations that require travel time and travel speed.

Validation

The initial O/D algorithm was tested by using one month of vendor A's data as the test case. This data set contained nearly 3 million GPS records and used about 1 GB of file space. The O/D algorithm was run on an Intel Pentium Xeron processor and required about an hour to run. The origin and destination pairs generated by the trip end identification algorithm from vendor A's one month data set included 358,692 trips of which 6,443 were abnormal trips flagged by the second round of trip processing.

One advantage of this algorithm is that it allows a range of summary statistics to be developed that will be valuable for a freight performance measure program. For example, the average trip distance between origins and destinations for the month of test data was 16 miles, the average travel time was 21 minutes, and average travel speed was 34 mph. Although the standard deviations for trip distance and travel time varied, they still fell within a reasonable range given the size of the Puget Sound study area.

Because computing efficiency will be critical for the large freight performance measures data set, the O/D algorithm was also tested on each of the three vendor's databases, which included 320 million GPS points (occupying 10 GB of space). The process required 24 hours.

To validate the accuracy of the assumptions behind the O/D algorithm, several truck IDs were randomly selected from the database, and their entire daily trips were geocoded onto a regional road network in GIS software. Each truck's trip was followed from stop to stop along the road network, and the team located the stops by using Google Earth™, Google Maps™, and aerial photos. This process looked at the O/D locations calculated by the algorithm to determine whether the trucks' stops corresponded to locations with truck terminals, warehouses, or other reasonable delivery and pick-up locations. Figure 2 shows a typical view of the origin and destination locations presented by the GIS.

In the example shown in Figure 2, one of the randomly selected trucks was assigned by the algorithm to an origin at 1:43 PM just west of Buckley and a destination approximately 10 miles away at 2:12 PM in Auburn. The travel time was reasonable, given the distance, and examination of the origin site address in Google Maps showed many trucks in a parking lot. The destination also showed a number of trucks near one side of a building. This information suggested that each stop was at a location that could be a logical stop for a truck, either to park at a terminal or to load or unload goods.

A second validation test of the O/D algorithm was conducted by using a GIS and assigning the GPS truck origins and destinations to the 938 transportation analysis zones (TAZs) that the Puget Sound Regional Council (PSRC) uses for travel demand modeling. Each GPS truck trip included a travel distance, time, and speed between an origin TAZ and destination TAZ (or O/D pair). The average travel distance and travel time for all GPS trucks between each pair of zones were compared to the PSRC's travel time and distance network statistic for same pair of zones.

When large difference occurred between GPS travel statistics and the network data, the data were examined in greater detail. Figure 3 shows one example that required additional evaluation, a set of GPS truck trips that originated in one TAZ in Kent Valley and ended in another TAZ in south Snohomish County. The PSRC TAZ network indicated that the shortest travel time trip would probably include travel on I-405. Because truckers are concerned about

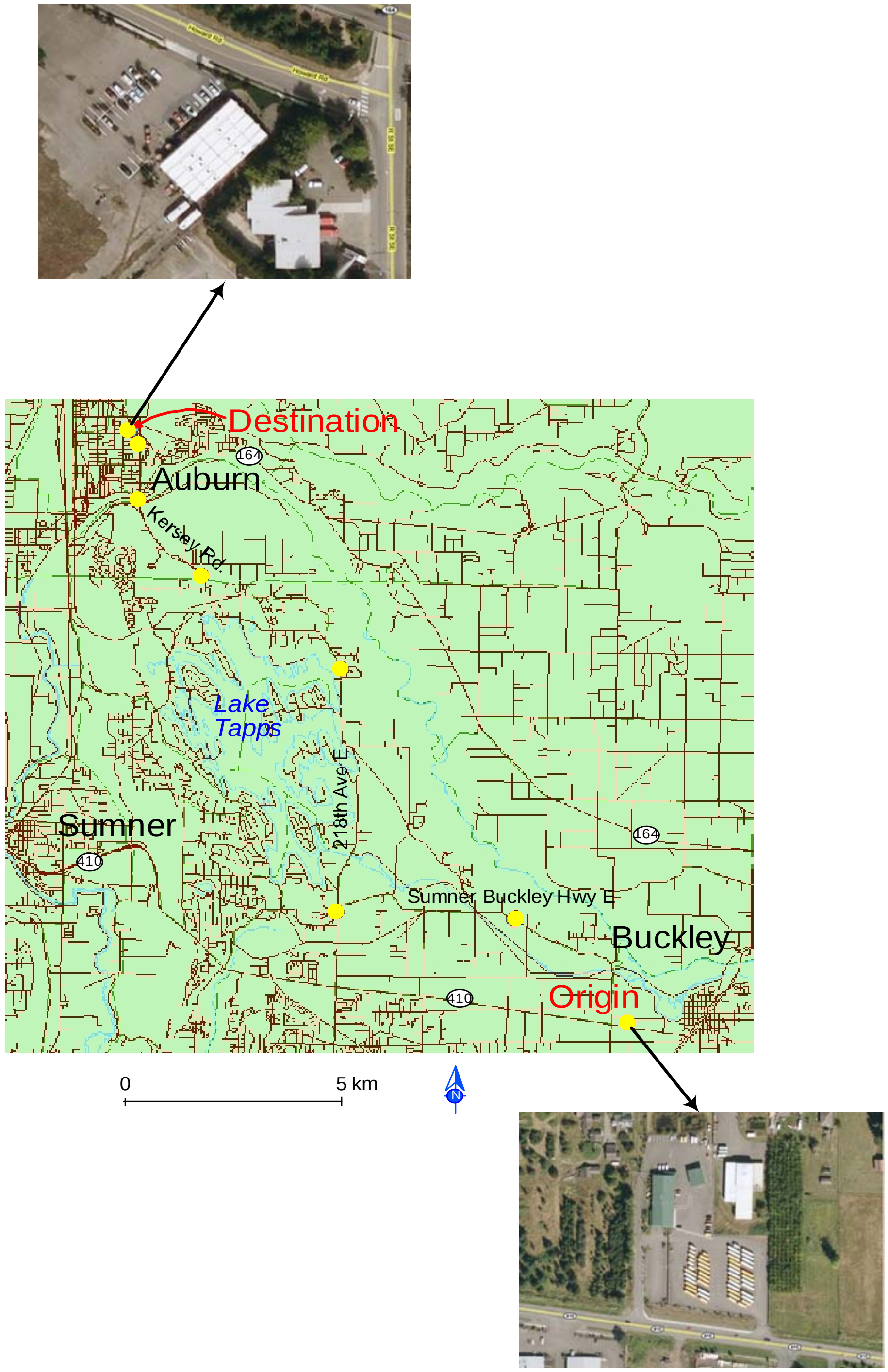

Figure 2: Origin and Destination Validation with GIS and Google Earth™

labor and fuel costs, logically I-405 would be the most efficient route for any trucker trying to travel between these two zones. Figure 3 shows all GPS truck trips for a year, with one trip

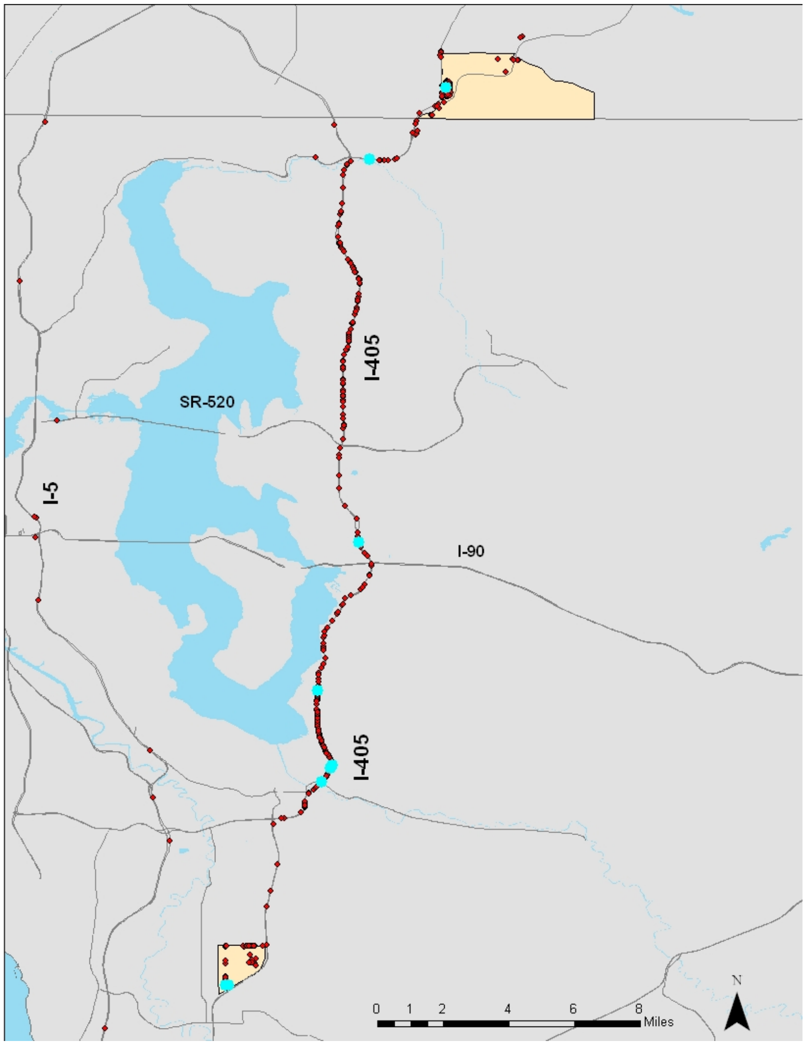

Figure 3: Origins and Destination Example

highlighted in blue. It is clear that most trips did occur on I-405, but some alternative routes are indicated by points along I-5. This long diversion from the most efficient route suggests that some truckers had a business stop (i.e., a destination) that was not captured by the O/D algorithm.

Through observation of outliers such as shown in Figure 3 and by isolating trips between specific TAZs, the route choice of trucks and performance of the O/D algorithm were evaluated. This comparison of GPS truck travel numbers with the TAZ network data revealed two issues with the dwell-based O/D algorithm that required addressing. First, the researchers found numerous short stops (stops of less than 3 minutes) for trucks in vendor A's database that were not recorded as destinations but should have been. The analysis revealed a number of trucks traveling significant distances to a different TAZ, making a short stop and then coming back to another TAZ for a final destination stop. This occurred when these trucks made short delivery stops that were less than 3 minutes, and the O/D algorithm, with a dwell time of 3 minutes, did not capture these short stops. This problem was addressed by taking advantage of the fact that vendors A's GPS data feed includes engine park and stop status information. A stop was identified as either the truck engine off or the truck in park with the engine on. The researchers determined that any stop where the engine remained in park could also be considered an origin or destination. The algorithm was revised to include all of these "park" stops.

The second problem was that some truck trips originated inside the study area, then left the area, and then returned to a final destination inside the study area. When they left the study area, GPS travel data were lost because they were not provided under the contract with vendor A; once they returned to the area, the GPS data feed was reacquired. These external truck trips were sources of errors because a truck could leave the study area and travel an additional 200 miles, whereas the TAZ O/D pair for that truck might only be 20 miles apart, leaving 180 miles of travel unaccounted for (because any origins and destinations outside the area were not captured). These trips were not common and were identified through a change in the algorithm's logic. This logic looked for consecutive records that indicated a truck was moving and whether that truck had traveled an extreme distance between those records. One hundred miles was set as this distance threshold, and if a trip fit these parameters, it was flagged as an external trip.

A final category of trips also needed to be flagged. A few trips had much longer GPS travel distances than were reasonable, given the PSRC zone to zone network distance, but remained internal to the study area, did not have any stops longer than 3 minutes, and during which the truck was never placed into a "park" status. Investigation of these data indicated that a few trucks made deliveries without the need to park. One example of this situation was a trash container company that delivered large empty containers to a construction site. Because the container were tilted off the back of the trucks, the driver never needed to stop or even get out of the vehicle to make a delivery, and the GPS never recorded park or stop status.

To separate different types of trips before the development of performance measures, a trip type variable was created to categorize each trip. The following summarizes the three different trip types that were used along with a percent of total trips derived from a sample month of GPS data which included 32,000 trips in April, 2009.

- *Local Trips.* These are trip during which the driver drops off packages without stopping longer than 3 minutes. These trucks are probably small package delivery trucks (38 percent of trips).
- *Loop Trips.* These trucks, as discussed above, complete their business without a need to stop or place the truck into park. Few trucks fit this category, and most are garbage or construction trucks (9 percent of trips).
- *Access Trips:* This is the most common trip type, in which the trucks have clearly defined origins and destinations, and each stop can be defined by a stop longer than 3 minutes (53 percent of trips).

DEVELOPING PERFORMANCE MEASURES

The GPS truck data are being collected in order to develop performance measures by quantifying truck travel in the central Puget Sound region, and to demonstrate the feasibility of a statewide truck performance monitoring program. By using cases studies, the data were evaluated as a tool to

- benchmark roadway performance before and after a construction project or major infrastructure changes
- locate roadway bottlenecks
- examine the performance of truck travel between major regional destinations.

An urban section of I-90 across Lake Washington and travel on SR 167 were used as case studies to explore the ability of the GPS data to create these measures.

Performance Measures for Benchmarking Construction

Can the GPS data be used to benchmark the performance impacts of construction? An I-90 bridge project was selected for the case study to explore these questions. I-90 was used because this roadway is the primary east-west corridor for truck freight in Washington, and a large number of vendor A’s GPS-equipped trucks traveled on I-90. This section of I-90 is also well equipped with FLOW system traffic monitoring devices, so the researchers could compare the GPS data with travel data from all vehicles. The FLOW system is a coordinated network of traffic monitoring and measuring devices that operates on urban state and interstate highways in the central Puget Sound area (Ishimaru and Hallenbeck 1999). The test section also includes several tunnels where GPS signals are blocked. This provided an opportunity to evaluate the impacts of GPS signal loss. Finally, a major construction project closed a portion of the roadway for several weeks, providing an opportunity to benchmark project impacts.

The construction project was intended to replace the bridge's expansion joints, which had deteriorated and presented a safety risk. To protect drivers, WSDOT replaced the aged and worn expansion joints on the center roadway reversible lanes from May 4 to 22, 2009. This closed the reversible bridge's express lanes, thereby reducing capacity on the bridge over Lake Washington for that period.

The project team analyzed GPS data for two to four weeks before construction, during the three-week construction period, and for two to four weeks after construction. The data from vendor A was selected because vendor A had the highest frequency of GPS reads and the most number of trucks traveling the Puget Sound region.

The data were broken down spatially to assess the impacts of truck traffic during these periods on the 1.5 miles of the bridge span, 1.5 miles west of the bridge, and 1.5 miles east of the bridge for travel in both directions. Figure 4 shows the geographic scope of the data points analyzed for both eastbound and westbound directions of the I-90 corridor. Data gaps just east and west of the bridge occurred because of the tunnels that blocked the GPS data signals.

The average truck speeds (in mph) were separated into four daily periods. They were as follows:

- Morning Peak: 6:00 am - 9:00 am
- Midday: 9:00 am - 3:00 pm
- Evening Peak: 3:00 pm - 7:00 pm
- Night: 7:00 pm - 6:00 am

While the purpose of this case study was to evaluate the usability of the GPS data, a discussion of the findings is relevant to understand the output and significance of the tool. The case study, by using speed data from the GPS equipped trucks, found that the lane closures during construction had a minimal impact on truck travel times across and near the bridge. On average, trucks traveled at or close to the posted speed limits. However, at some times of the day and on some specific road segments this study did reveal significant variation between truck spot speeds and all other traffic performance captured by FLOW data. For example, westbound

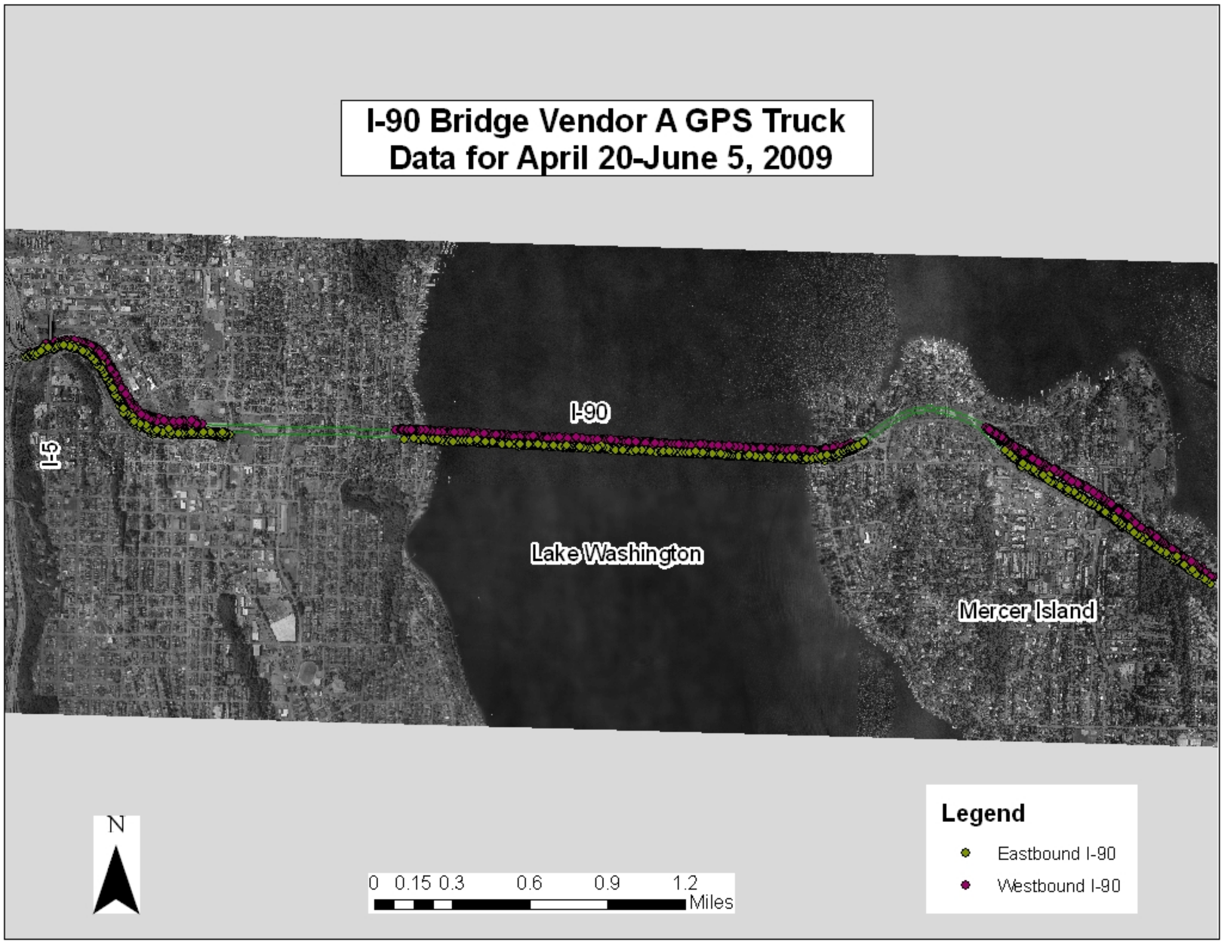

Figure 4: I-90 Geographic Scope of Eastbound and Westbound Data Points

truck speeds west of the I-90 Bridge dropped below pre-construction truck speeds by about 8 mph during construction (see Figure 5). Appendix 4 contains a detailed analysis of travel time findings for the case study.

The I-90 project was announced well before it started, so diversion to alternative routes was possible and could have reduced the impact of the construction. Truck diversion was evaluated to account for trucks that diverted to SR 520 as an alternative route during the construction period. Findings were inconclusive, as the overall truck volumes of GPS-equipped trucks in vendor A's fleet did not appear to drop significantly on I-90 or to rise significantly on SR 520 during construction . A significant drop in daily truck volumes two weeks after construction may be explained by Memorial Day occurring on May 25. Note that further

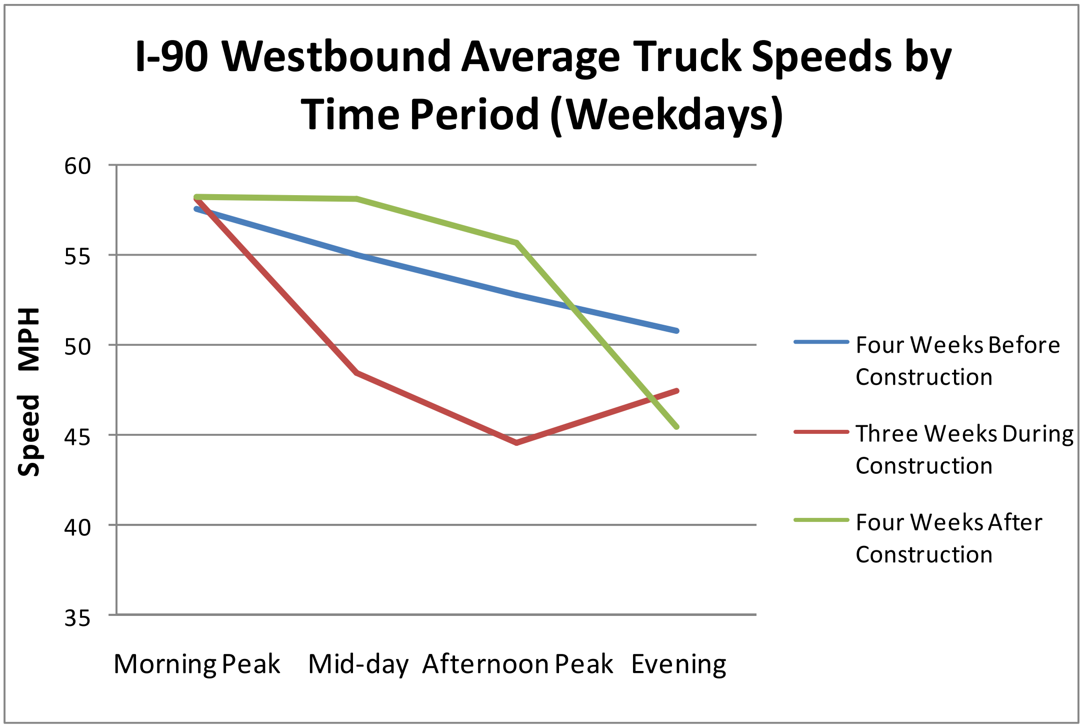

Figure 5: Example of Benchmark Data

analysis on diversion routing may be necessary to fully understand truck performance on I-90. A diversion analysis would not only need to look at the diversion patterns of the project's GPS-equipped trucks but also to use count data, if available, for all trucks from sources such as WSDOT automatic classifiers and portable truck counters.

GPS Data Issues

The analysis of the data for the I-90 case study highlighted the need to carefully process the GPS data before they are used as part of a freight performance measures program. Some data limitations that should be considered include the following:

- **Uncertainty about Truck Type.** The GPS data do not contain information about truck types. In some cases, this may result in confusing results. For example, a number of concrete trucks carry vendor A's GPS devices. As a result, some of the captured data included trucks that were part of the construction effort on the I-90 Bridge at the time of the road closures. The express lanes were completely closed from May 4 to 23, but the project data contained trucks in the express lanes that moved at a very slow speed and returned multiple times. Closer scrutiny suggested that these were concrete trucks participating in the construction project, and the data were removed from the analysis.

- **GPS Signal Loss.** As mentioned earlier, processing algorithms need to account for an occasional loss of the GPS signal. Two tunnels to the east and the west of the I-90 Bridge block GPS signals. The GPS devices could not register their locations while in the tunnels, so the data were registered at the last known location before each tunnel, creating data point clusters near the tunnels heading both eastbound and westbound (see Figure 6 for data issues at the east end of the bridge GPS).
- **Roadway Network Inaccuracy**. The PSRC network data used for the case study are a set of mapped lines that do not necessarily indicate the midpoint of the actual roadway. Therefore, spatial analysis had to be checked manually to ensure that the points were geo-located to the correct roadway. For example, in the PSRC network database, the express lanes and the eastbound lanes at the east end of the I-90 bridge come closer together than the actual roadways do. Because of this inaccuracy, the GIS package geocoded data points to the express lanes, even though the points were actually from the eastbound lanes (see Figure 6). These location points required manual cleaning to re-assign them to the correct lanes. This manual processing would be difficult to complete on a large scale.
- **Data Cleaning Difficulties.** For this analysis, a manual cleaning methodology was used. Data were organized on the basis of their direction heading and visual confirmation of their roadway location. As a large amount of GPS data needed to be analyzed, the system required automated processing for efficiency. The system had to organize data for analysis and clean the data on the basis of direction and proximity to important truck routes. Additionally, a number of data points had heading values that were unrealistic (a 382-degree heading on a 0- to 360-degree scale). After the researchers talked with the GPS vendor, these points were eliminated for this analysis because they were errors in the GPS devices due to position wander or GPS satellite misreads.

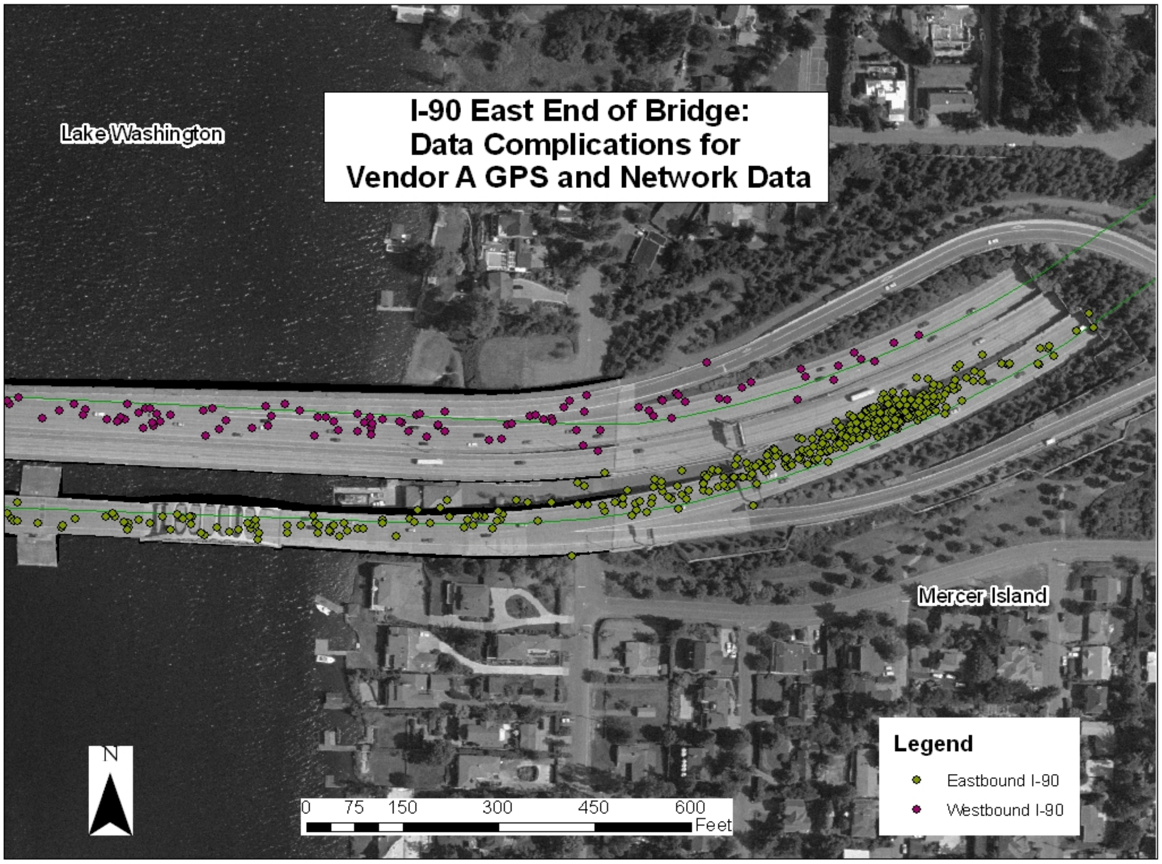

Figure 6: East End of I-90 Bridge: Clustering before the tunnel and express lane.

Location of Travel Bottlenecks

The GPS data were examined for their usefulness in quantifying the mobility and reliability of truck travel at specific roadway locations. For most organizations, mobility is a function of travel speed that relates directly to travel times on selected corridors or segments of a roadway network. Reliability is also related to speed, but the speed is compared to a typical or uncongested (free-flow) roadway travel time. Once roadway performance on individual segments have been quantified, these roadway segments can be linked to examine truck travel along multiple segments on high-use truck routes.

Previous research has focused on estimating link travel times on the basis of GPS data (Lee, Lee, and Yang 2006; Sananmongkhonchai, Tangamchit, and Pongpaibool 2008). These studies have used GPS spot speeds to compute the link speeds. However, most of their algorithms have been based on only a few GPS probe cars, with GPS location sampling rates of every few seconds. This research explored ways to calculate link travel times from GPS with a low sampling rate but a large number of trucks.

Vendor A's data provide position reads when a truck starts and stops and approximately every 15 minutes when a truck is moving. The 15-minute position reports also include a spot speed, which was the basis for the I-90 case study. These speeds are the instantaneous travel speeds recorded by the GPS at the time that the truck's GPS device records and transmits its location. As a result, the speed data from the trucks are not necessarily representative of the travel performance of trucks for the roadway segment. The spot speeds, for example, may reflect a truck that is slowing to allow cars to merge from a ramp. Spot speed measures are more variable than true roadway performance because they report the variations that occur as individual vehicles travel. While these variations are representative of the actual vehicle performance, they tend to overstate the variability of the roadway.

This phase of the project evaluated whether GPS spot speeds from vendor A on one segment or link provided a sufficient sample to develop a performance statistic that was a reasonable tool for locating recurring bottlenecks for trucks. To do this, the researchers compared spot speeds with 5-minute speeds calculated by WSDOT's FLOW system for the I-90 test section. Researchers used FLOW speeds calculated from loops in the far right lane only because that is where trucks are required by state law to travel unless they are passing slower traffic. The researchers randomly selected 210 truck spot speeds from I-90 and then matched them to FLOW speed data with the same time stamp (within 5 minutes). When a FLOW speed was higher than the 60 mph threshold, the FLOW speed was reported as 60 mph, since this could be regarded as free flow speed. Figure 7 shows a case in which most of vehicles were at free flow speed, but 24 trucks with the same timestamp had GPS speeds slower than or equal to the flow speed. Only 11 percent of trucks traveled at a speed the same as or higher than the FLOW speed with the same timestamp. When a truck's GPS speed was higher than 60 mph, it was likely that traffic flow speed was higher than 60 mph but was truncated. This situation represented overall free flow traffic conditions.

These findings were reasonable, since an earlier Puget Sound study of trucks equipped with GPS devices found that truck speeds were slower than the FLOW loops indicated for all freeway traffic (Hallenbeck, McCormack, Nee, and Wright 2003). In that study, cars equipped with GPS devices were sent out to travel in the far right lane to mimic the performance of fully loaded trucks. These runs indicated that trucks were more significantly affected by merge-related congestion at ramps than cars, since trucks could not change lanes to avoid the congestion, had to decelerate early to avoid the congestion, and also accelerated slowly after being forced to slow down. (This phenomenon is also supported by informal discussions with truck drivers.) These findings led to the conclusion that the FLOW speed data often overestimate roadway performance, whereas trucks' GPS spot speeds underestimate overall freeway travel speeds but can indicate the travel speeds of individual trucks. This difference is exacerbated because some FLOW data loops are close to ramps, where significant congestion occurs. Thus, the FLOW loops tend to not capture the slowest vehicle speeds, whereas the GPS devices capture travel on all locations on the freeway. As the report concluded, "averaged loop data may slightly over-estimate road performance (as experienced by heavy trucks), and truck-based GPS data tend to under-represent it" (Hallenbeck et al. 2003, p. 44).

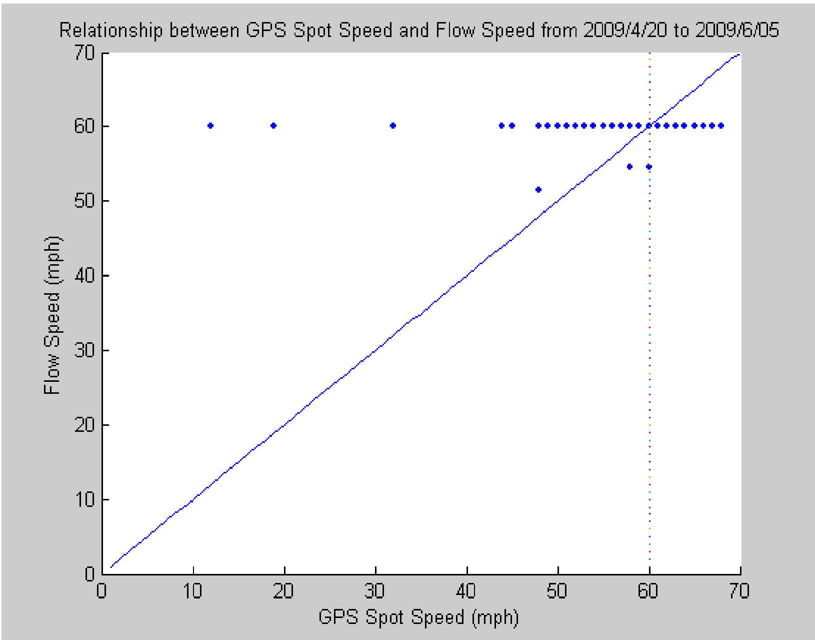

Figure 7: GPS Spot Speeds and Flow Speeds

When spot speed data were compared to FLOW data over four daily time periods for I-90 (results are shown in Appendix 4), the averaged spot speeds occasionally matched the general purpose speeds extracted from FLOW data. However, there were also situations in which the spot speeds differed from FLOW speeds. For example, in the evening period, westbound, during the third week of construction, the FLOW speed for the roadway segment east of the bridge was

35 mph, whereas the average spot speed for trucks was 54 mph. These differences again indicate that spot speeds may provide different information about freeway travel than do FLOW data.

This analysis suggested that spot speeds can capture roadway performance when conditions are at free flow. That is, if a number of trucks are traveling at maximum speed on a roadway segment, it's a fairly good indication that truck travel is flowing freely. However, spot speeds may need to be used with some care if they are to accurately represent congested conditions. If a GPS read shows that a truck is traveling at less than free flow speeds, it could suggest either congestion or that a particular truck was decelerating during the GPS reporting period. In essence, when a heavily loaded truck is monitored in urban conditions, the GPS data are often measuring the slowest vehicle, since these trucks are the most constrained by road conditions because of their inability to accelerate, decelerate, or change lanes in moderate to heavy congestion. This finding also suggests that FLOW data do not always represent actual travel conditions experienced by trucks and that GPS data could offer a more accurate truck performance measurement alternative.

These findings also suggest that GPS spot speeds, because of the limited volume of GPS-equipped trucks, may not always be the best tool for measuring short-term road performance. More specifically, the analysis of I-90 highlighted data limitations with the GPS probe trucks. Even on a major travel corridor such as I-90, the number of GPS-equipped trucks on a given segment of roadway during a given analysis period is limited. For example, the greatest volume of GPS-equipped trucks (westbound) was 70 trucks per day, which made it difficult to establish statistical reliability in the short term.

The best use of spot speeds is to indicate travel performance over time. This project is collecting GPS data for a period of a year or more. Average truck spot speeds analyzed over long periods, rather than just several weeks as in the I-90 case study, will result in a more accurate indicator of recurring roadway performance for trucks. The extreme speed values due to truck acceleration or deceleration will average out. Portions of roadways with trucks traveling at speeds consistently lower than free flow speeds can be identified as bottlenecks. This use of average spot speeds over longer periods is how ATRI calculated truck bottlenecks in its national study. Spot speeds were averaged for one year and then compared to free flow speeds (Short, Picket, and Christianson 2009).

Figures 8 and 9 show that I-90 GPS spot speeds can be aggregated over a year to identify potential bottlenecks on a roadway. By using a year of GPS data for I-90 (10,990 trucks westbound and 11,100 trucks eastbound), the research team was able to support a hypothesis that there is a bottleneck west of the I-90 Bridge. By using three speed categories (>50 mph, 25-50 mph, and <25 mph) and comparing three similar road segments, the figures show that the greatest percentage of slow truck travel occurs on the roadway segment west of the bridge. Westbound GPS truck speeds were less than 25 mph 25 percent of the time and were 25 to 50 mph 36 percent of the time. This finding is further supported when the average spot speeds are compared to average FLOW speeds, as demonstrated in Figure 10. Trucks at the roadway segment west of the bridge traveled notably slower than did the general purpose traffic, implying that this roadway segment is difficult for trucks. These examples demonstrate how spot speeds can be used to identify bottlenecks, and can indicate travel performance over time, and are

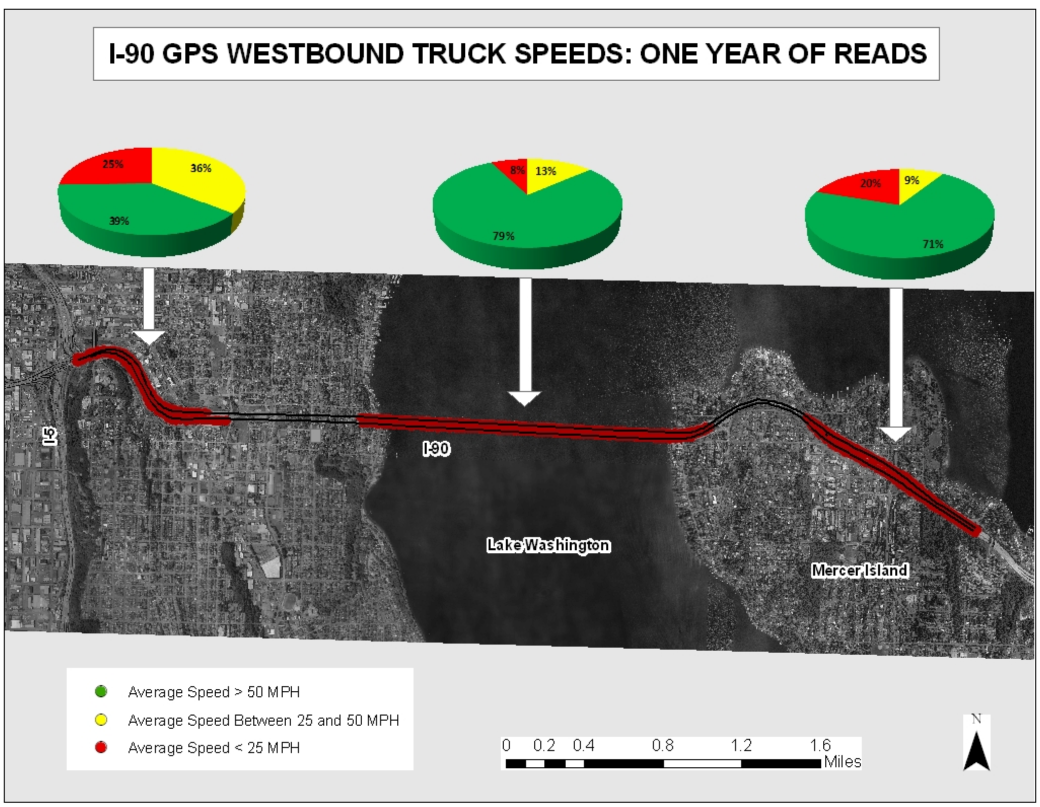

Figure 8: I-90 Westbound Truck Speed (Averaged over One Year)

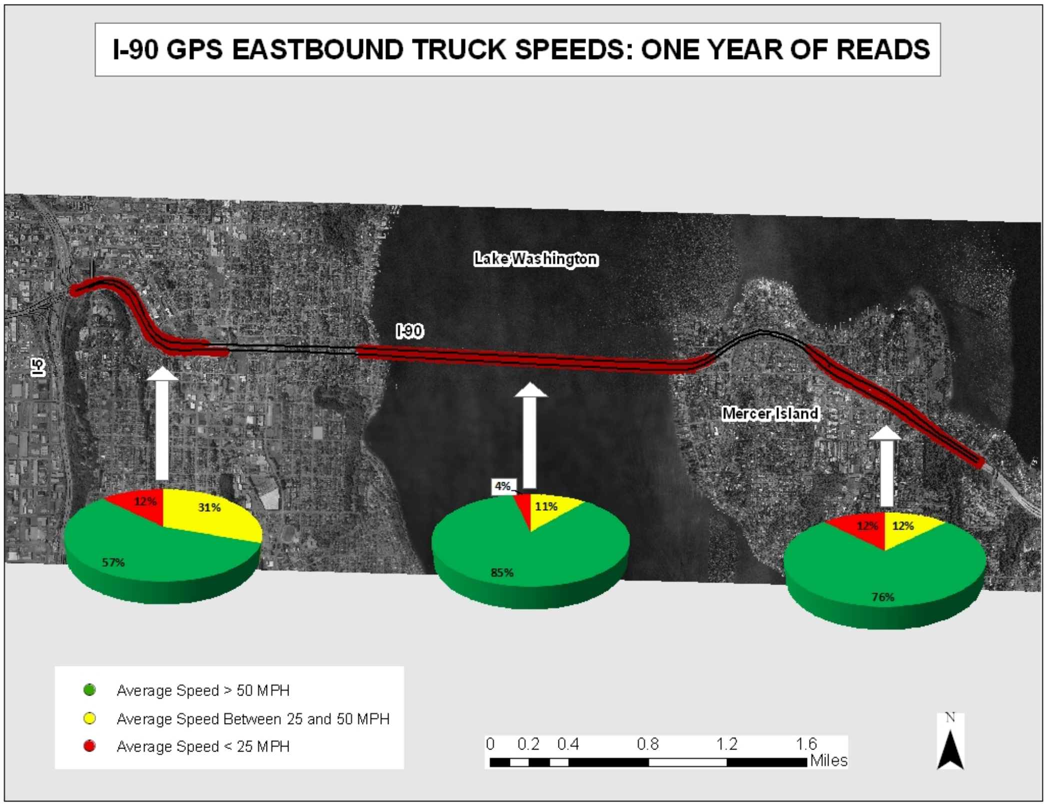

Figure 9: I-90 Eastbound Truck Summaries (Averaged over One Year)

reliable when sufficient data are used. As a result, the use of spot speeds will be most relevant on high-level roadway sections (freeways and interstates) with larger volumes of trucks. Spot speeds are less useful on arterial-level roads with traffic control devices, since it difficult to differentiate intended stops or slowdowns at intersections from slowing due to bottlenecks and roadway problems. The spot speeds from vendor A's GPS-equipped trucks will not be as useful for examining arterial-level roadway performance as they are for calculating travel performance over long distance routes.

SR 167 Comparison Study

A second case study used the same GPS-based analytic tools and applied the methods of the I-90 case study to a 15-mile-long segment of SR 167. The truck spot speed data were averaged over the course of a year (September 2008 to August 2009) and compared to general traffic speeds (FLOW speeds) on three road segments. (Note that no FLOW speed data were available for the south segment south of SR 18.) As expected, the average truck speeds were typically slightly lower than general traffic flow. The biggest difference in speeds was 6 miles per hour in the middle segment, southbound (Figure 11). This comparison study demonstrated that the spot speed analysis methodology is more consistent and can offer more accurate performance measurement over a period of one year.

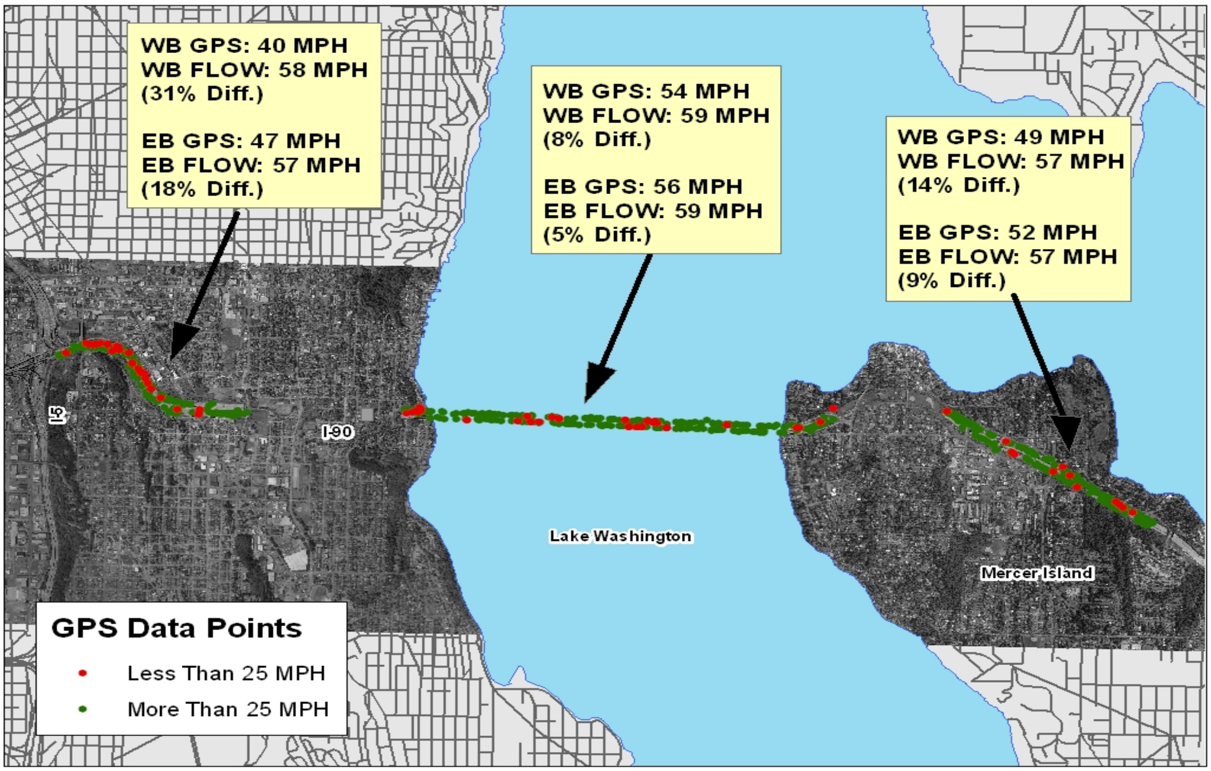

Figure 10: GPS and FLOW Speeds on I-90 (Averaged over One Year)

Zonal Performance Measures

Because the truck trip O/D data were assigned to traffic analysis zones (TAZs), regional travel could be quantified between these zones, providing an effective platform for evaluating regional truck travel. The advantage of examining zone-to-zone travel with GPS data is that the data can be used to monitor network performance between economically important areas even when truck drivers choose multiple connecting routes. Another benefit of the zone-level analysis is that the most economically important zones with the most truck trips are also ones that should be the most represented by the largest numbers of O and Ds. Figure 12 shows the most active truck origin and destination zones based on the GPS data from vendor A.

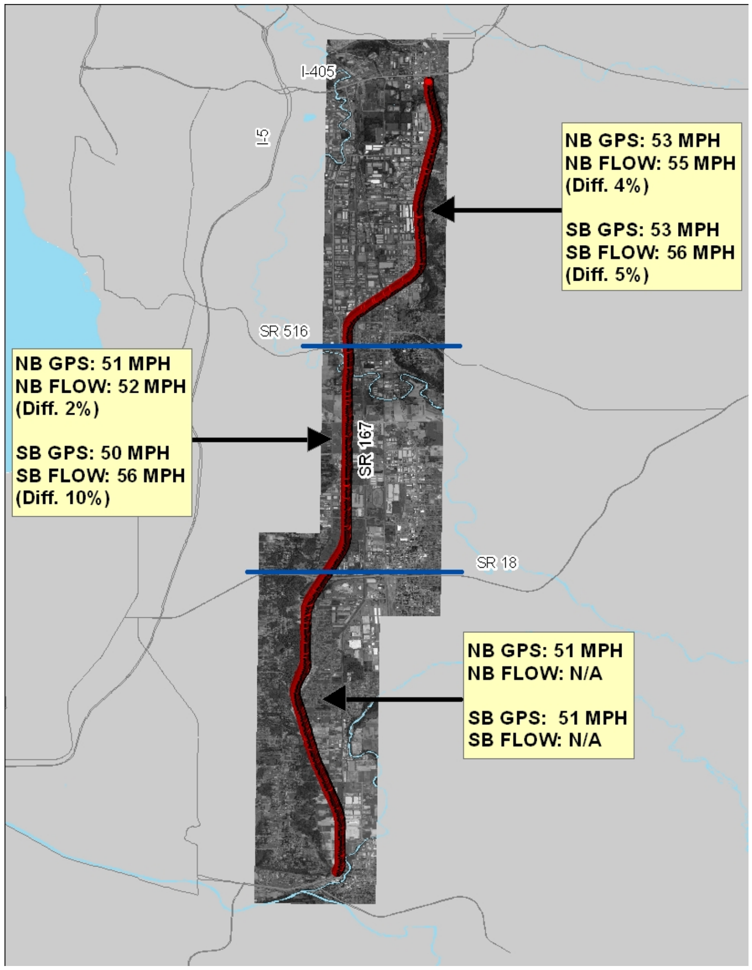

Figure 11: SR 167 GPS Speeds Compared to Total Traffic Speed

A range of quantitative performance measures can be calculated for the zone to zone travel. Because the GPS data provide, at base level, travel distance and speed for the probe trucks, all the measures are derived from these numbers and can be compared to ideal or free flow travel conditions. A previous literature survey completed as part of this project, as well as works by others (IBI Group 2009), suggest a number of truck oriented performance metrics that can be used to analyze freight performance between zones., Table 2 demonstrates the application of these freight performance measures to a year's travel (September 2008 to September 2009) from TAZ 340 and TAZ 385 on the basis of data from vendor A. These zones are both in the Kent Valley, where a number of manufacturing and distribution centers create numerous truck trips.

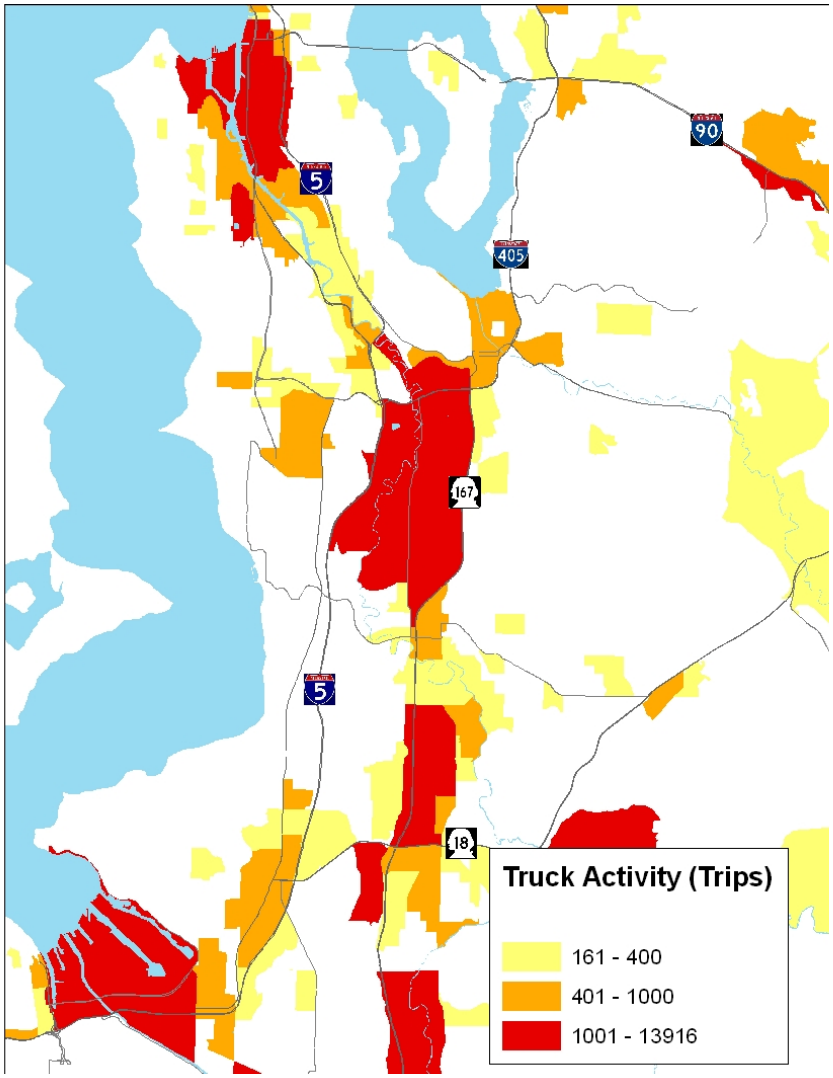

Figure 12: Truck Trip TAZ Activity by Zone

The GPS data processing program developed for this project automatically calculate these measures for zone-to-zone trips. The free-flow speeds between the TAZs used in these performance measures are based on network data from the Puget Sound Regional Council. The PSRC uses posted speed limits and distances to calculate the ideal travel speed between each pair of TAZs, given the time of day. There are two types of PSRC TAZ travel speeds: with and without terminal times. Terminal times represent the time to travel from within a zone to the geometric center of the TAZ (known as a centroid). Therefore, free flow speed without terminal times is measured from one TAZ centroid to another TAZ centroid, and free flow speed with terminal times is considered the traffic condition from the TAZ edge to the centroid, which is a more accurate measurement for free flow speed. Because the PSRC speed data account for congestion by time of day, the project team used the nighttime travel with minimal congestion (10:00 P.M. to 4:00 A.M) free flow speed with terminal time as the base.

Table 2. Sample Output—Performance Measures between TAZ 340 and TAZ 385

Performance Measure	Results
Total Trips	2394
Total AM Peak Trips (6:00 AM-9:00AM)	30
Total Midnight Trips(9:00AM-3:00 PM)	190
Total PM Peak Trips (3:00PM-6:00PM)	401
Total Evening Trips (6:00PM-9:00PM)	426
Total Overnight Trips (9:00PM-6:00AM)	1347
Average Travel Time (minutes)	13
Variability of Travel Time (minutes)	4.7
95th Percentile Travel Time (minutes)	19
Average Travel Speed (mph)	30.7
Variability of Travel Speed (mph)	7.4
Average Travel Distance (mile)	6.5
Travel Time Index	0.92
Planning Time Index	1.34
Buffer Time Index	1.46

Sample Sizes

Using GPS to track truck movements can be a cost-effective method for monitoring the freight system (IBI 2009). However, a sufficient number of trucks must be on a roadway to develop valid performance measures. The project team calculated that the total number of useable GPS-equipped trucks represents about 1.5 to 2.5 percent of all trucks on a State Route or Interstate Highway (the I-90 and SR167 examples are shown in tables 3 and 4). This percentage may vary, depending on the number of lanes on the roadway, its location, and its capacity. Smaller highways or arterial streets have far fewer trucks than large interstates, so the number of GPS-equipped trucks will be too small to represent all trucks.

Truck numbers are taken from WSDOT truck counts from automatic vehicle classifier (road loops) over the course of a year. The WSDOT equipment counts trucks as they pass over a given point. The GPS-equipped trucks were counted on the basis of the number of reads in a given day over a 4-mile stretch of roadway, including the WSDOT FLOW counter location. This 4-mile threshold was established because the points are taken at 15-minute intervals, and if a truck is traveling 60 mph, it will be registered within a 4-mile area. However, when more frequent reads are taken, truck counts will be more accurate, as smaller areas can be examined.

Determining the percentage of GPS-equipped trucks in comparison to all trucks is valuable, but in some cases a GPS truck sample size must be determined to statistically estimate a reliable link speed. Fortunately, a number of studies have been conducted that determined the appropriate sample size for estimating link travel time and travel speed (Cheu et al. 2002, Chen and Chien 2000, Nezamuddin et al. 2009, Li et al. 2002, Quiroga and Bullock 1998b). The information from these studies was used to develop statistical significance measures for the GPS truck data that the processing software developed for this program can calculate. Appendix 6 presents the underlying equations used for sample size calculation. For example, applying this equation to the Kent Valley travel data between TAZ 340 and TAZ 385 results in a minimum sample size of 16 trips required to calculate an average travel speed within 10 percent for a confidence level of 95 percent. Appendix 6 includes more details about this calculation.

Table 3. SR167 Daily Truck Totals (Both Directions)

Time Period	GPS Equipped Trucks Number/Day	Total Trucks Number/Day from Counts	Percent of Total
North Segment (MP 2.5)	266	10370	2.5%
Middle Segment (MP 17.5)	209	14270	1.5%
South Segment (MP 11.8)	283	12040	2.4%
TOTAL	758	36680	2.1%

Table 4. I-90 Daily Truck Totals (Both Directions)

Time Period	GPS Equipped Trucks Number/Day	Total Trucks Number/Day	Percent of Total
TOTAL	109	7716	1.4%

PROGRAM SOFTWARE

The research team developed software for this project that automates many of the algorithms and processes discussed above. The software, used in conjunction with a GIS package,

- identifies origins and destinations and creates an O/D matrix
- flags errors and incomplete and external trips

- categorizes trips as loop, local, or access
- creates a range of trip to trip performance measures, and
- calculates sample size confidence statistics.

To facilitate the OD identification process, the research team developed an automatic program named “OD Detector” to implement the OD algorithms. The team also developed "OD Generator" to calculate the performance measures necessary to automate the freight performance measure statistics and generate data to track truck movements between TAZs. The roles of these two programs in the database processing flow are shown in Figure 1. Appendix 7 contains the software’s input screens.

PROGRAM COSTS AND RECOMMENDATIONS

The cost of an ongoing GPS data program can be estimated. On the basis of discussions with several vendors, as well as the data collection costs for this effort, the typical cost of data from one truck for one month is between $0.80 and $1.20. Different coverage options and costs are estimated for a State of Washington GPS program in Table 5. The network coverage relates both to the geographic area of the GPS data and to the number of data points (truck reads) that can be used to calculate performance measures.

Table 5. GPS Program Cost and Coverage

Program Size	Estimated Monthly Program Cost	Coverage	Network Coverage
2,500 trucks (1 vendor)	$3,000	Puget Sound region	Major interstates and other high volumes roads during peak travel time, travel between major trip generators
6,000 trucks (1 vendor)	$6,000	Statewide	Major interstates and other high volumes roads during peak travel, travel between major trip generators, major cross state interstates,
12,000 trucks (2 vendors)	$10,000	Statewide	Major cross state interstates, urban interstates and high level urban roadway during peak travel time, travel between major trip generators
16,000 trucks (3 or more vendors)	$14,000	Statewide	Major cross state interstates, urban interstates and high level and medium urban roadway, travel between major trip generators for larger cities in Washington state, travel performance across state borders

In addition to data collection costs, there are ongoing database maintenance and processing costs. These costs will vary, depending on the geographic coverage of the database information and the size of databases, but can be expected to run between $2,000 to $6,000 per month. Active use of the GPS data will require more in-depth studies to fully understand their statistical usefulness for freight performance measurement. Such an analysis will need to evaluate spot speeds on both highways and freight arterials.

The GPS data sets, as used, have some limitations that potentially can be addressed in a future freight performance measures program. The main limitation is that the data set with the best geographic coverage (vendor A) only has location reads, when trucks are moving, every 15 minutes. This limited read frequency is partially countered by the larger number of vendor A's trucks that travel the Puget Sound study every day. However, this rate still limits the ability to look at roadway performance on lower level or smaller volume streets. This limitation can be reduced both by acquiring data from more trucks (both as the vendor's client database grows and by buying data from multiple vendors) and as the GPS equipment becomes more capable and costs drop, resulting in more frequent location reads. Given the size of the database, there is a need to avoid any manual data cleaning and to automate many of the geocoding and processing steps. With this automation, inevitably some "good" data will be discarded, so having a large number of data points will help to mitigate any data loss.

A second limitation is simply that little is known about each probe vehicle. Classification of the vehicles is important since, for example, two-axle trucks potentially require different performance measures than do tractor trailer combinations. Fortunately, contacts at vendor A have indicated that truck classification is possible with additional software development. It is recommended that future data acquisition agreements with vendor A and other GPS vendors request truck classification information.

Of the various performance measures that have been developed, the research team recommends that x^{th} *Percentile Travel Time* be considered as a primary statewide performance measure. This measure is a relatively easy to understand indicator of reliability (a 95th percentile travel time reliability means that 19 out of 20 trips are on time) but can also be modified to account for different segments of the trucking industry. For example, businesses such as those that haul fresh fish may require a 95th percentile reliability, whereas those that carry scrap paper may only require a 60th percentile reliability.

CONCLUSIONS

This research project used data from GPS devices installed in trucks for fleet management purposes to build the foundation for a Washington state freight performance measure program. After meeting technical obstacles in acquiring GPS data from individual trucking companies, the researchers found that purchasing GPS data and services from three fleet management vendors worked well. While the cost to acquire the data from vendors and privacy restrictions limited some uses, this relationship had a number of advantages, including access to an ongoing data stream, good technical support, and access to reads from a large number of trucks that travel the regional network. The main disadvantage is that commercially available GPS data are collected for truck fleet management and are not designed for a public sector performance measure program. As result, the data sets included a large number of individual trucks but less frequent GPS location reads than necessary to quickly evaluate urban network travel times.

Processing the resulting raw GPS data into a form that could be used for a performance monitoring program required a number of steps. Because each vendor's dataset was large and organized differently, the team developed an automated process to acquire and format the raw GPS data into a working database to locate truck travel patterns on the roadway network and to analyze the trucks' travel times and speeds. Because identifying the travel patterns of truck trips is an important aspect of monitoring truck performance, a rule-based algorithm that automatically separates traffic-related truck stops from origin and destination stops was developed. The algorithm also flags abnormal trips so they can be accounted for when performance measurement statistics are developed. This O/D algorithm was validated and modified by using network maps and aerial photos in GIS software and by making comparisons to network travel statistics used for regional travel models.

Because some of the GPS data included spot speeds for trucks, this effort also evaluated the value of these speeds for their usefulness in examining roadway segment performance both to locate recurring congestion (bottlenecks) and to quantify the impacts of construction projects. The results of several case studies in which these speeds were compared with freeway speed data from loops suggested that spot speeds have value but also some limitations, especially when truck travel performance is evaluated over a short periods of time and on short segments of roadway. The research indicated that spot speeds are best averaged over longer periods or on roadway segments traveled by large numbers of the GPS-equipped trucks. The future use of spot speeds will require more statistically meaningful studies to fully understand their usefulness for freight performance measurement.

The GPS data can also be used to monitor regional transportation network performance between economically important areas or zones. The advantage of GPS data is that they can capture the routes trucks use to travel between zones. A number of different quantitative performance measures can be calculated for regional travel.

On the basis of the research findings, the research team developed a multi-step processing program. This program identifies trucks' origins and destinations, flags errors due to GPS signal problems or external trips, categorizes trips by several travel categories, calculates a range of zone-to-zone performance measures, and calculates sample size confidence statistics.

The development of the GPS data platform, as well as testing of the data in several case studies, indicated that with enough data, the truck GPS probes can help to effectively measure congestion and highway network performance. The benefit of an ongoing truck freight performance measures program lies in its ability to quantify truck travel characteristics over long periods of time.

REFERENCES

American Transportation Research Institute (2005) "Methods of Travel Time Measurement in Freight-Significant Corridors." Submitted to the Transportation Research Board Annual meeting January, http://www.atri-online.org/research/results/Freight%20Performance%20Measures%20TRB%20for%20atri-online.pdf.

Chue, R., C. Xie and D. Lee (2002) "Probe Vehicle Population and Sample Size for Arterial Speed Estimation", *Computer-Aided Civil and Infrastructure Engineering,* 17 pp. 53-60.

Czerniak, R. (2002) *Collecting, Processing, and Integrating GPS Data into GIS,* NCHRP 301, Transportation Research Board, Washington, D.C.

Du J. and L. Aultman-Hall (2007) "Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues", *Transportation Research Part A*, 41, pp. 220-232.

Fitzgerld, C., J. Zhang and P.R. Stopher (2006) "Processing GPS Data from Travel Surveys." *Proceedings - International Global Navigation Satellite systems Society (IGNSS) Symposium* (CD-ROM). Gold Coast, Australia.

Greaves, S. P. and M. Figliozzi (2008) "Collecting Commercial Vehicle Tour Data with Passive Global Positioning System and Technology." *Transportation Research Record: Journal of the Transportation Research Board, No. 2049,* TRB, National Research Council, Washington, D.C., pp. 158-166.

Hallenbeck, M., E. McCormack, J. Nee, and D. Wright (2003) "Freight Data from Intelligent Transportation System Devices." WA-RD 566.1, Washington State Department of Transportation.

Hunter M., S. Wu, and H. Kim (2008) "Practical Procedure to Collect Arterial Travel Time Data Using GPS-Instrumented Test Vehicles." *Transportation Research Record Journal of the Transportation Research Board, 1978,* TRB, National Research Council, Washington, D.C., pp. 160-168.

IBI Group (2009) "Final Report: Continental Gateway Road Network Performance." Transport Canada, February.

Ishimaru, J. and M., Hallenbeck (1999) "Flow Evaluation Design Technical Report." Technical Report for Washington State Department of Transportation, WA-RD 466.2, March.

Lee, S., B. Lee, and Y. Yang (2006) "Estimation of Link Speed Using Pattern Classification of GPS Probe Car Data." *Proceeding - International Conference on Computational Science and its Applications*, pp. 495-204.

Li, S., K. Zhu, B. van Gelder, J. Nagle, and C. Tuttle (2002) "Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices." *Transportation Research Record, Journal of the Transportation Research Board, 1804*, Paper No. 02-2129.

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, No. 1957*, TRB, National Research Council, Washington, D.C. pp. 43-50.

Mei, C. and S. 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,* 1719, Paper No. 00-1334.

Nezamuddin, J., L. Crunkleton, P. J. Tarnoff, S. E. Young (2009) "Statistical Patterns of Traffic Data and Sample Size Estimation." http://www.catt.umd.edu/documents/variance_analysis_v1.pdf (Accessed December 12 2009)

Quiroga, C. and D. Bullock (1998a) "Travel time studies with global positioning and geographic information systems: an integrated methodology", *Transportation Research Part C*, 6, pp. 101-127.

Quiroga, C. and D. Bullock, (1998b) "Determination of sample sizes for travel time studies." *ITE Journal*, 68(8), pp. 92-88.

Sananmongkhonchai, S., P. Tangamchit and P. Pongpaibool (2008) "Road Traffic Estimation from Multiple GPS Data Using Incremental Weighted Update." In: Proc. of the 8th International Conference on ITS Telecommunications, pp. 62–66.

Schuseeler, N. and K. W. Axhausen (2008) "Processing GPS Raw Data without Additional Information", Preprint CD-ROM for the 88th Annual Meeting of Transportation Research Board, Washington, D.C.

Short, J., R. Pickett and J. Christianson (2009), "Freight Performance Measures Analysis of 30 Freight Bottlenecks", The American Transportation Research Institute, Arlington, VA, March.

APPENDIX 1: DATA DICTIONARY

Vendor A

Column	Description
SEQ_ID	Intended to be the row ID initially, but the value is Null
DEVICE_ID	Unique Truck ID
DATA_TYPE	There are four types, moving, park, other, maintenance
DATA_DESC	More detailed description to DATA_TYPE
SPEED	GPS spot device speed
SPEED_UOM	The unit of speed, KPH
DIRECTION	The direction of truck
DURUTION	How long the truck has stopped
LATITUDE	Latitude of the truck
LONGITUDE	Longitude of the truck
GPS_STATUS	Can be 0 or 1. 0 means good GPS signal and 1 means bad GPS signal , thus the lat/long is the last known good GPS and not the true location
LOCATION_TIMESTAMP	The time of the event from the real-time clock chip.
TIMEZONE	Time zone for the truck
DST	Daylight saving time offset, i.e. if the device is configured to observe DST or not.
STATUS	Can be 0, 1, 2, or 3. This is the stop duration where 1 means "short" stop, 2 means "medium" stop, and 3 means "long" stop. These meanings are configurable per customers/vehicles. The reason for this is that if the vehicle is coming up to a stop light, you would not consider it as a stop unless it is not moving for over 2 minutes. The default definition of a "short" stop is over two minutes but less than a "medium" stop. Other customer such as long-haul trucking, they may have a different definition of what a stop is because they are mostly moving. In their case, a "short" stop may be half an hour because it takes over half an hour just to fuel up. 0 obviously means the vehicle is moving.
MILEAGE	The distance shown in the truck's Odometer
QUEUE_TIME	The time when the truck waits in the queue
OBSERATION_TIME	The time from the GPS satellite at the time of the event (or GPS fix). In the case where there is no GPS, then the observation time will be the last known GPS time lock.

Vendor B

Column	Description
alias	Alternate equipment name or identifier
auxID	Qualcomm-assigned auxiliary customer account identifier of this transaction's publisher
companyID	Company identifier of this transaction's publisher
deviceID	Equipment on-board device identifier
equipmentID	Unique customer-defined equipment identifier
equipType	A value indicating the type of equipment. Values include "trailer" or "tractor"
eventTS	Event timestamp expressed in the following format in GMT timezone: yyyy-mm-ddThh:mm:ssZ
VendorBID	No Definition
ignitionStatus	The ignition status of the tractor, where 1=On, 2=Off.
lat	Latitude expressed in signed degrees with floating-point decimal precision.
lon	Longitude expressed in signed degrees with floating-point decimal precision.
mobileType	A value indicating the type of mobile communication device. Values include: 0 = Unknown 1 = MCT 2 = TMCT 3 = OmniOne 4 = MCP 10 = UTT 11 = Tethered
posTS	Position GMT timestamp expressed in the following format: yyyy-mm-ddThh:mm:ssZ
posType	Identifies the hardware that was used to determine the position information contained in the transaction, where: 0=Unknown, 1=LORAN, 2=QASPR, 3=GPS
VIN	OEM Vehicle Identification Number

Vendor C

Column	Description
ID	Row ID, can be the primary key
Vehicle_ID	Unique Truck ID
Date	The time of the event from the GPS Device
Latitude	Latitude of the truck
Longitude	Longitude of the truck
Distance	The truck's travel distance , -1 means the truck parks
Odometer	The mileage shown in the Odometer in the truck
GPSSpeed	GPS device speed
ECMSpeed	Truck engine speed
Name	Location name, address
StoreID	Store name
IsDepot	True means the location is the depot, the truck will load or unload. False means it is not the depot.
IsGeoCode	Whether the truck location can be geo-coded
VehicleName	The name of truck

APPENDIX 2: DATABASE FEEDS

The database feed from each GPS vendor is different. ATRI emails a CSV file, a flat file with comma-separated values. The research team has written a PHP script to parse the data in this file and load it into a database table. Vendor A provided the install file for a JavaScript client, which also loads data into a single database table. This client only requires the connection information for the MySQL database, and now it automatically retrieves live data.

Vendors B and C each provided a sample program written in C#, both of which required considerable modification before they would retrieve data properly. Vendor C's program retrieves data from its Web service in a compressed format and saves this directly to a ZIP file. It was desirable to decompress the data stream in the C# program and then use C# to load the data into MySQL. The database administrator was unable to find and install the proper decompression classes into Visual Studio so instead wrote a script that uses the C# program to download a ZIP file, invokes another program to decompress the ZIP (which contains a CSV file), then runs yet another script to open the CSV file, parse it, and insert it into a single database table. Vendor B's C# application downloads data from its Web service in an XML object. The administrator decided to use C# to parse the XML and write the data out as a CSV file, which is then parsed and inserted into a single database table in the same way as vendor C's data.

Our largest dataset receives about 3 million rows of data each month. Having such a large amount of data in a single table makes processing very time consuming. Also, each dataset provides special information. This special, extended information varies from vendor to vendor. To speed up processing, at the end of every month the project team creates two new tables for each dataset. These tables contain only data for the previous month. The data are split into "core" values, which we need for our processing, and "extended" values, which comprise everything else we are provided. These tables are easy to join together to do processing that utilizes data points in the extended table; however, to speed processing time we keep all our active data points in the core tables.

For origin and destination (OD) processing, the project team runs a script directly on these MySQL database tables, which have been split into month-long datasets. For GIS processing, we then export the core table for each month into CSV files. These CSV files are easier to work with because they leave off data points we are not able to use in GIS (those stored in the extended tables). These CSV files also contain O/D trip information we generated. When

GIS processing has been completed, transportation analysis zone and other information is written into a new CSV file, which is then parsed and used to update the database to contain the GIS information.

APPENDIX 3: GEOCODING

For the I-90 case study, the research team geocoded the Geographic Positioning System (GPS) dataset from vendor A to identify trucks traveling on the I-90 bridge and the approach to the bridge. While the geocoding process was automatic within the geographic information systems (GIS) package, the resulting output required data cleaning to ensure that trucks on frontage roads, overpasses, and other data anomalies were excluded. An additional complication was that there were no reads from trucks within the tunnels on either end of the I-90 Bridge because of the timing of the connection to the GPS satellites. After the appropriate days and points on the I-90 segment had been isolated, there were 7,373 data points within the seven-week study period. Appropriate data for the analysis were selected from a larger set of GPS responses on the basis of the following GIS selection steps:

1. I-90 was given a 100-foot horizontal buffer from its midpoint to identify all vendor A's data on I-90.

2. Points were selected and grouped from the buffer on the 1.5-mile bridge segment and 1.5 miles east and 1.5 miles west of the bridge.

3. Data were then grouped into eastbound and westbound categories on the basis of a heading variable calculated by the GPS device (a value of between 0 and 360 degrees). If this direction variable was illogical (e.g., heading due north on an east/west road segment), it was excluded. Orthographic photos were used to determine the precise locations of points near freeway overpasses or parallel roadways.

4. Outliers included data with a 382-degree value (on a 0- to 360-degree scale) for direction. These values were due to poor GPS satellite signal or, in a few cases, trucks that had been stopped or parked for a significant period and then began moving, producing an inconclusive direction value. Because the speeds for these data were always zero, the average speeds for the roadway may have been skewed and the direction of the truck unknown. Further analysis revealed that many of the 382-degree direction values were recorded by the same trucks multiple times in the same day. These may indicate that some GPS devises may have malfunctioned regularly during the study period.

5. Data with a 0-degree value for direction were included in the study. These values were given when a truck was idle for a period of time and the GPS device was unable to record

6. Data from the express lanes were included only if they were associated with an appropriate travel heading during non-construction weeks, as the express lane directions are reversible.

After the data had been cleaned, 4,513 data points (truck positions) were successfully assigned to I-90 (a 61 percent rate).

APPENDIX 4: I-90 CASE STUDY RESULTS

The findings from the analysis of the May 2009 closure of the I-90 express lanes are summarized below.

1) **Overall, not as many GPS-equipped trucks traveled in the AM peak as in other time periods (tables A4-1 and A4-2).** There were only 206 AM peak (6:00 AM to 9:00 AM) responses during the entire five-week study period out of 6,595 responses. That represents about 3 percent of all responses in contrast to the 13 percent of total hours per day represented by the three hours of the AM peak. This indicates that GPS-equipped trucks were traveling much less during the morning peak than during other times of the day. This may be a function of the type of truck used in the study. For example, if the trucks were used for regular deliveries, the drivers may have driven before the morning rush hour to avoid traffic and unload during the am peak. Additional study of other vendors that have different types of trucking services may provide more AM peak responses. With such a low volume of GPS reads, statistically valid analysis of truck performance on I-90 during the AM peak was not possible.

2) **The daily GPS-equipped truck volumes increased on I-90 westbound and I-90 eastbound during the construction period (tables A4-1 and A4-2).** The numbers of Vendor A GPS-equipped trucks reported on I-90 during the four weeks before construction were 49 per day westbound and 54.3 per day eastbound. During the three weeks of construction, the numbers went up to 59 per day westbound and climbed to 69 per day eastbound. The week after construction the numbers increased to 64 per day westbound and decreased to 67 eastbound. These small changes in daily volumes suggest that construction had a minimal impact on GPS-equipped truckers' choice of routes during construction, but it is important to note that truck volumes may have changed for a variety of reasons. For example, if there was construction-related congestion, the truck counts may have gone up, and trucks may have been counted multiple times in a road segment in the same trip. Or if there was no construction, trucks may have chosen the route, increasing volumes. Overall truck volumes on this route and on alternative routes should be analyzed more closely in comparison to regular traffic patterns and along with spot speeds to better determine highway performance.

3) **Average daily truck speeds on I-90 decreased slightly overall during construction (tables A4-1 and A4-2).** The four weeks before construction the average truck speeds were 52 mph westbound and 55 mph eastbound. During construction they fell slightly to 47 mph westbound and 52 mph eastbound. The four weeks after construction they increased to 51 mph westbound and 54 mph eastbound. This suggests that average truck speeds were affected during the construction period but that these impacts were minimal. These overall construction speed decreases were statistically significant (at a 95 percent confidence interval based on two-tailed t-tests).

4) To determine detailed impacts, the data had to be analyzed independently by time of day and using smaller road segments to determine where and when the most significant truck travel time impacts occurred. Below are statistically significant findings (at a 99 percent confidence interval based on two-tailed t-tests) from analysis of the data by time of day and road segment:

   a) During construction, westbound truck speeds dropped below pre-construction truck speeds by about 8 mph during afternoon peak periods and about 7 mph during midday time periods (Figure A4-1).

   b) Eastbound construction speeds dropped below pre-construction truck speeds by 8 mph during evening time periods (Figure A4-2).

   c) Westbound truck speeds east of the I-90 Bridge dropped below pre-construction truck speeds by about 12 mph during the first two weeks of construction (Figure A4-3).

   d) Eastbound truck speeds west of the I-90 Bridge dropped below pre-construction truck speeds by about 20 mph during the first two weeks of construction (Figure A4-4).

   e) Heading both eastbound and westbound on all road segments, truck speeds increased to near average speeds during the third week of construction (figures A4-3 and A4s-4).

Table A4-1. I-90 Westbound Volumes and Average Truck Speeds by Time Period

Time Period	**Four Weeks Before Construction**		**Three Weeks During Construction**		**Four Weeks After Construction**	
	Number	Avg. Speed (MPH)	Number	Avg. Speed (MPH)	Number	Avg. Speed (MPH)
Morning Peak	32	57.5	33	58.1	40	58.2
Midday	116	55	126	48.4	162	58.1
Afternoon Peak	283	52.7	279	44.6	323	55.6
Evening	542	50.8	446	47.4	751	45.4
Total	**973**	**52**	**884**	**47.1**	**1276**	**50.5**

Table A4-2. I-90 Eastbound Volumes and Average Truck Speeds by Time Period

Time Period	**Four Weeks Before Construction**		**Three Weeks During Construction**		**Four Weeks After Construction**	
	Number	Avg. Speed (MPH)	Number	Avg. Speed (MPH)	Number	Avg. Speed (MPH)
Morning Peak	31	58.5	38	58.5	32	56.3
Midday	259	56.5	219	56.9	324	56.7
Afternoon Peak	389	55.4	318	54.6	495	54.6
Evening	407	54.3	457	46.3	493	52.2
Total	**1086**	**55.3**	**1032**	**51.5**	**1344**	**54.2**

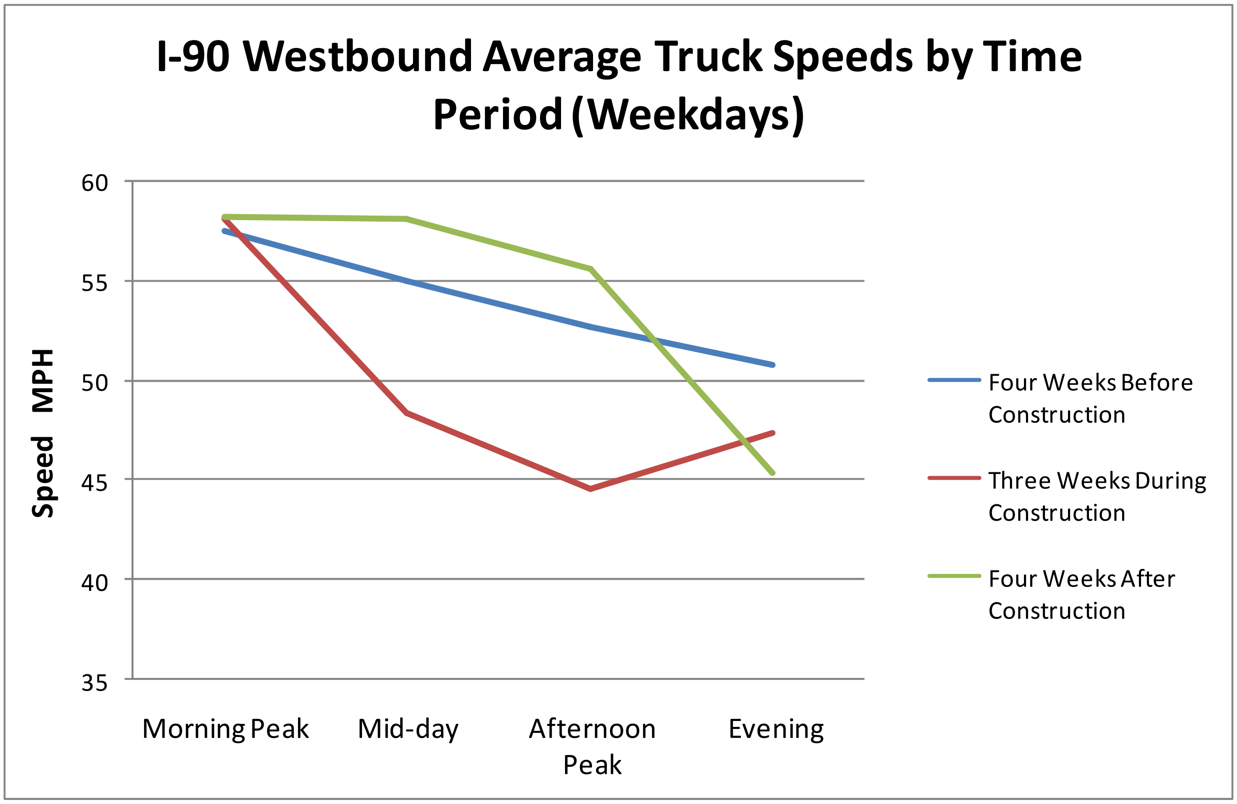

Figure A4-1: Westbound Truck Speed Comparison by Time Period

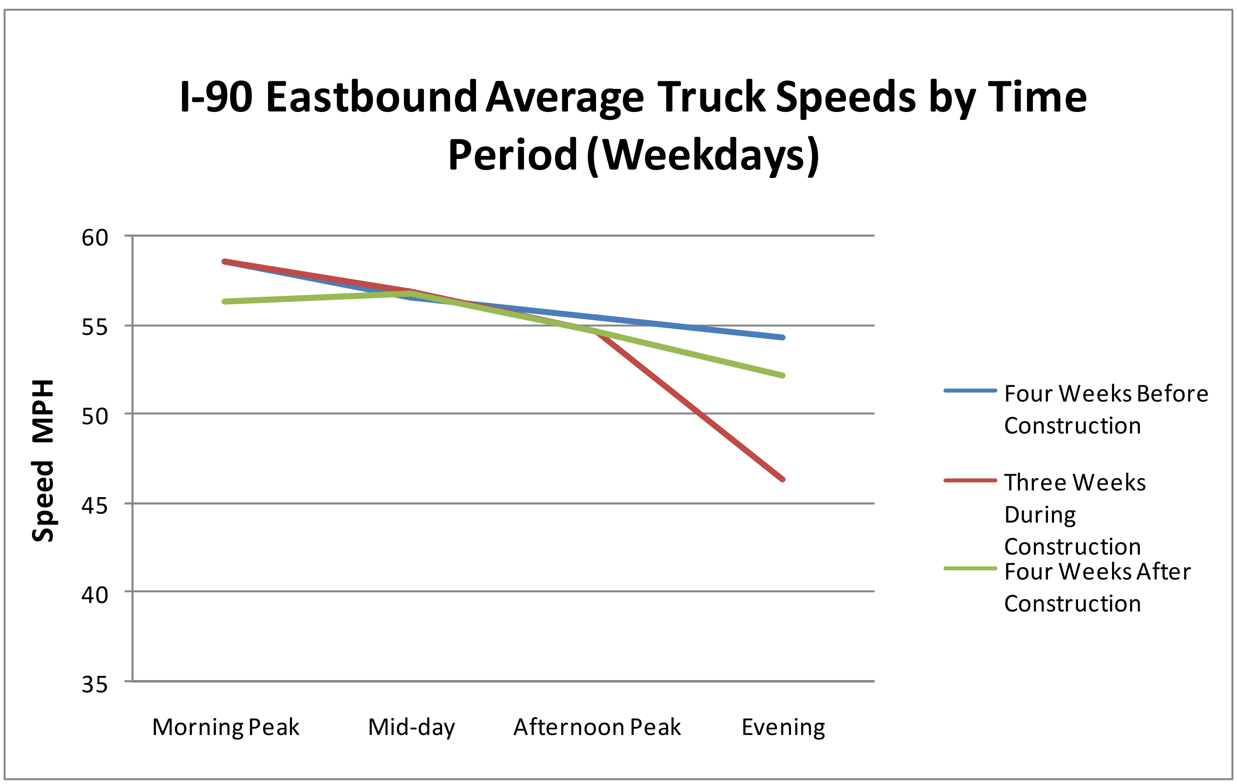

Figure A4-2: Eastbound Truck Speed Comparison by Time Period

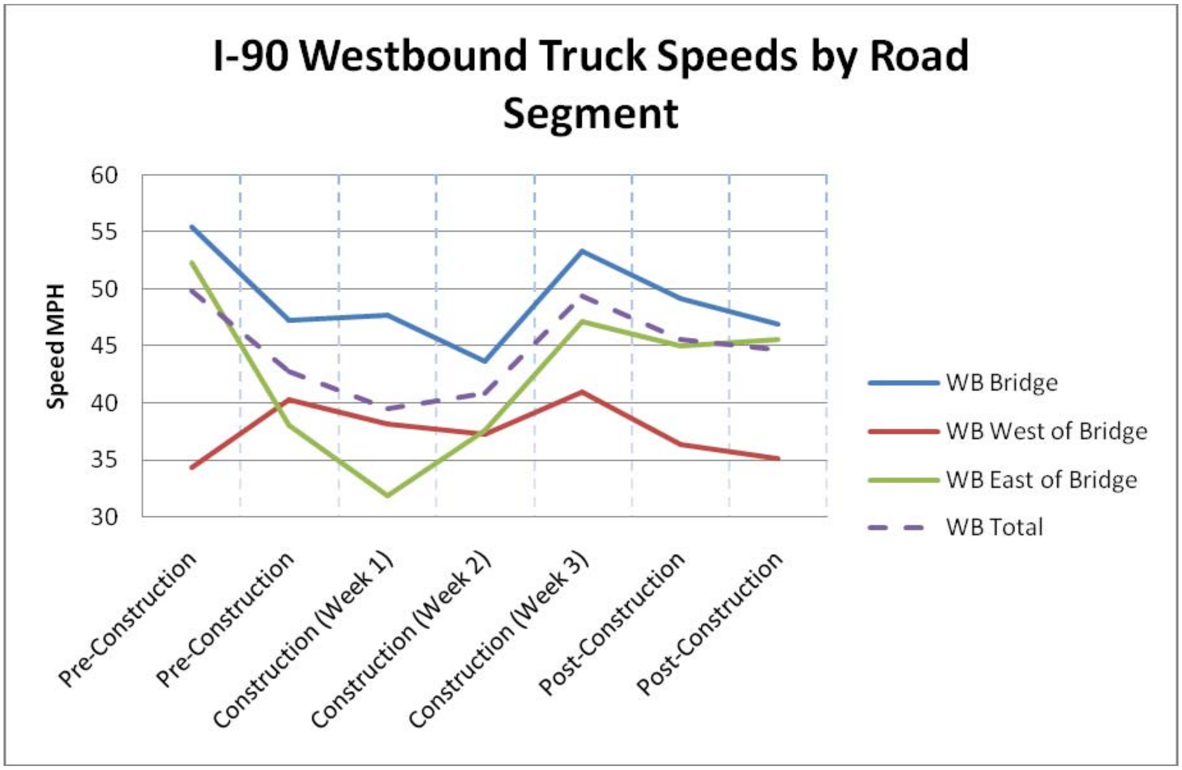

Figure A4-3: Westbound Truck Speed Comparison by Road Segment

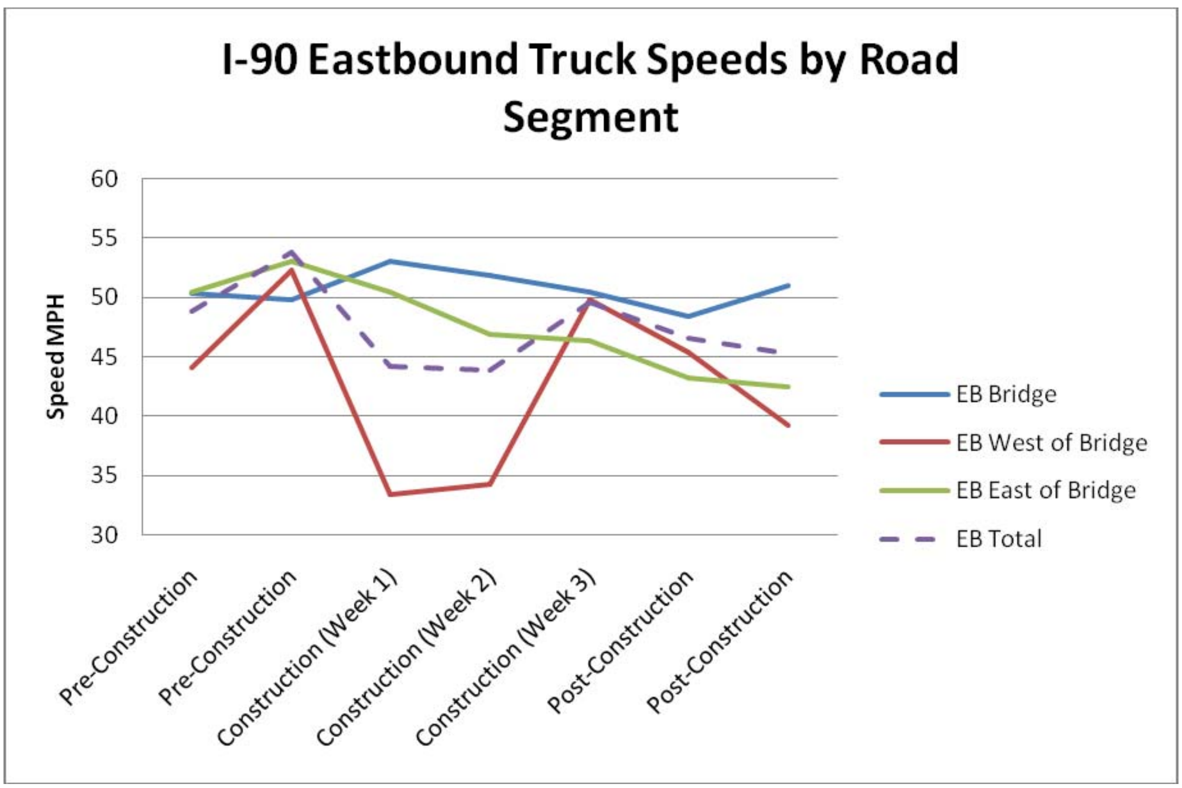

Figure A4-4: Eastbound Truck Speed Comparison by Road Segment

I-90 Roadwork Impacts on Truck Traffic April 20-24: Two Weeks before Construction

Westbound
Total Volume: 287 Mean Speed: 49.8 St. Dev: 18.5

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation
Bridge	Morning	7	58.3	3.6
	Midday	14	43.9	24.5
	Evening	43	58.3	6.6
	Night	73	55.8	10.9
Total	-	137	55.5	12.5
West of Bridge	Morning	0	0	0
	Midday	8	42.4	18.6
	Evening	19	50.1	7.1
	Night	37	24.5	23.3
Total	-	64	34.3	22.6
East of Bridge	Morning	4	58.5	1.1
	Midday	9	58.1	5.8
	Evening	27	56.1	11.3
	Night	46	48.5	20.1
Total	-	86	52.3	16.7

Eastbound
Total Volume: 316 Mean Speed: 48.8 St. Dev: 19.7

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation
Bridge	Morning	4	58.8	1.5
	Midday	32	40.8	24.8
	Evening	55	59	5.5
	Night	68	47.2	22.5
Total	-	159	50.3	20
West of Bridge	Morning	0	0	0
	Midday	22	46.1	16.6
	Evening	42	42.4	20.9
	Night	34	45	21
Total	-	98	44.1	20.1
East of Bridge	Morning	0	0	0
	Midday	13	44.8	24.8
	Evening	20	57.2	6.5
	Night	26	48.1	19.5
Total	-	59	50.5	18.5

I-90 Roadwork Impacts on Truck Traffic April 27-May 1: One Week before Construction

Westbound

Total Volume: 445 Mean Speed: 42.7 St. Dev: 22.9

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	7	58.6	2.5	59.2
	Midday	25	44.6	23.1	59.4
	Evening	59	47.1	20.3	42.1
	Night	116	47.4	19.1	59.3
Total	-	207	47.3	19.8	
West of Bridge	Morning	0	0	0	59.1
	Midday	10	51.9	12	58.5
	Evening	16	45.7	19.4	46.7
	Night	35	34.5	33.8	59.9
Total	-	61	40.3	30.1	
East of Bridge	Morning	1	58	0	59.1
	Midday	18	59.9	4.6	59.9
	Evening	82	29.1	27.4	35.3
	Night	76	42.1	23.8	59.3
Total	-	177	38	26.2	

Eastbound

Total Volume: 418 Mean Speed: 53.8 St. Dev: 13.5

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	5	58.6	5.1	59.4
	Midday	59	49.4	21	60.0
	Evening	61	58.5	4.7	57.9
	Night	68	56.9	5.2	60.0
Total	-	225	49.8	29.5	
West of Bridge	Morning	1	58.4	0	54.4
	Midday	30	46	21.5	58.6
	Evening	58	53.9	7.9	55.1
	Night	48	54.3	6.2	59.1
Total	-	137	52.3	12.4	
East of Bridge	Morning	0	0	0	58.1
	Midday	10	42.1	27.6	58.6
	Evening	20	58.3	6.4	52.7
	Night	26	53.2	16.1	59.9
Total	-	56	53	17.4	

I-90 Roadwork Impacts on Truck Traffic May 4-8: The First Week of Construction

Westbound
Total Volume: 365 Mean Speed: 39.5 St. Dev: 21.7

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	9	57	2.4	54.4
	Midday	19	47.3	15.4	59.2
	Evening	52	47.4	18.3	54.0
	Night	75	46.8	18.6	60.0
Total	-	155	47.7	17.7	
West of Bridge	Morning	0	0	0	59.3
	Midday	6	32.5	23.3	59.0
	Evening	9	51.1	4.1	49.9
	Night	39	35.9	18.7	60.0
Total	-	54	38.1	18.7	
East of Bridge	Morning	0	0	0	35.2
	Midday	25	29.8	25.4	55.7
	Evening	67	29.6	21.8	34.1
	Night	64	35.1	23.8	59.9
Total	-	156	31.9	23.4	

Eastbound
Total Volume: 409 Mean Speed: 44.2 St. Dev: 22

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	10	59	2.8	58.7
	Midday	44	46.4	24	60.0
	Evening	51	57.9	4	58.0
	Night	78	52.8	13.6	60.0
Total	-	183	53	15.6	
West of Bridge	Morning	2	57.5	.5	53.7
	Midday	40	46.1	19.7	58.0
	Evening	32	48.2	13.2	35.6
	Night	103	23.5	23.2	59.4
Total	-	177	33.4	23.9	
East of Bridge	Morning	0	0	0	56.1
	Midday	20	49.2	21.2	58.4
	Evening	16	53.7	11.2	56.1
	Night	13	48.2	19.1	59.8
Total	-	49	50.4	18.1	

I-90 Roadwork Impacts on Truck Traffic May 11-15: The Second Week of Construction

Westbound

Total Volume: 292 Mean Speed: 40.9 St. Dev: 22.7

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	10	56.6	5.1	52.5
	Midday	29	42.3	22.5	58.2
	Evening	59	34	25	51.0
	Night	67	50.7	16.8	59.6
Total	-	165	43.6	22.2	
West of Bridge	Morning	1	55.9	0	59.2
	Midday	4	53.5	5.7	58.3
	Evening	6	44	5.1	50.7
	Night	27	32.6	24.6	59.9
Total	-	38	37.2	22.3	
East of Bridge	Morning	2	58.4	1.5	38.0
	Midday	20	29.6	23.8	56.9
	Evening	36	29.8	22.5	42.0
	Night	31	50.7	16.7	59.4
Total	-	89	37.6	23.2	

Eastbound

Total Volume: 376 Mean Speed: 43.9 St. Dev: 23.1

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	8	59.2	3.8	56.3
	Midday	45	47.5	22.4	59.0
	Evening	48	48.8	20.7	56.9
	Night	64	56	9.3	60.0
Total	-	165	51.8	17.7	
West of Bridge	Morning	4	56.4	3.6	53.2
	Midday	26	44.6	20.8	57.4
	Evening	40	49	19.1	35.0
	Night	83	23	22.9	58.5
Total	-	153	34.3	24.7	
East of Bridge	Morning	1	60.3	0	55.8
	Midday	13	46.3	26.1	58.1
	Evening	25	49.7	20.3	54.1
	Night	19	42.9	24.5	60.0
Total	-	58	46.9	23.2	

I-90 Roadwork Impacts on Truck Traffic May 18-22: The Third Week of Construction

Westbound
Total Volume: 309 Mean Speed: 49.4 St. Dev: 18.4

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	9	60.7	2.5	59.5
	Midday	31	55	14.9	59.6
	Evening	32	55.6	14.8	46.3
	Night	93	51.2	16.3	59.7
Total	-	165	53.3	15.6	
West of Bridge	Morning	0	0	0	59.8
	Midday	2	55.6	2.5	58.3
	Evening	11	48.7	4.3	48.7
	Night	39	38.1	18.7	59.9
Total	-	52	41	17.2	
East of Bridge	Morning	2	59.5	.5	59.5
	Midday	8	53.6	20.6	60.0
	Evening	35	54.1	16.9	34.5
	Night	47	40.2	23	59.5
Total	-	92	47.1	21.6	

Eastbound
Total Volume: 377 Mean Speed: 49.6 St. Dev: 19.5

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	6	61.3	5.6	59.8
	Midday	42	40.5	28.8	60.0
	Evening	62	54.5	14.8	58.3
	Night	81	51.9	17.1	60.0
Total	-	191	50.5	20.3	
West of Bridge	Morning	7	55.9	4.1	57.9
	Midday	23	55.6	6.2	59.4
	Evening	42	49.5	17.2	52.2
	Night	49	46.6	18.6	58.9
Total	-	121	49.8	16.2	
East of Bridge	Morning	0	0	0	57.9
	Midday	12	58.3	4.2	59.4
	Evening	17	59	4.8	52.2
	Night	36	36.4	25.5	58.9
Total	-	65	46.3	22.2	

I-90 Roadwork Impacts on Truck Traffic May 25-29: One Week after Construction

Westbound

Total Volume: 311 Mean Speed: 45.6 St. Dev: 21.3

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	7	60.0	1.9	59.4
	Midday	19	60.4	3.6	59.8
	Evening	40	46.2	23.5	48.0
	Night	80	47.1	20	59.8
Total	-	146	49.2	20	
West of Bridge	Morning	1	55	0	58.0
	Midday	8	49.8	15.3	58.2
	Evening	8	46.1	6.5	50.3
	Night	32	30	21.1	60.0
Total	-	49	36.3	20.3	
East of Bridge	Morning	2	59.0	1	59.3
	Midday	9	60.6	3.6	59.7
	Evening	37	45.6	23.7	40.0
	Night	68	42.1	21.7	59.7
Total	-	116	45	22	

Eastbound

Total Volume: 312 Mean Speed: 46.6 St. Dev: 20.8

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation	FLOW Avg. Speed
Bridge	Morning	2	61.5	3.5	59.5
	Midday	59	40.3	27	60.0
	Evening	52	55.6	14.3	53.1
	Night	56	49.9	14.6	60.0
Total	-	169	48.4	20.7	
West of Bridge	Morning	0	0.0	0	57.2
	Midday	28	41.7	19.9	58.5
	Evening	24	50.4	16.2	49.7
	Night	32	44.8	16.7	58.1
Total	-	84	45.4	18	
East of Bridge	Morning	0	0	0	57.2
	Midday	26	39.8	24.7	58.5
	Evening	8	60.3	4.5	49.7
	Night	25	41.2	24.7	58.1
Total	-	59	43.2	32.5	

I-90 Roadwork Impacts on Truck Traffic June1-5: Two Weeks after Construction

Westbound
Total Volume: 401 Mean Speed: 44.7 St. Dev: 22.6

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation
Bridge	Morning	10	59.4	1.9
	Midday	38	49.2	23.3
	Evening	59	46.5	22.1
	Night	109	45.2	21.6
Total	-	216	46.9	21.8
West of Bridge	Morning	0	0	0
	Midday	9	51.4	13.5
	Evening	17	32.5	24.3
	Night	37	32.3	21.8
Total	-	63	35.1	22.6
East of Bridge	Morning	1	58	0
	Midday	14	61.3	3.5
	Evening	31	52.7	17.7
	Night	76	39.2	24.7
Total	-	122	45.6	22.9

Eastbound
Total Volume: 400 Mean Speed: 45.3 St. Dev: 22.2

Road Segment	Time Period	Number	Avg. Speed	Std. Deviation
Bridge	Morning	5	23.4	26.1
	Midday	31	53.5	14.8
	Evening	65	55	14.8
	Night	85	48.7	19.5
Total	-	186	51	18.3
West of Bridge	Morning	5	59	1.1
	Midday	25	52.6	12.3
	Evening	66	32.2	24.6
	Night	44	39.7	23.4
Total	-	140	39.2	23.5
East of Bridge	Morning	0	0	0
	Midday	16	48.4	23.3
	Evening	31	41.4	26.7
	Night	27	40.3	22
Total	-	74	42.5	24.5

APPENDIX 5: SR 167 CASE STUDY RESULTS

These data are for weekdays from September 24, 2008, to July 31, 2009, for the SR 167 case study.

Northbound Totals

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	1490	58	6.6	N/A
Midday	19465	47.2	15.9	N/A
PM Peak	21788	50.4	15.4	N/A
Evening	15957	56.2	11.7	N/A
Night	5215	56.3	11.6	N/A
Total	63855	51.6	14.8	N/A

Northbound North of SR 516 Segment

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	774	57	7.2	46.9
Midday	8073	53.2	12.8	50.9
PM Peak	9021	50.9	5.8	54.1
Evening	6608	53.3	15.2	59.2
Night	2596	53.8	14.4	58.3
Total	27072	52.7	14.5	54.5

Northbound South of SR 516 North of SR18 Segment

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	186	58	5.6	35.9
Midday	3780	44.3	17.9	48.3
PM Peak	4378	49	16.3	55.5
Evening	3483	57.9	7.3	58.8
Night	1033	58.3	5.8	56.8
Total	12860	50.9	15.3	52.2

Northbound South of SR18 Segment

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	470	59	5.8	N/A
Midday	7612	42.3	15.7	N/A
PM Peak	8389	50.7	14.7	N/A
Evening	5866	58.6	8	N/A
Night	1586	59.2	7.7	N/A
Total	23923	50.7	14.8	N/A

Southbound Totals

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	2319	53.6	13.8	N/A
Midday	9623	58.8	7.2	N/A
PM Peak	23603	53.6	7	N/A
Evening	36914	48.2	16.2	N/A
Night	15341	44	18.7	N/A
Total	87801	51.4	15.1	N/A

Southbound North of SR 516 Segment

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	996	47	18.4	57.1
Midday	2812	57.7	9.1	51.5
PM Peak	7150	56.3	8.2	49.7
Evening	10474	52.4	12.2	59.7
Night	4760	47.1	17.2	59.9
Total	26192	52.9	12.9	55.7

Southbound South of SR 516 North of SR 18 Segment

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	638	59	4	58.9
Midday	3163	58.9	6.9	53.7
PM Peak	7735	58.7	7	44.6
Evening	12659	45.6	18.2	59.9
Night	4726	43	20.2	60.0
Total	28921	50.4	16.7	55.7

Southbound South of SR 18 Segment

Time Period	Number	Average Speed	Standard Dev.	Flow Speeds
AM Peak	685	58	6.2	N/A
Midday	3648	59.6	5.3	N/A
PM Peak	8718	59.2	5.6	N/A
Evening	13781	47.3	16.2	N/A
Night	5855	42.2	18.3	N/A
Total	32687	51.2	15.1	N/A

APPENDIX 6: SAMPLE SIZE CALCULATIONS

With the GPS probe data, the travel time between transportation analysis zones (TAZs) fluctuates more than the travel speed because of to the variability of the TAZ's area. In comparison to travel time, link speed is independent of link length and can be measured easily and objectively.

Cheu et al., (2002) and Mei and Chien (2000) investigated probe GPS vehicle population and sample size for speed estimation and provided an equation to calculate the sample size as follows:

$$n = \left(\frac{t_{\alpha/2,n-1} s}{\varepsilon_r \bar{x}}\right)^2 \quad (1)$$

where

$\alpha =$ significance level

$t_{\alpha/2,n-1} = t$ value from two-tailed t distribution with $n-1$ degrees of freedom for a confidence level of $1-\alpha$

$\bar{x} =$ mean travel speed

$\varepsilon_r =$ user-selected allowable relative error in the estimate of the mean speed

$s =$ sample speed standard deviation

However, equation (1) is not closed form, and an iterative procedure has to be applied because the t-statistic is dependent on sample size.

Nezamuddin et al. (2009) and Li et al. (2002) demonstrated in their sample size study that if a large sample is available, the z-statistic can be used instead of the t-statistic, which requires a one-step calculation, and the equation can be written as the follows:

$$n = \left(\frac{z_{\alpha/2} s}{\varepsilon_r \bar{x}}\right)^2 \quad (2)$$

Here, $z_{\alpha/2}$ is the z-statistic for a given confidence level, which doesn't rely on sample size. However, it should be kept in mind that equation (1) remains the most reliable way of calculating sample size and is preferred whenever a reliable estimate of standard deviation can be obtained.

To simplify the calculation, we used equation (2) to estimate the sample size because of the large number of GPS trucks in our database. There is one caution to applying this sample size estimation equation. The above equations are based on the assumption that the speed of vehicles (or travel time) follows a normal distribution. Mei and Chien (2000) found that other factors can affect the distribution, including the roadway geometrics, as well as traffic volumes on the link. For example, in semi-congested or congested conditions, the travel speed may not follow a normal distribution, and the above estimation equation may not be applicable. However, the above equations are widely used and are expected to give a reliable sample size estimate (Quiroga and Bullock, 1998).

The researchers calculated the minimum sample size by using the Kent Valley data between TAZ 340 and TAZ 385 as an example. Using a relative error of ε_r =10 percent, which resulted in $\varepsilon_r \overline{x}$ of 3.07 mph, they then applied equation (2) in Appendix 6 to calculate the minimum sample size as 16, which implied that the average travel speed can be determined within 10 percent for a confidence level of 95 percent.

References

Chue, R., C. Xie and D. Lee (2002)"Probe Vehicle Population and Sample Size for Arterial Speed Estimation", *Computer-Aided Civil and Infrastructure Engineering*, 17 53-60.

Li, S., K. Zhu, B. van Gelder, J. Nagle, and C. Tuttle (2002) "Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices", *Transportation Research Record* 1804, Paper No. 02-2129.

Mei, C. and S. Chien (2000) "Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation", *Transportation Research Record* 1719, Paper No. 00-1334.

Nezamuddin, J., L. Crunkleton, P. J. Tarnoff, S. E. Young (2009) "Statistical Patterns of Traffic Data and Sample Size Estimation", http://www.catt.umd.edu/documents/variance_analysis_v1.pdf (2009-12-22 Accessed)

Quiroga, C. and Bullock, D. (1998) "Determination of sample sizes for travel time studies", *ITE Journal*, 68(8), pp. 92-88.

APPENDIX 7: PROGRAM SOFTWARE

The research team developed a series of programs that automated many of the algorithms and processes required to clean the GPS data and develop performance measures. These programs' software, used in conjunction with GIS software,

- identifies truck origins and destination; this information is used to create a TAZ-based O/D matrix
- flags errors, uncompleted and external trips
- categorizes trips as loop, local, or access trips
- creates a number of zone to zone performance measures, and
- calculates sample size confidence statistics.

To complete the O/D identification process, an automatic program named "Trip Detector" was developed to implement the OD algorithm. The research team also developed a program "OD Generator," which automates the calculation of freight performance measure statistics and generates data to track the truck travel between TAZs.

Figures A7-1 and A7-2 show a snapshot of the O/D Detector and the O/D Generator's input screens. The insert O/D Matrix button can flag errors, label uncompleted and external trips, and categorize trips between origins and destinations as loop, local, or access trips. The generate zonal performance measures button can create a range of trip to trip performance measures, and calculates sample size confidence statistics. In addition, by assigning origin and destination TAZs, the O/D Generator software can extract detailed GPS records for the further analysis in either Excel or CSV format.

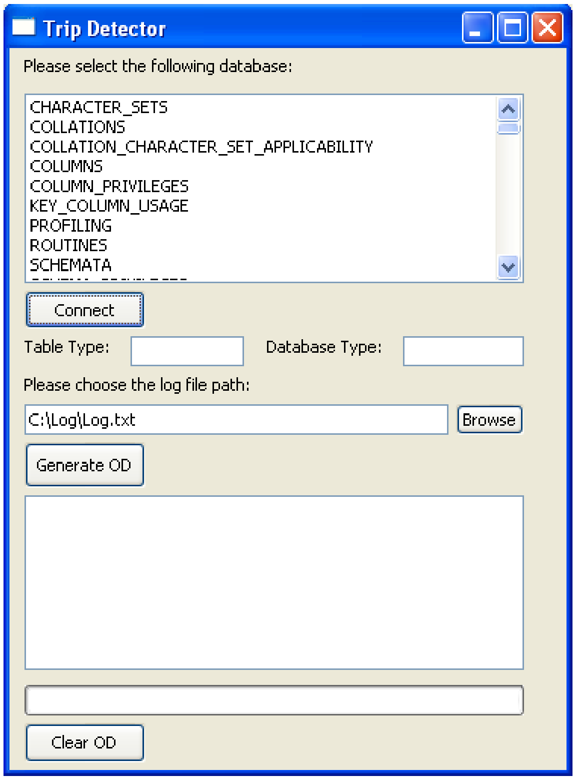

Figure A7-1: A Snapshot of the Trip Detector Input Screen

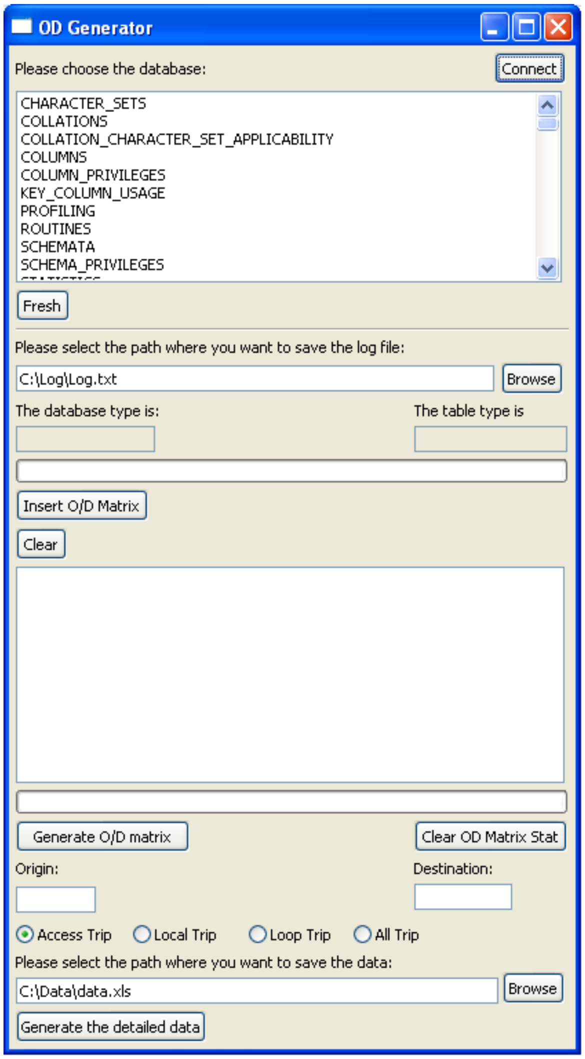

Figure A7-2: A Snapshot of the O/D Generator Input Screen