Processing Commercial GPS Data to Develop a Web-Based Truck Performance Measures Program

Xiaolei Ma
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Washington, Box 354802
Seattle, WA 98105
Tel: (518) 618-6775
Email: xiaolm@u.washington.edu

Edward McCormack, Ph.D. (Corresponding Author)
Research Assistant Professor
Department of Civil and Environmental Engineering, Box 354802
University of Washington
Seattle, WA 98195, 2700
Tel: (206) 543-3348
Fax: (206) 685-0767
Email: edm@u.washington.edu

Yinhai Wang, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering, Box 352700
University of Washington
Seattle, WA 98195-2700
Tel: (206) 616-2696
Fax: (206) 543-1543
Email: yinhai@u.washington.edu

Word count: 6429 + (1 Table+ 4 Figures) × 250 =7679Submitted for Publication in Transportation Research Record: Journal of the Transportation Research Board

Submitted on August, 1st, 2010
Resubmitted on November 14th, 2010
Final Submitted on March 15th, 2011

ABSTRACT
Although trucks move larger volumes of goods than other modes of transportation, public agencies know little about their travel patterns and how the roadway network performs for trucks. Trucking companies use global positioning systems (GPS) from commercial vendors to dispatch and track their equipment. This research collected these GPS data from approximately 2,500 trucks in the Puget Sound region and evaluated the feasibility of processing these data to support a statewide network performance measures program. The program monitors truck travel time and system reliability, and will guide freight investment decisions by public agencies.

While other studies have used a limited number of project-specific GPS devices to collect frequent location reads, permitting a fine-grained analysis of specific roadway segments, this study used data that involved less frequent reads but were collected from a larger number of trucks for over a year. The research team used automated processing to clean and format the data because they included millions of data points. Because a performance measurement program ultimately monitors trips generated by trucks as they travel between origins and destinations, an algorithm was developed to extract this information and geocode each truck's location to the roadway network and to traffic analysis zones. Measures were developed to quantify the truck travel characteristics and performance between zones. To simplify the process and provide a better communications platform for the analysis, the researchers developed a Google Maps-based online system to compute the measures and graphically show the trucks' routes.

INTRODUCTION

The commercial global positioning systems (GPS) used by trucking companies for fleet management have only recently been explored as a source of truck performance data for public agencies. Although these systems are the most comprehensive national source of truck travel data available for quantifying truck performance, their data cannot be directly used to calculate truck performance measures, and considerable processing is required. Problems include limitations in terms of reading frequency, format, and quality. Adequately processing these probe data and further developing useful truck performance information generation and visualization tools have proved to be ongoing challenges for researchers (*1*).

The Washington State Department of Transportation (WSDOT), Transportion Northwest at the University of Washington, and the Washington Trucking Associations partnered on a research effort to collect and analyze GPS truck data from commercial in-vehicle fleet management systems used in the central Puget Sound region. This GPS information, which is in essence a by-product of trucking industry operations, was collected and evaluated for its usefulness in supporting a long-term, truck-based freight network performance monitoring program. WSDOT intends to use this program to guide planning and construction decisions. The data, from GPS devices installed for the convenience of the trucking industry, were collected infrequently but involved a relatively large number of trucks over a long period of time. In this effort's study area, data were collected from approximately 2,500 trucks per day.

This paper discusses the steps required to process these data, address data limitations, and develop usable performance measures. Processing the data required acquiring it from diverse sources and developing software that could handle extremely large databases. Bad data due to GPS signal loss or inaccuracy had to be removed, and the data had to be effectively assigned (geocoded) to the Puget Sound area roadway network.

A crucial element of the project was developing a method to identify truck travel patterns, since the success of a truck-oriented performance measures program depends on extracting useful truck travel time and speed, roadway location, and stop location information. A performance measures program ultimately monitors travel generated by trucks as they respond to shippers' business needs, picking up goods at origins (O) and dropping off goods at destinations (D). Recognizing the importance of individual truck O-D pairs, the authors developed an algorithm to extract the pairs from the GPS reads that formed the basis for the performance measures.

Once each truck's geocoded location was mapped to the Puget Sound region's roadway network and then assigned to traffic analysis zones (TAZs), truck freight performance measures that focused on zonal travel times and speeds could then be calculated. These measures, analyzed over time, would determine a freight system's reliability. A Google Maps-based truck performance system was implemented to provide a user-friendly and interactive interface for agency staff to use in accessing the GPS data and monitoring trucking movements in the Puget Sound region.

LITERATURE REVIEW

Several previous efforts have used GPS data to measure truck performance. A recent study by the American Transportation Research Institute (ATRI) concluded that GPS data from trucks can be confidentially processed to provide average travel rates along major long distance U.S. corridors (*2*). The data used for the ATRI study were purchased from GPS vendors, and the spot speeds from GPS devices were aggregated over time to identify truck bottlenecks. McCormack and Hallenbeck (*3*) tested the use of truck GPS data for measuring freight movements along specific roadway corridors in Washington state. The study, which used 25 portable data collection GPS loggers with 5-second location reads, concluded that GPS data can provide an indication of roadway performance. Greaves and Figliozzi (*4*) described algorithms for processing 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. The Ontario Ministry of Transportation has an ongoing arrangement with a private vendor to use GPS data from its commercial fleet management systems. This vendor provides GPS data from approximately 30,000 trucks each weekday (*5* and *6*). The IBI group utilized the GPS data from the same vendor to monitor truck freight performance at Ontario and Québec border crossings (*7*). They developed a series of metrics to examine freight system performance.

A number of researchers have documented the processing of GPS data and the retrieval of trip information to detect O-D 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. They documented the data collection, reduction, and reporting procedures used to produce measures of effectiveness for travel times at various levels of resolution, including the roadway system, corridor, and local road levels (*8*). Du and Aultiman-Hall (*9*) developed an automatic trip end identification algorithm that uses a combination of maximum and minimum dwell times, heading changes, and a check for distance between the GPS points. This process was used to increase the accuracy of trip rate information. Schuessler and Axhausen (*10*) described a post-processing procedure for cleaning and smoothing raw GPS data and identified trip activity and trip modes automatically by using fuzzy logic. Hunter et al. (*11*) used GPS-instrumented test vehicles and developed an algorithm that identified the traversal time between intersections for a GPS device mounted in a probe vehicle to calculate travel times on urban arterial streets. Zheng et al. (*12*) used GPS to monitor transit vehicle movements along signalized arterial corridors to evaluate a transit signal priority system deployed in the Snohomish County in Washington state.

A number of these research efforts relied on GPS devices with high reporting frequencies but small numbers of probe vehicles to identify trip ends and calculate travel times. While these studies typically applied some level of automated processing to clean and organize the data, the post-processing also included manual processes to fix or remove data with problems.

DATA ACQUISITION

The first step of this study was to collect truck GPS data. Obtaining the data directly from trucking companies was initially attempted but proved to be unworkable. While most companies agreed to share their data, it was difficult to work out the technical details with the companies' in-house data staff (if they existed) for transferring the data. The authors reconsidered and determined that it would be easier to obtain data from the GPS device vendors. This approach had a number of advantages:

- The vendors had technical staff experienced in sending out GPS data.
- Each vendor was able to provide GPS data from multiple trucking 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 for the benefit of the trucking companies and might not be ideal for the freight performance measures program.

- The data presented an ongoing cost.
- Because the vendors were not accustomed to selling data to a university, each relationship required a sometimes protracted contracting process.
- Because protecting the privacy of clients for the vendor's trucking company was critical, non-disclosure agreements were required, which involved attorneys and slowed the project.

Data acquisition contracts were signed with three GPS vendors. Each vendor collected and managed truck GPS data differently. An overview of the data acquired from each vendor is shown in Table 1. Common to each data set were variables that included longitude, latitude, truck ID (hashed for privacy in some cases), travel heading, and a date and time stamp. These three vendors jointly provided GPS data from approximately 2,500 trucks traveling in the Puget Sound region each day.

DATA STORAGE AND PROCESSING

To manage and analyze the truck GPS data, an efficient mechanism to receive, clean, store, and manipulate the GPS data on a database server was essential. Because of the diversity and large volumes of the GPS data, a scalable database with flexible data formats needed to be created. This required relational tables storing different vendors' GPS data, which could be combined to jointly supply the data needed for truck O-D identification, GIS processing, and performance measurement. The truck GPS data processing flow chart is shown in Figure 1.

Database Development

An initial challenge with the data was that each vendor's data used different formats and communication methods. Data from vendors A and B were periodically obtained via File Transfer Protocol (FTP) services, and data from vendor C were acquired manually by email every month. The research team developed a database system that played a crucial role in acquiring and archiving these heterogeneous GPS data. For both vendors A and B, a scheduled fetch program was run at the back end to connect with the vendors' FTP servers and retrieve and upload the data into databases at one-month intervals. For vendor C, a separate program was written to periodically check data availability. When data were received, this program automatically imported them into databases. Along with these programs, the researchers developed a data feed and archiving service to place each vendor's data in a separate database.

Because more than 100,000 location reads were downloaded into the database each day, querying and analyzing such a large and growing database as part of a performance measures program could be cumbersome. To optimize the database, a program was developed to automatically split the three vendors' databases into three monthly data sets. Common information from each vendor's data set was formatted into standard columns, which contained only the fields necessary for a freight performance measures development program. This reduced data processing time. After several programming iterations, it became possible to query nearly 10 million rows of data within a second.

Truck Location Mapping

The trucks' GPS locations (determined on the basis of the latitude and longitude reported by the GPS) needed to be geo-coded onto the road network for analysis by geographic information system (GIS) software. However, the accuracy of the GIS in mapping geocoded location data to a road network typically suffered from errors caused by spatial mismatches between the base network and the latitude and longitude points, as well as other problems such as assignment confusion related to overpasses and frontage roads. These issues typically required post-processing and error checking. Because of the large size of the dataset, manual checking for errors was impossible. A computer program was developed to automatically identify suspect points for further checking or elimination. For data quality control, a set of rules was developed and implemented in the computer program, which was also designed to use a GIS scripting language to assign each truck's GPS reads to the correct roadway segments.

A section of Interstate-90 through the Seattle area served as a case study to develop and test the geo-coding process. GPS data from vendor A on I-90 from May 4 to May 23, 2009, were retrieved and processed in the computer program. Use of a 100-foot (30-meter) buffer from the road midpoint helped in assigning each geocoded location read to the nearest roadway. Heading data for the trucks were then checked to associate the GPS travel bearing with the road segment's bearing. A 15-degree heading range was selected because this reasonably captured trips along the roadway while excluding truck travel on overpasses and intersecting roads. The process proved effective in mapping trucks' geocoded locations to roadways, with about 60 percent of the points passing the screening process.

ORIGINS AND DESTINATIONS IDENTIFICATION

Developing a performance measures program requires an understanding of the travel patterns of trucks as they use the roadway network. This requires an efficient and accurate process to identify each location where a truck stops to complete the productive transaction that defines the purpose of each truck's trip. Because GPS devices record a range of stops, a methodology was needed to differentiate between traffic-based (unintended) stops at intersections or in congestion and intended stops that corresponded to a truck's origin and destination. 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) (*4,9,10,11*). In these studies, the O-Ds identified by stop duration were also manually examined to identify any unusual situations or problems.

Because of the large size of the GPS data sets in this study, an automatic and efficient O-D identification algorithm was required. The research team initially developed a dwell time threshold to identify trucks' origins and destinations. Determining the dwell time setting was crucial, as a threshold setting that was too long or too short could either miss or incorrectly segment trips. An appropriate dwell time depends on traffic conditions in a given area (*9*). McCormack and Hallenbeck's (*3*) truck GPS research completed previously in Washington state determined 3 minutes to be a reasonable dwell threshold. The 3-minute period filtered 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 is stationary for more than 3 minutes is unusual in the Puget Sound region. In some cases, a 3-minute stop might include a driver's break or fueling stop, but these stops were considered part of a truck's daily travel pattern.

However, an observation of outliers and examination of travel between specific TAZs, revealed that a number of short stops of less than 3 minutes were not recorded as destinations but should have been. Further analysis indicated that a number of trucks traveled significant distances to a different TAZ, made a short stop, and then continued to another TAZ for a final destination stop. This problem was addressed by taking advantage of the fact that the largest database's (vendor A) GPS data feeds included engine park and stop information. A stop was therefore defined as either a truck with its engine off or in park status with the engine on. In this case, any parked stop when the engine remained on could also be considered an origin or destination. The algorithm was consequently revised to include all of these "parked" stops. Moreover, the revised O-D identification algorithm included programming that successfully resolved issues of signal loss, signal jiggle, and abnormal trips.

Signal Loss

GPS signals may be blocked when overhead obstructions such as tall buildings prevent GPS devices from communicating effectively with the GPS satellites (*4, 13*). In this study, this resulted in a few lost O-D data pairs. 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 the short blockage period, it is reasonable to assume that a truck continues to travel at a constant speed, as the GPS points recorded before and after the signal loss can be used to calculate the average speed. For this study, a threshold speed limit of 5 mph was selected. If the average calculated speed was below this threshold, the program determined that a trip had ended in the area of signal loss.

Signal Jiggle

In many cases, when a series of GPS points for a truck is mapped in a GIS system, the points may fluctuate around a position when a truck idles (known as jiggle), creating a false report of movement (*14*). This is 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 was less than 0.000051 degrees (around 65 feet or 20 meters), a flag was tagged to the record marking it as a possible destination point. When the average speed for this trip was calculated, the delay time caused by the fluctuations was subtracted to achieve a more accurate result.

Abnormal Trips

GPS data often contain abnormal trips caused by vehicles leaving an area or GPS multipath effects, in which the GPS signals are reflected off the surface of objects located between the GPS satellites and the GPS device (*13, 14*). In an effort to deal with these data problems, the O-D identification algorithm was also programmed to detect abnormal trips. These trips were flagged. This secondary process detected the following:

- extremely short trips between GPS reads for an individual truck
- false trips in which the elapsed truck travel time was zero
- trips with extremely high speed
- trips in which a truck's O or D was external to the study area and could not be captured by the O/D algorithm
- trips in which a truck left the study area and then returned.

These flagged trips could be removed from any truck performance program calculations that required travel time and travel speed. Figure 2 illustrates the O-D identification algorithm as implemented in Java.

The O-D algorithm was validated by using one month of vendor A's data. This data set contained 3 million GPS records and filled about 1 GB of file space. The O-D algorithm ran on an Intel Pentium Xeron processor and required about an hour to complete; it resulted in 358,692 O-D trips, with 6,443 abnormal trips flagged during the second-round trip processing. To validate the accuracy of the fundamental assumptions behind the O-D algorithm, several truck IDs were randomly selected from the database, and their entire daily trips were geo-coded onto a road network in GIS software. Each truck's trip was followed from stop to stop along the road network, and stops were explored by using Google Earth, Google Maps, and aerial photos. This process looked at the O-D locations detected by the algorithm to determine whether the trucks' estimated travel corresponded to locations with truck terminals, warehouses, or other reasonable freight delivery and pick-up locations. In this test, the O-D algorithm was successful.

TRIP TYPE CLASSIFICATION

Different types of trucks have different trip characteristics. This distinction is useful for targeting performance measures. Distance statistics calculated from the roadway network of the Puget Sound Regional Council's (PSRC) travel demand model were used to distinguish among different truck trip types. The PSRC data provide trucks' optimal travel time and distance between zones, and an important assumption is that most truckers travel an economically efficient (optimal) route between an origin and destination. On the basis of this assumption, the average travel distance and travel time for all GPS trucks between each pair of zones were compared to the PSRC's travel distance for the same pair of zones.

This process revealed that a few trips had notably longer GPS distances than the PSRC network distances. These truck trips remained internal to the study area, did not include any stops longer than 3 minutes, and did not contain instances in which the truck was placed into a "parked" status. Investigation indicated that a few GPS trucks made deliveries without the need to park. One example was a trash container company that delivered large empty containers to a construction site. Because the containers 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 a parked or stopped status.

This comparison between the average trip travel distance and PSRC network distance indicated that the truck trips could be classified into several general categories:

- *Access Trips:* Trucks had clearly defined origins and destinations, and each stop could be defined by a park event or a stop longer than 3 minutes. This was the most common trip type (53 percent of all trips).
- *Local Trips.* If the average trip travel distance was smaller than 0.5 mile, then a trip was identified as a local trip. These were trips during which the driver dropped off packages without stopping longer than 3 minutes. These trucks, typically small package delivery trucks, were involved in about 38 percent of all trips.
- *Loop Trips.* If the average trip travel distance was at least two times longer than the PSRC network distance, a trip was identified as a loop trip. These trucks completed their business without the need to stop or place the truck into park. Only a few trucks fit this category (9 percent), and these were typically garbage or construction trucks.

DEVELOPING TRUCK PERFORMANCE MEASURES

The GPS truck data were collected to quantify truck travel in order to develop performance measures. Assigning the truck trip O-D data to traffic analysis zones makes it possible to quantify regional travel between these zones, and this can provide an effective platform for evaluating regional truck travel. The advantage of examining zone-to-zone travel with GPS data is that this process 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 should also be the ones that are represented by the largest numbers of O-D pairs.

A number of quantitative performance measures were calculated for zone-to-zone travel. Because the GPS data provided, at a basic level, travel distance and speed for the probe trucks, all the measures were derived from these numbers and were often compared to ideal or free flow travel conditions. The literature survey completed as part of this research, as well as studies by others (*7*, *15*), suggested a number of performance metrics that can be applied to the GPS data to analyze truck performance between zones.

Free Flow Speed and Nominal Travel Time

Free flow speed and nominal travel time performance measures are defined as the truck travel speed and time under uncongested traffic conditions. In this research, the calculation of free flow truck speed was derived from the PSRC truck freight analysis model. The free-flow travel time excluded intra-zonal travel to the origin and destination TAZs, so free-flow travel time was calculated on the basis of centroid-to-centroid (zone center to zone center) travel time during the night (from 11:00 PM to 5:00 AM), when roadways were likely to be uncongested (*16*). The free-flow travel speed can be derived by dividing the centroid-to-centroid distance by the free-flow travel time.

Average Travel Time

Average travel time (ATT) represents the averaged trucks' travel time (arithmetic mean) between one origin and destination pair. The ATT depends on trucks' route choices and the traffic conditions. For a single trip, a travel time performance measure was calculated by taking the difference between the origin timestamp and the destination timestamp recorded by the GPS device and comparing that to an average of all trips with the same origin and destination.

Average Travel Speed

For each trip between one O-D pair, trucks normally do not travel the same distance; therefore, a large variance in travel time may be observed. To provide a more accurate and descriptive measure, average travel speed (ATS) is used. This measure was calculated as the arithmetic mean of all trip speeds. Each trip speed was calculated by dividing the distance by the travel time between each O-D pair. The distance between each O-D pair was precisely recorded by the GPS device.

Average Travel Distance

Average travel distance is the arithmetic mean of travel distances for all the trips. For most trips, travel distance can be read directly from the odometer recorded by the GPS device mounted in each truck.

Variability of Travel Time (VTT) and Variability of Travel Speed (VTS)

These two measures determine the reliability of the freight movement for all trips between each zonal O-D pair. They are the ratios of the standard deviation of travel time (travel speed) to the average travel time (travel speed) aggregated over a sample of road sections.

x^{th} Percentile Travel Time

For this research, 95th percentile travel time was adopted.

Travel Time Index (TTI)

Travel Time Index is the ratio of the average travel time to the travel time under free-flow conditions.

Buffer Time Index (BTI)

Buffer Time Index is the ratio of the 95th percentile travel time to the average travel time. This index indicates how much additional time is needed to ensure an on-time arrival with 95 percent probability.

Planning Time Index (PTI)

Planning Time Index is the ratio of the 95th percentile travel time to the nominal travel time (corresponding to free-flow speeds), which can be thought of as the 95th percentile Travel Time Index. For example, a Planning Time Index of 1.6 means that for 95 percent of all trips to arrive on time, truckers should allow a travel time that is 1.6 times longer than the nominal travel time.

Minimum GPS Sample Size

To estimate link travel time or speed, it was necessary to determine the minimum required GPS truck sample size. This information was used to develop statistical significance performance measures. With the GPS probe data, the travel time between TAZs fluctuated more than the travel speed because of the variability in TAZ area size. In comparison to travel time, link speed was independent of link length and could be measured easily and objectively. Cheu et al., (*17*) and Chen and Chien (*18*) investigated probe GPS vehicle population and sample size for speed estimation and provided an equation to calculate the sample size as follows:

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

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

Nezamuddin et al. (*19*) and Li et al. (*20*) 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:

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

Here, $z_{\alpha/2}$ is the z-statistic for a given confidence level 1-α, which does not rely on sample size. However, note 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, the authors used Equation 2 to estimate the sample size because of the large number of GPS trucks in the database. One caution to applying this sample size estimation equation is that both equations are based on the assumption that the speed of vehicles (or travel time) follows a normal distribution. Chen and Chien (*18*) found that other factors can affect the distribution, including roadway geometrics and traffic volumes on the link. For example, in congested conditions, the travel speed may not follow a normal distribution, and the above estimation equations may not be applicable. However, the above equations are widely used and are expected to provide reliable sample size estimates (*21*).

ON-LINE TRUCK PERFORMANCE MEASURES SYSTEM IMPLEMENTATION

Once a series of performance metrics had been developed, an online system was designed to make the measures readily usable by staff from WSDOT and other public agencies. A Web-based freight module was developed from the basis of an existing data interoperation and analysis systems platform known as Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net). DRIVE Net is a community-based program developed by the Smart Transportation Application and Research Laboratory (STAR) Lab at the University of Washington (*22*).

Figure 3 shows the online truck freight performance measures system. The GPS data server stored the raw data received from the GPS vendors. After the O-D identification program processed these data, they were automatically imported into the online database. This program detected each trip end, calculated the travel time and speed, and completed data quality checks. Zonal freight travel time, speed, and distance information used by the PSRC travel demand model was collected in the PSRC database module, which was used as the road network reference for developing truck performance measures. The PSRC mapping data were also stored in the TAZ GIS database module for visualization in Google Maps. On the server side, the truck performance measure module calculated statistics by using data from the truck GPS and PSRC databases. The calculated performance measure metrics were transferred to the client side through a Google Web Toolkit (GWT) module, which is a development tool for building and optimizing complex browser-based applications (*23*). The inputs received from users were sent to the JavaScript Interpreter to display in Google Maps or were further processed and calculated on the server side.

The resulting online system allows users to easily calculate performance measures between zones. By specifying the origin and destination of interest in the online systems, users are able to track the trucks' movements and generate performance metrics, such as numbers of trips, average travel time and speed, and Travel Time Index. An example of the output is shown in Figure 4. That screen capture shows truck performance measures calculated between a warehouse district and the Port of Seattle and displayed along the left side of the screen. When users click a marker on the map, a small window pops up (also shown in Figure 4) with a brief description of the recorded GPS point, including device ID, spot speed, time stamp, moving status, mileage, and direction. Users can track any one truck's movement by clicking a Show GPS Route button, which will display the routing information for an individual truck. The sample size was based on a 95 percent confidence level and incorporated into the on-line system. This allows users to determine the data reliability and accuracy for a selected zone pair.

A zone-based network O-D matrix tool is also available in this system. By specifying the date, users can generate a variety of O-D matrices based on total trip numbers, average travel time, and average travel distance, and output a CSV file designed for use by modelers.

CONCLUSIONS

This research developed a series of processing steps and algorithms to apply to data collected from commercial GPS devices installed in trucks. These efforts established a foundation for a Washington state freight performance measures program. The main challenge with the commercially available GPS data was that they were collected for truck fleet management business needs and not designed for a public sector performance measures program. As result, the data sets included a large number of individual trucks but collected GPS location reads less frequently than has been previously done to evaluate some urban network travel times. This research demonstrated the feasibility of obtaining commercial GPS truck data from relatively low cost sources and that these data can be processed to create usable truck performance measures designed for public agencies.

Acquiring the data was initially a problematic, but the researchers found that purchasing GPS data and services directly from GPS fleet management vendors worked well.

Processing the raw GPS data into a form that could be used for a performance monitoring program required a number of steps. Each vendor's data set was large and organized differently, and this required developing an automated process to acquire and format the raw GPS data into a working database that would allow locating truck travel patterns on the roadway network and analyzed their 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 distinguished operational truck stops from origin and destination stops was developed. The algorithm also flagged abnormal trips so they could be accounted for when performance measurement statistics were developed. This O-D algorithm was validated and modified by using network maps and aerial photos in GIS software and by comparing their results to network travel statistics used for a regional travel model.

The GPS data were then capable of being used to monitor transportation network performance between economically important areas or zones. The advantage of GPS data is that they can capture the routes used by trucks to travel between zones. A Web-based performance measures system was designed so that public agency staff can use the GPS data to evaluate truck performance. The system is based in Google Maps and provides a user-friendly interface for visualizing the GPS data and generating zonal truck performance measure metrics. In addition, this system can be used to track the trucks' movements and their routing choices.

Although the Web-based truck performance measures platform has resulted in a usable foundation for a statewide performance monitoring program, future improvements and system expansions are still under way. The current data stream will be expanded to include new GPS vendors and statewide data. This will require the development of more robust and versatile data processing and quality control steps. With these additional data, segment-based truck performance measures may be possible and these will require new processing steps.

Because of privacy concerns, some aspects of the Web-based truck performance program will not be made available to end users. A version that outputs aggregated data will be designed for agency staff and public users. Further research will be completed to investigate the utility of this system to transportation professionals.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support and funding from the Washington State Department of Transportation, Transportation Northwest at the University of Washington, and the Washington Trucking Associations.

REFERENCES

1. Figliozzi, M.A. and Tufte, K.A., A Prototype for Freight Data Integration and Visualization Using Online Mapping Software: Issue, Applications, and Implications for Data Collection Procedures. Preprint CD-ROM for the 88th Annual Meeting of Transportation Research Board, Washington, D.C. 2009.
2. American Transportation Research Institute, "Methods of Travel Time Measurement in Freight-Significant Corridors", http://www.atrionline.org/research/results/Freight%20Performance%20Measures%20TRB%20for%20atri-online.pdf, Accessed: Jul. 06, 2010.
3. McCormack, E. and Hallenbeck, M. H., ITS Devices used to Collect Truck Data for Performance Benchmarks, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1957,* TRB, National Research Council, Washington, D.C., 2006, pp. 43-50.
4. Greaves, S. P. and Figliozzi, M. A., Collecting Commercial Vehicle Tour Data with Passive Global Positioning System and Technology, In *Transportation Research Record: Journal of the Transportation Research Board, No. 2049,* TRB, National Research Council, Washington, D.C., 2008, pp. 158-166.
5. Tardiff, R. Using Operational Truck Location Data to Improve the Understanding of Freight Flows, North American Freight Transportation Data Workshop, Transportation Research Record. May 15, 2007,
6. Update on DMAO GPS Data Tool – Planning and Assessment Applications, Policy and Planning Division, Transportation Planning Branch, Province of Ontario, PowerPoint file, Apr. 2007.
7. IBI Group, *Congestion Measurement and Assessment Framework- A Canadian Perspective*, Transport Canada, final Report, Jun. 2007.
8. Quiroga, C. and Bullock, D. Travel time studies with global positioning and geographic information systems: an integrated methodology, *Transportation Research Part C*, 6, 1998, pp. 101-127.
9. Du, J. and Aultman-Hall, L., Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues, *Transportation Research Part A*, 41. 2007, pp. 220-232.
10. Schuseeler, N., and Axhausen, K. W., Processing GPS Raw Data without Additional Information, Preprint CD-ROM for the 88th Annual Meeting of Transportation Research Board, Washington, D.C. 2008.
11. Hunter, M.P., Wu, S. K. and Kim, H. K., Practical Procedure to Collect Arterial Travel Time Data Using GPS-Instrumented Test Vehicles, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1978,* TRB, National Research Council, Washington, D.C., 2008, pp. 160-168
12. Zheng, J. Zhang, G., Wang, Y., and Briglia, P., Evaluation of Transit Signal Priority Using Observed and Simulated Data. *ITE Journal*. Vol. 79, No. 11, 42-29, 2009.
13. Zheng, J., Wang, Y., and Nihan, N., Quantitative Evaluation of GPS Performance under Forest Canopies. *Proceedings of the 2005 IEEE International Conference on Networking, Sensing, and Control*. 777-782. Mar. 2005.
14. Czerniak, R.J., Genrich, R. L., Collecting, Processing and Integrating GPS data into GIS, National Cooperative Highway Research Program, Synthesis 301, 2002.

15. IBI Group, *Continental Gateway Road Network Performance*, Transport Canada, Assignment # T8080-05-0159N, December, 2008.
16. Puget Sound Regional Council, Analysis and Forecasting in PSRC, Jan. 2010.
17. Cheu, R., Xie, C. and Lee, D., Probe Vehicle Population and Sample Size for Arterial Speed Estimation, *Computer-Aided Civil and Infrastructure Engineering*, 17 53-60. 2002
18. Chen, M. and Chien, S. I. J., Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation, In *Transportation Research Record: Journal of the Transportation Research Board.* No. 1719, TRB, National Research Council, Washington, D.C., 2000, pp. 61-68.
19. Nezamuddin, J., L. Crunkleton, P. J. Tarnoff, S. E. Young Statistical Patterns of Traffic Data and Sample Size Estimation, 2009, http://www.catt.umd.edu/documents/variance_analysis_v1.pdf Accessed: Jul. 06, 2010.
20. Li, S., Zhu, K., Van Gelder, B., Nagle, J. and Tuttle, C., Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices, In *Transportation Research Record: Journal of the Transportation Research Board.* No. *1804*, TRB, National Research Council, Washington, D.C., 2002, pp. 17-22.
21. Quiroga, C. and Bullock, D. Determination of sample sizes for travel time studies, *ITE Journal*, 68(8), pp. 92-88. 1998
22. Ma, X., Wu, Y., and Wang Y. DRIVE Net: An E-Science of Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis. In *Transportation Research Record: Journal of the Transportation Research Board.* In Press, TRB, National Research Council, Washington, D.C., 2011.
23. Google Web Toolkit, http://code.google.com/webtoolkit/. Accessed July 02, 2010.

LIST OF TABLES AND FIGURES

TABLE 1 GPS Data Overview by Vendor
FIGURE 1 GPS Data Processing Flow Chart
FIGURE 2 O-D Identification Algorithm Flowchart
FIGURE 3 Online Truck Freight Performance Measures System Flow Chart
FIGURE 4 Truck Freight Performance Measures Screen Shot

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	In-vehicle GPS with a cellular connection
Vendor B	12,000	25	0.5	In-vehicle GPS with a cellular connection
Vendor C	3,000	60	1-5	GPS cell phone

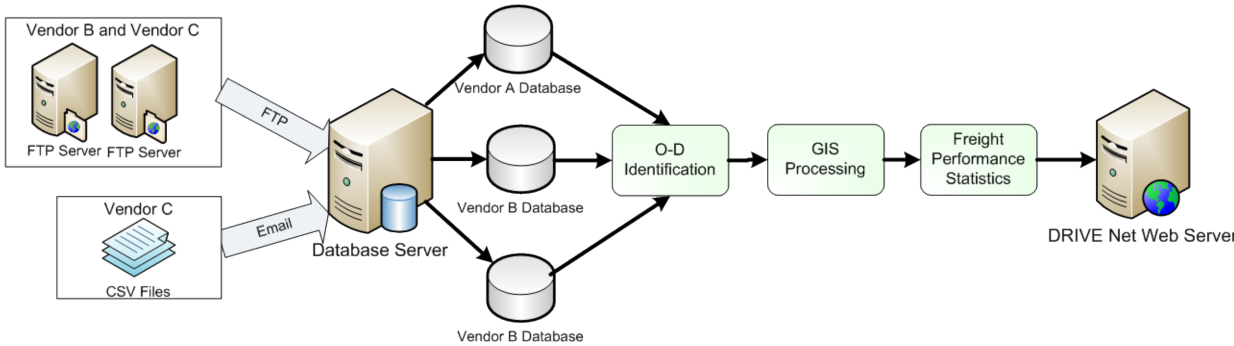

FIGURE 1 GPS Data Processing Flow Chart

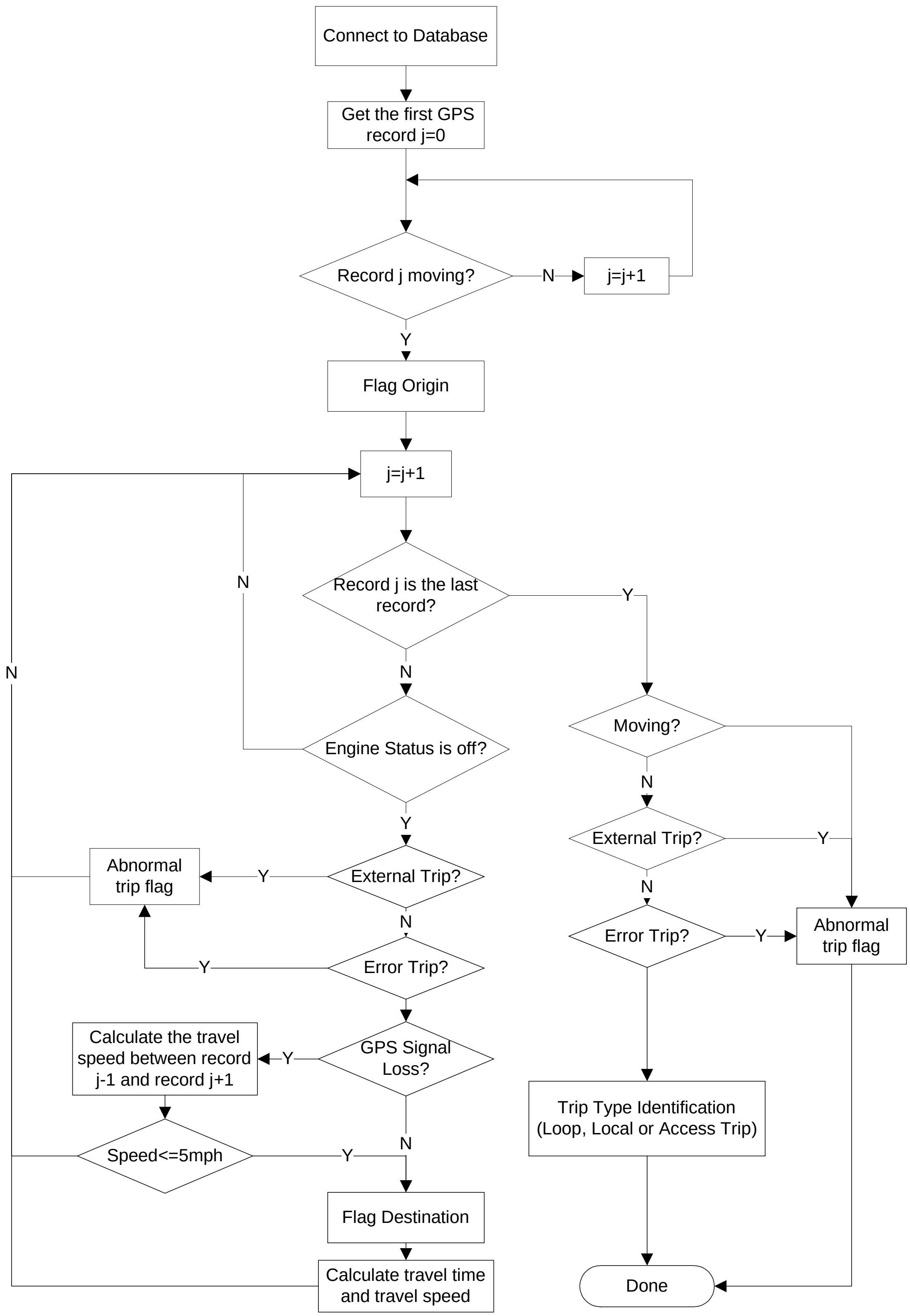

FIGURE 2 O-D Identification Algorithm Flowchart

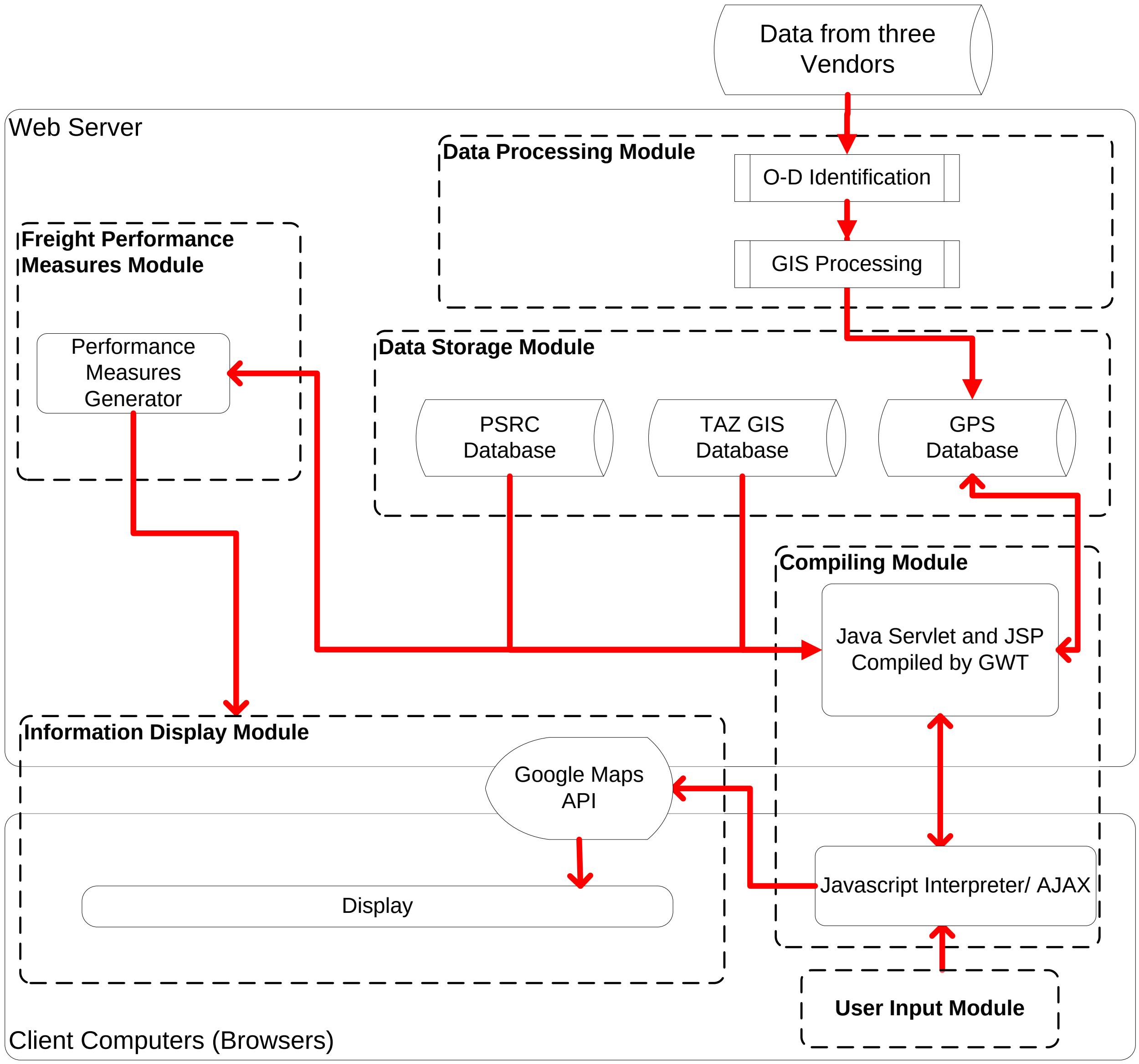

FIGURE 3 Online Truck Freight Performance Measures System Flow Chart

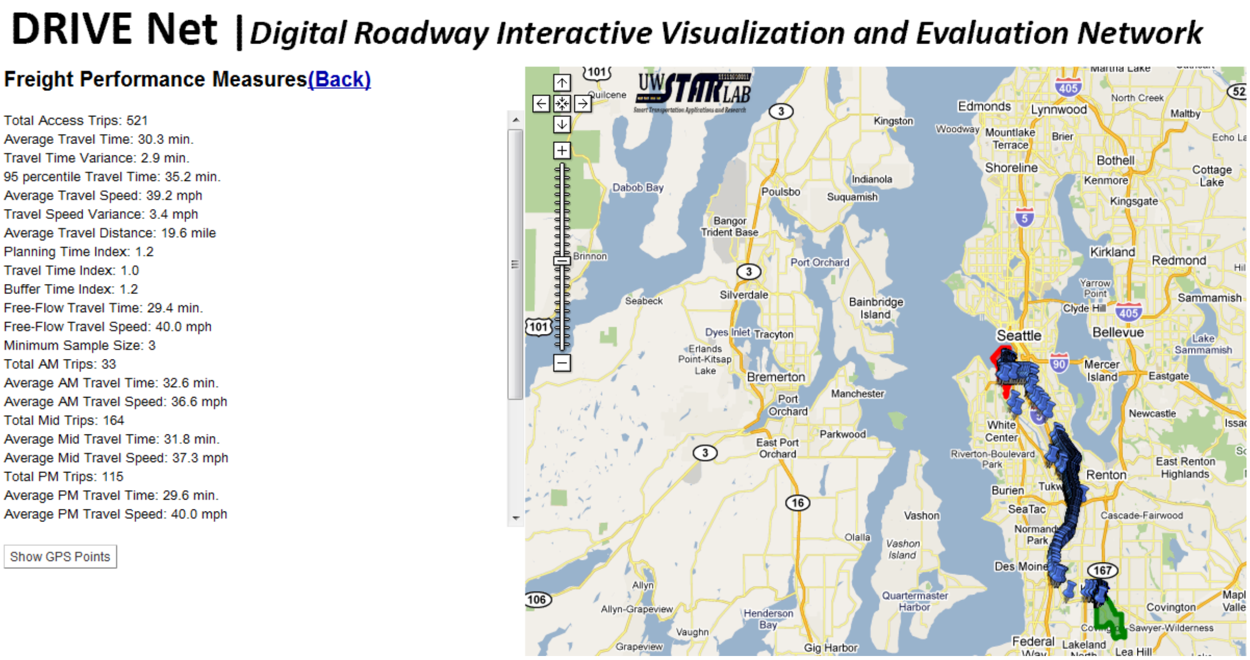

FIGURE 4 Truck Freight Performance Measures Screen Shot