Use of Truck GPS Data for Freight Forecasting

Alon Bassok[1], Ed McCormack[2] and Maren Outwater[3]

Introduction

Given trucks' importance in moving freight, relatively little is known about their travel patterns particularly in urban areas. Global positioning systems (GPS) used by trucking companies are a source of truck probe data that is just starting to be utilized for a variety of applications including modeling. The Puget Sound Regional Council (PSRC), Washington State Department of Transportation (WSDOT), and the University of Washington (UW) have 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. Among other uses, this GPS information, which is in essence a by-product of trucking industry operations, is evaluated for its feasibility to supply quantitative information that will support truck and freight models.

Data Acquisition

An earlier study determined that GPS data collection could quantify truck movements along specific roadway corridors (1). Based on this previous effort, the project team contacted approximately twenty national and local trucking and distribution companies about sharing their truck's data. With only one exception, every company agreed to provide access to their trucks' GPS data. However, actually obtaining the data from these individual companies has proven to be a challenge. One major difficulty has been trying to work out the technical details of transferring the data with their in-house data staff (if they exist).

The project team realized that it was be easier to obtain truck GPS data directly from the device vendors as opposed to the companies operating the trucks. This has several advantages: these companies have the technical staff that can set up and send routine GPS data reports; several of them are interested in our data program simply because they are trying obtain additional value from their own GPS data; and we can pay them, so we have a business relationship instead of voluntary participation.

This resulted in data acquisition contracts with three GPS vendors and the receiving of GPS trucks data on a daily basis for approximately 2,500 trucks that travel in the Puget Sound region. Since the data is from GPS devices installed for the convenience of the trucking industry, truck location reads are typically collected relatively infrequently (at starts, stops and around every 5 to 15 minutes when moving) but involves collecting data over a year.

Geo-Coding

Once the database became available, the first data processing task was to relate (geocode) each truck's location (as expressed by a GPS latitude and longitude) to its actual location on the Puget Sound region's

[1] Associate Freight/Economics Analyst, Puget Sound Regional Council, abassok@psrc.org

[2] Senior Research Engineer, Washington State Transportation Center; Research Assistant Professor, Transportation, University of Washington, Civil and Environmental Engineering, edm@u.washington.edu

[3] Director of Data Systems and Analysis, Puget Sound Regional Council, moutwater@psrc.org

roadway network as well to the traffic analysis zones (TAZs) used for transportation modeling by the PSRC. The use of Geographic Information Systems (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. This result in errors such as assignment confusion regarding overpasses and frontage roads. These issues typically require post-processing and error checking. 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 language to assign each truck's GPS reads to roadway segments. Each latitude and longitude read was assigned 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. The use of headings, which vary for each roadway segment, required that the geocoding process had to first identify the roadway corridor of interest and then assign the trucks trips to that roadway.

Identifying Trip Destinations

Because a truck model ultimately monitors travel generated by trucks as they respond to economic needs, picking up goods at origins (O) and dropping off goods at destinations (D), it was critical to develop algorithms that would extract individual truck's O/D information from the GPS data. Because GPS devices used in this effort record all stops, a methodology was needed to differentiate between traffic-based (unintended) stops at intersections or in congestion and intended stops at origins and destinations. Several studies that have analyzed GPS truck data have attempted to separate traffic stops from O/D stops based on the duration of stops (2,*3*,*4*,*5*). 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. McCormack and Hallenbeck's truck GPS data benchmarks research completed earlier in Washington State determined that 180 seconds is a usable dwell threshold (1). That amount of time 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 is completely stationary for more than 3 minutes is unusual. The processing also needs to compensate for occasional GPS signal loss due to overhead obstructions. 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.

The O/D algorithm was proofed by using one month of GPS data. This data set contained nearly 3 million GPS records. The origin and destination pairs generated by the trip end identification algorithm from the data set for one month included 358,692 trips. One advantage of this algorithm is that it allows summary statistics to be developed For example, the average trip distance between origins and destinations 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 for the Puget Sound urban area. These statistics also will be valuable for calibrating freight models. In addition, it is possible to calculate tour distance, speed and travel time for any trip type of interest.

Estimating Truck Trip/Tour Rates

One of the uses of the obtained GPS trip data is to identify trip generation rates for specific industry sectors. As a test case, a database was developed of all major grocery stores and distribution centers in the Puget Sound region. Origin and destination data for individual trucks was related to the individual grocery stores. Over a period of 91 days, nearly 2,400 trucks made 22,000 tours, consisting of over 215,000 individual trips. This results in an average number of 9 tours per truck and 0.1 tours per truck per day and 10 trips per tour. Of the trips in the tours, an average of 2 trips was made to major grocery stores. For example, a vehicle delivering beverages may stop at several grocery stores but also stop at gas stations, convenience stores and restaurants. This type of information could easily be retrieved for any employment sector and can be utilized as an input for activity based models. Further, comparable GPS data for passenger vehicles can yield similar information by any land use or activity types.

Once the information on grocery related truck trips was compiled, an attempt was made to determine the trip generation rate for grocery stores. Since the GPS data available represents a subset of all trucks on the road network, the trip generation rate derived from the GPS data was compared to manual traffic counts at select grocery stores as well as interviews with store managers. The trip rates from the GPS data compare favorably to those from interviews, which report 10 to 12 daily trucks. However, both of the figures are roughly half of those observed by the manual counts. There are several reasons that may help explain this discrepancy. First, the GPS data collected for this project represents a subset of all trucks and interviews with stores generally lead to an undercount of trucks, since only trucks that use a loading facility are described neglecting smaller deliveries—e.g. bread, flowers, etc. On the other hand, the manual counts may overestimate delivery trucks since they count all trucks, including those that are there for non-delivery purposes—e.g. a commercial vehicle stopping to pick up lunch.

A further investigation considered grocery store trip generation rates by different land use intensity types. Land uses were grouped from lowest intensity (rural) to highest intensity (metropolitan cities). As expected, higher intensity land uses receive more daily trips than those with lower intensity of use as shown in Table 1. In addition, daily truck trips have a positive relationship with the size of the store as measured by square footage, a potential indicator of the amount of stock-keeping units (SKU's). While the trucks represented in this dataset are a subset of all trucks, this analysis provides an accurate sense of travel patterns in terms of speed, times and distances. Using the manual counts as a benchmark, a relationship can be established between observed trucks volumes and those represented from the GPS data from which a figure for total grocery store truck trips could be generated. Although this investigation considered grocery stores, it demonstrates the utility and ease of using the same methodology to develop trip rates for other industries.

Table 1: Average daily truck trips by area type from GPS dataset.

Land Use	Daily Trucks
Metropolitan Cities	12.4
Core Cities	12.1
Larger Cities	8.4
Smaller Cities	6.6
Unincorporated Urban Areas	7.3
Rural	3.9

Other Potential Uses for GPS Data

We are just beginning to explore all the potential uses of the GPS data for supporting freight mobility planning activities. There are potential uses for truck forecasting models and system performance measurement. These are described below and will be tested using the GPS data over the next several years as time and resources permit.

If sufficient data is available over a long period of time, several meaningful observations can be drawn from the GPS data. Average trip and tour lengths can be estimated using the origin/destination data as well as GPS travel distance readings. This information can be further disaggregated by employment sectors, land use types and times of day. This information can be useful for estimating disaggregate distribution models or calibrating aggregate distribution models. The same is true for estimating disaggregate tour pattern models for a specific industry sector or using these data to calibrate aggregate trip generation models.

Corridor and interzonal travel speeds can be derived in order to determine speed differences for different vehicle types and for use in delay functions. Again, this is contingent on a large dataset since data over a short period of time from few vehicles may overly emphasize non-recurrent roadway problems. These data are useful for validation of existing models, but could also be considered as a performance metric for ongoing monitoring of the system.

Another promising analysis would consider developing an origin/destination matrix for the sake of comparison with a model's assignment outcome. Since the GPS data is not exhaustive, such an analysis requires sufficient observed traffic count information for trucks in order to develop a relationship between observed volumes and the GPS subset of vehicles on the roadway.

Finally, it is also possible to identify route choice splits in order to understand the relationship between vehicles on arterials versus the freeway network. In combination with other data on time and cost for alternative routes, it may also be useful to understand the relationships between route choice and time and cost tradeoffs. One example would be the time and cost tradeoffs that trucks face in crossing Puget Sound using the ferry or by driving around.

Conclusion

The use of this GPS dataset in our current research highlights the potential and merit for further use of similar datasets and development of new analytical tools for a variety of travel model and performance measure related activities. While this particular set of GPS data only considers trucks and does not differentiate different truck types, the same sort of analysis conducted here could be expanded as additional data sources become available. In addition, the relationships with the GPS vendors have thus far been successful, suggesting the possibility of continued data sharing opportunities in the future. Furthermore the vendor's GPS devices are becoming less expensive and increasing capability with more frequent location reads and vehicle performance data. Conducting these analyses with GPS data is in a nascent stage. Additional research and data development will be useful in improving the methodology and approach and in identifying further uses for this data.

References

(1) Edward McCormack and Mark E. Hallenbeck, "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.

(2) Stephen P. Greaves and Miguel A. Figliozzi, "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.

(3) Jianhe Du and Lisa Aultman-Hall, "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.

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

(5) Michael P. Hunter, Seung Kook Wu, and Hoe Kyong Kim, "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