Abstract

The availability of large samples of truck GPS data has presented a new, unprecedented source of
ingformation for understanding truck travel patterns and forecasting truck demand. This data is now
being incorporated in the development of statewide models, beginning with Indiana’s in 2012 and now
with Iowa and Tennessee’s. In order to use the data in modeling and forecasting, it is necessary to
expand the GPS-based sample data to represent all truck movements. This presents a challenge, since
the sample is not randomly drawn and cannot therefore be presumed to be representative. For
example, it is believed that short-haul movements, while present in the data, are under-represented;
without correcting for this, it is not possible to produce accurate information regarding average trip
lengths, etc. This research brief presents on-going work to understand the representativeness of the
American Transportation Research Institute’s (ATRI) truck GPS data and develop a methodology to
produce factors for expanding it to represent truck travel patterns in general.

Statement of Financial Interest

Resource Systems Group, Inc., could benefit from the presentation of this work in so far as it promotes
similar consulting engagements. The American Transportation Research Institute is a non-profit
organization, but could benefit from the presentation of this work since they supply truck GPS data for a
fee.

Statement of Innovation

The availability of large samples of truck GPS data has presented a new, unprecedented source of
information for understanding truck travel patterns and forecasting truck demand. This data is now
being incorporated in the development of statewide models, beginning with Indiana’s in 2012 and now
with Iowa and Tennessee’s. In order to use the data in modeling and forecasting, it is necessary to
expand the GPS-based sample data to represent all truck movements. This presents a challenge, since
the sample is not randomly drawn and cannot therefore be presumed to be representative. For
example, it is believed that short-haul movements, while present in the data, are under-represented;
without correcting for this, it is not possible to produce accurate information regarding average trip
lengths, etc. This research brief presents on-going work to understand the representativeness of the
American Transportation Research Institute’s (ATRI) truck GPS data and develop a methodology to
produce factors for expanding it to represent truck travel patterns in general.
EXPANDING TRUCK GPS-BASED PASSIVE ORIGIN-DESTINATION DATA IN IOWA AND TENNESSEE

Vincent L. Bernardin, Jr., Ph.D.
Senior Consultant, RSG
2709 Washington Ave, Ste. 9
Evansville, IN 47714
Vince.Bernardin@RSGinc.com
Ph: 812-200-2351

Steven Trevino
Analyst, RSG
2709 Washington Ave, Ste. 9
Evansville, IN 47714
Steven.Trevino@RSGinc.com
Ph: 812-200-2351

Jeffery Short
American Transportation Research Institute
2060 Franklin Way, Suite 201
Marietta, GA  30067
JShort@trucking.org
Ph: 770-432-0628

Submitted: March 28, 2014
STATEMENT OF INNOVATION
The availability of large samples of truck GPS data has presented a new, unprecedented source of information for understanding truck travel patterns and forecasting truck demand. This data is now being incorporated in the development of statewide models, beginning with Indiana’s in 2012 and now with Iowa and Tennessee’s. In order to use the data in modeling and forecasting, it is necessary to expand the GPS-based sample data to represent all truck movements. This presents a challenge, since the sample is not randomly drawn and cannot therefore be presumed to be representative. For example, it is believed that short-haul movements, while present in the data, are under-represented; without correcting for this, it is not possible to produce accurate information regarding average trip lengths, etc. This research brief presents on-going work to understand the representativeness of the American Transportation Research Institute’s (ATRI) truck GPS data and develop a methodology to produce factors for expanding it to ensure it is representative of truck travel patterns in general.

STATEMENT OF FINANCIAL INTEREST
Resource Systems Group, Inc., could benefit from the presentation of this work in so far as it promotes similar consulting engagements. The American Transportation Research Institute is a non-profit organization, but could benefit from the presentation of this work since they supply truck GPS data for a fee.
OBJECTIVES, MOTIVATION AND INNOVATION

The availability of large samples of truck GPS data has presented a new, unprecedented source of information for understanding truck travel patterns and forecasting truck demand. This data is now being incorporated in the development of statewide models, beginning with Indiana’s in 2012 and now with Iowa and Tennessee’s. However, in order to use the data in modeling and forecasting, it is necessary to weight or expand the GPS-based sample data to represent all truck movements.

The American Transportation Research Institute (ATRI) is a 501(c)3 not-for-profit research organization. As an independent part of the American Trucking Associations (ATA) Federation, it receives truck GPS position data from many of its members. The primary source of the data is on-board communications and navigation equipment installed on commercial trucks. The resulting dataset is extremely large and growing, containing several billion truck positions annually, representing several hundred thousand individual trucks out of the total 2.4 million trucks registered in the US. (1) ATRI compiles the data as part of the Freight Performance Measures Initiative, an effort sponsored in part by the Federal Highway Administration.

ATRI’s truck data has been used for nearly a dozen years in many analyses to identify truck bottlenecks, congestion and speeds on major highway freight corridors across the country. More recently it was used to better understand truck origin-destination patterns as part of a feasibility study of dedicated truck lanes on an 800-mile section of I-70 through Missouri, Illinois, Indiana and Ohio with funding from FHWA’s Corridors of the Future program. The GPS data proved to be extremely useful in understanding the patterns of truck movements utilizing the corridor as well as “ground-truthing” and improving travel demand forecasts. However, the data was not formally incorporated in any models.

The Indiana Statewide Travel Demand Model was the first travel forecasting model to formally incorporate the data when it was updated in 2011-2012. (2) The Indiana Department of Transportation obtained an eight week sample of ATRI’s truck GPS data drawn from each of the four quarters of 2010. The resulting sample contained over 16 million records representing over 2 million trips by over 300,000 trucks. The data was analyzed and while it was determined that the ATRI data clearly includes short-haul movements, it was determined that the preponderance of the data related to medium- and longer-haul truck trips and it was therefore used to improve the commodity-flow based truck movements in the model. The incorporation of the ATRI data into the model resulted in substantial improvements in the truck model validation, decreasing the root mean squared error (RMSE) against roughly 6,000 truck counts from 69.3% to 60.6% and decreasing the mean absolute percentage error (MAPE) from 74% to 42%. (3)

Despite the very positive results on the truck model performance, it was recognized that further improvement was likely possible if the representativeness of the ATRI data was better understood so that it could be expanded more realistically.
Inspired by the success in Indiana, both the Iowa and the Tennessee departments of transportation have acquired truck GPS data from ATRI in order to support the update of their statewide travel models. As part of these efforts, RSG has undertaken research with the support of ATRI, to study the representativeness of ATRI’s GPS truck data and develop more refined methods of weighting or expanding it for modeling purposes. In particular, the desire is to better understand both geographic and distance-based biases in the sample relative to total truck traffic.

**METHODOLOGY**

Both Iowa and Tennessee obtained an eight week sample of trucks passing through their respective statewide model study areas drawn from each of the four quarters of 2012. For each model, a GIS file representing travel analysis zones (TAZ) was provided to ATRI. ATRI selected all truck GPS traces entering, exiting, traveling within or passing through the overall area for each of the sampled weeks. A data management and analysis software package was used to further prepare the dataset for integration into the truck trip table. Truck positions for each unique vehicle were sorted into a time series, and within each series each truck position was matched with the subsequent truck position to produce a set of truck position pairs. The geodetic distance between the first and second truck positions for each of the truck position pairs was then calculated. ATRI then replaces the precise GPS location data in its records with the TAZ. In addition to supporting the ultimate development of a trip table, this process also offers some benefit of further ensuring the anonymity of the data by associating truck positions with geographic areas far more generalized than a discrete latitude/longitude position – which could allow for the development of an address-specific customer list. The dataset is then reformatted so that each record represents the movement of a truck between GSP ‘pings’. ATRI then delivers a dataset containing an anonymous truck identifier, the distance between pings, the TAZ position of the beginning and ending ping and the timestamp of the beginning and ending ping (see Figure 1).

<table>
<thead>
<tr>
<th>from TAZ to TAZ</th>
<th>distance</th>
<th>time</th>
<th>elapsed time</th>
<th>speed</th>
<th>status1</th>
<th>status2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 101032</td>
<td>66.0</td>
<td>57.7</td>
<td>57.7</td>
<td>68.6m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>101032 101033</td>
<td>18.3</td>
<td>14.3</td>
<td>72.0</td>
<td>68.6m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>101033 101015</td>
<td>26.8</td>
<td>27.9</td>
<td>99.9</td>
<td>57.5m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>101015 101015</td>
<td>0.0</td>
<td>5.0</td>
<td>5.0</td>
<td>0 stopped</td>
<td>stopped</td>
<td></td>
</tr>
<tr>
<td>101015 101015</td>
<td>0.2</td>
<td>2.7</td>
<td>7.7</td>
<td>5.2 stopped</td>
<td>stopped</td>
<td></td>
</tr>
<tr>
<td>101015 101015</td>
<td>0.3</td>
<td>9.8</td>
<td>17.5</td>
<td>2.0 stopped</td>
<td>stopped</td>
<td></td>
</tr>
<tr>
<td>101015 101015</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>28.2m stopped?</td>
<td>stopped</td>
<td></td>
</tr>
<tr>
<td>101015 2035</td>
<td>37.1</td>
<td>60.0</td>
<td>60.3</td>
<td>37.1m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>2035 18099</td>
<td>67.8</td>
<td>65.4</td>
<td>125.7</td>
<td>62.2m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>18099 27006</td>
<td>5.9</td>
<td>5.4</td>
<td>131.1</td>
<td>65.3m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>27006 18023</td>
<td>10.0</td>
<td>15.9</td>
<td>147.0</td>
<td>37.8m</td>
<td>moving</td>
<td>moving</td>
</tr>
<tr>
<td>18023 18023</td>
<td>0.0</td>
<td>5.0</td>
<td>5.0</td>
<td>0 stopped</td>
<td>stopped</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1 - Identifying Stops from GPS Trace Data*

This dataset must then be processed and further reduced to represent trips between origin-destination pairs. This is done in two steps, first identifying for each pair of GPS pings or
movement record, whether the truck was in motion or stopped. This determination is made based on complex criteria of a minimum travel speed and a minimum elapsed time and/or distance. The complex criteria are necessary to avoid including brief stops at traffic signals or brief repositioning movements within a single site (see Figure 2 for an illustration).

Figure 2 - Brief Re-positioning Movements to be Distinguished from Trips

Once the moving records and stopped records are identified, the records are processed to identify the origin and destination for each sequence of moving records. When a stop record was found in the list, it signified the destination of the trip and the origin for the subsequent trip (see Figure 1). The result is a list of trips by origin-destination pair, which can be aggregated by origin-destination pair to produce a trip-table in flat/list format, which, in turn, can be read into a matrix format file by most travel modeling software.

The focus of this effort is on the expansion of this resulting raw trip table. The simplest method is to scale the trip table to reflect the total number of truck trips or truck VMT. However, this fails to account for differences in the portion of the universe of trucks represented in the sample. For example, it is believed that short-haul movements, while present in the data, are under-represented; without correcting for this, it is not possible to produce accurate information regarding average trip lengths, etc. There is no initial hypothesis as to whether there are any geographic biases in the data, but it is a possibility, since the sample is not randomly drawn. If there are geographic biases it would be important to correct for this to avoid distorting the spatial distribution of trips.

The approach taken in this effort begins by simply scaling the raw ATRI trip table to represent the proper amount of truck VMT. Then, origin-destination matrix estimation (ODME) algorithms are applied using truck counts on the network and the scaled trip table as a seed.
Several methods of ODME are being tested as well as differing convergence criteria with close attention paid to the magnitude of the adjustments from the seed matrix. Although the effort is commenced using some of the most common algorithms (4, 5), other more specialized algorithms (6) may also be explored if necessary, particularly to explore geographic biases. The results of the ODME are not used directly, but analyzed to identify significant patterns. Once patterns of over- and/or under-representation have been identified, a categorization scheme is developed to address these systematic issues and average adjustment factors for each category are calculated from the ODME results judged most reliable. The final expansion factors are then the product of the scaling factor and the adjustment factors.

While ODME algorithms are an imprecise method, they provide a consistent, structured and logical framework for relating the information contained in truck count data to origin-destination patterns. Truck counts, while also imperfect, are widely available and believed to be the best source of unbiased data on total truck traffic.

The Iowa statewide model’s (iTRAM) network includes 84,090 links in Iowa, representing 28,563 miles of roadway, of which 36% have a truck count or count derived estimate of truck traffic. This should provide a rich dataset for ODME analysis and expansion. The Tennessee network has a similarly rich set of truck counts and count derived estimate of truck traffic. The results of both analyses together will provide a valuable cross-check on results as well as some sense of the transferability of results between various parts of the country.

MAJOR RESULTS
As of the writing of this brief, the Iowa data has been analyzed and expanded, and the Tennessee data was being processed to produce a raw trip table. The expanded Tennessee truck trip table is due to be delivered by the end of April, so it hoped its results will also be available in time for presentation at the conference.

The Iowa dataset was spread over eight weeks in 2012 and two in early 2013. It included over 50 million movement records by over 135,000 individual trucks making over 3.1 million trips. Due to quality issues with GPS signals, it was necessary to remove roughly 400,000 trips with suspect GPS data. It was judged better to be conservative and throw out a few possible good trips than to include a few very bad ones. The resulting final dataset contained over 2.7 million trips, including over 60 million truck miles of travel observed in Iowa over the 56 days in the sample. This translated into 1,083,152 truck VMT in Iowa on a daily basis. Total daily truck miles of travel in Iowa are estimated by Iowa DOT to be 10,731,507 based on 2012 HPMS data. This represents an average, overall sampling rate of roughly 10.1%, corresponding to a scaling factor of 9.908. Since the sample was for 56 days, however, this means that the raw daily OD table was produced by factoring the total down by 0.1769 to represent a daily number.

The raw OD table was used as seed matrix for ODME and the results were analyzed to evaluate possible biases based on geographic regions or trip length. Although there are some anecdotes of
regional differences in ATRI sample coverage around the US, no evidence of systematic geographic differences in sample coverage were identified in Iowa or its immediate halo area. Figure 3 illustrates the OMDE implied weights within Iowa and clearly demonstrates there are no systematic biases. It is still possible that geographic or regional biases exist in other parts of the US, but in and around Iowa this does not appear to be an issue. Therefore, geography was not used in the weighting and expansion scheme.

Figure 3 - ODME based weighting showing no evidence of systematic geographic bias in Iowa

The analysis did, however, confirm suspected bias towards longer haul truck trips over short haul truck trips. At the same time, the data demonstrated that there were still a significant number of short haul movements in the date, such that they could be reasonably weighted to produce a more representative dataset. Figure 4 compares truck trip length frequency distributions from the raw ATRI data and resulting from ODME. It can clearly be seen that trips less than 60 miles in length are factored up in the ODME process, while trips greater than 60 miles in length are factored down very slightly.
A weighting scheme was therefore developed based on trip length, and can be seen in Figure 5. Short haul truck trips (<60 miles) were weighted up while long haul trips (>60 miles) were slightly weighted down. The resulting weights ranged from slightly over 0.7 to just over 2.5; the latter value only applying to truck trips less than 10 miles in length. Trips even just 10 – 20 miles in length only had to be factored up by less than 1.8. While this does indicate the need to correct the data for this bias, it also suggests that a simple and reasonable weighting scheme can produce reasonably good results, explaining roughly half the variation in ODME weighting.
IMPLICATIONS FOR TRAVEL MODELING

Given the importance of freight planning, it is critical to have accurate freight planning tools and models, but this is impossible without high quality data. The ATRI GPS truck data provides a rich and growing source of information on truck travel patterns. However, it is still only a sample, and like any sample care and an understanding of its representativeness is required in expanding it. Failure to correctly account for the under- or over-representation of certain categories of trucks could lead to faulty analyses and false conclusions.

This research presents an important step forward in our understanding of the representativeness of ATRI’s truck GPS data and how to weight or expand it for use in forecasting models and other analyses. With proper understanding and expansion techniques, ATRI’s truck GPS data stands to support a new generation of truck models with substantially better accuracy than earlier models and methods based on very limited data. The research presented here is critical to ensure that the data is used in an appropriate way.

REFERENCES