January 2014

Deriving Statewide Freight Truck Flows from Global Positioning System (GPS) Data

Akbar Bakhshi Zanjani
University of South Florida, akbzanj@gmail.com

Deriving Statewide Freight Truck Flows from Global Positioning System (GPS) Data

by

Akbar Bakhshi Zanjani

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Civil Engineering
Department of Civil and Environmental Engineering
College of Engineering
University of South Florida

Major Professor: Abdul R. Pinjari, Ph.D.
Pei-Sung Lin, Ph.D.
Jason J. Bittner, M.P.A.

Date of Approval:
July 2, 2014

Keywords: ATRI, Conversion Algorithm, Florida Statewide Model (FLSWM), ODME, Cube Software

DEDICATION

I dedicate my thesis to my family. My father, Asghar Bakhshi Zanjani, for his unabated support and endless efforts to provide us with the best possible quality of life, and my mother, Esmat Samadi, who filled my life with her care, devotion and love, Also my brother and sister, Mehdi and Negin, who have never left my side and are very special.

I would also like to dedicate this thesis to a special friend of mine, Alisa Anna Vasserman, whose inspiration, encouragement and optimism helped me through some of the most difficult stages of my life as a graduate student.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Abdul Pinjari for the continuous support of my Master's study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Master's study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Pei-Sung Lin and Mr. Jason Bittner for their encouragement, insightful suggestions and comments.

I would like to acknowledge the financial support of the Florida Department of Transportation (FDOT) during my entire research at the University of South Florida. However, "the opinions, findings, and conclusions expressed in this publication are those of the author and not necessarily those of the Florida Department of Transportation or the U.S. Department of Transportation."

I would also like to acknowledge the former and current members in our research team at the University of South Florida: Aayush Thakur, Anissa Nur Irmania, Jonathan Koons, Vaibhav Rastogi, Anirban Ghosh, Mohammadreza Kamali and Bertho Augustin for the support and help they provided me with. Also I want to thank my former and current officemates at the University of South Florida: Sujan, Vijay, Sashi and Lukai for their support and all the fun we have had in the last few years.

Last but not least, I would like to thank my family for their unconditional love and support. This would have never been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

An accelerated growth in the volume of freight shipped on Florida's highways has led to a significant increase in truck traffic, influencing traffic operations, safety, and the state of repair of highway infrastructure. Traffic congestion in turn has impeded the speed and reliability of freight movement on the highway system. Appropriate planning and decision making processes are necessary to address these issues. However, a main challenge in establishing such processes is the lack of adequate data on statewide freight movements. As traditional data sources on freight movement are either inadequate or no longer available, new sources of data must be investigated.

A recently available source of data on nationwide freight flows is based on a joint venture by the American Transportation Research Institute (ATRI) and the Federal Highway Administration (FHWA) to develop and test a national system for monitoring freight performance measures on key corridors in the nation. This data is obtained from trucking companies who use GPS-based technologies to remotely monitor their trucks. The database contains GPS traces of a large number of trucks as they traveled through the national highway system. This provides unprecedented amounts of data on freight truck movements throughout the nation (and in Florida). Such truck GPS data can potentially be used to support planning, operation, and management processes associated with freight movements. Further, the data can be put to better use when used in conjunction with other freight data obtained from other sources.

The overarching goal of this thesis is to investigate the use of large streams of truck-GPS data from the American Transportation Research Institute (ATRI) for the estimation of statewide

freight truck flows in Florida. To this end, first, an algorithm was devised to convert large streams of raw GPS data into a database of truck trips. The algorithm was applied to four months of ATRI's truck-GPS data comprising over 145 Million GPS records to derive a database of more than 1.2 million truck trips starting and/or ending in Florida. This database was used to analyze truck travel characteristics and origin-destination truck flow patterns for different geographical regions in Florida. The resulting database was used in conjunction with the GPS data to analyze the extent to which ATRI's data represents observed truck traffic flows in the state. It was found that at an aggregate level, almost 10% of heavy truck traffic flows in Florida is captured in the ATRI data.

Finally, the database of truck trips derived from ATRI's truck-GPS data was combined with observed heavy truck traffic volumes at different locations within and outside Florida to derive an origin-destination (OD) table of truck flows within, into, and out of the state. To this end, first, the truck trip database developed from ATRI's truck-GPS data was converted into a seed OD table at the TAZ-level spatial resolution used in FLSWM. Subsequently, a mathematical procedure called origin-destination matrix estimation (ODME) method was employed to combine the OD flow table generated from the ATRI data with observed truck traffic volume information at different locations within and outside Florida. The OD table of truck flows estimated from this procedure can be used for a variety of purposes, including the calibration and validation of the heavy truck modeling components of FLSWM.

CHAPTER 1: INTRODUCTION

1.1 Background

Freight is gaining increasing importance in transportation planning and decision making at all levels of the government – MPOs, state, and federal – for several reasons. An accelerated growth in the volume of freight shipped on American highways has led to a significant increase in the truck traffic. This has put enormous pressure on national highways impacting traffic operations, safety, highway infrastructure, port operations, and distribution center operations. Besides, traffic congestion impedes the speed and reliability of freight movement on the highway system and leads to direct economic costs for producers and consumers, passenger traffic congestion, safety issues, and environmental impacts.

As freight movement continues to grow within and between urban areas, appropriate planning and decision making processes are necessary to mitigate the above-mentioned impacts. However, a main challenge in establishing these processes is the lack of adequate data on freight movements such as detailed origin-destination (OD) data, truck travel times, freight tonnage distribution by OD pairs, transported commodity by OD pairs, and details about truck trip stops and paths. As traditional data sources on freight movement are either inadequate or no longer available (e.g., the Vehicle Inventory and Use Survey), new sources of data must be investigated.

1.2 Motivation

Unlike passenger cars, for other modes such as trucking, OD matrix estimation from traffic counts is not widely studied and much work needs to be done (Gonzalez-Calderon et al. 2012). During past decades, collecting data to estimate OD matrices for trucks has been highly costly as well as labor intensive. However, recognizing the need for better freight data, several efforts are underway to exploit advanced technologies and form innovative partnerships with freight stakeholders to gather freight movement data. Many trucking companies use advanced vehicle monitoring (AVM) systems that allow remote monitoring of their fleets using Geographical Positioning Systems (GPS) technology-based Automatic Vehicle Location (AVL) systems. To tap into such truck GPS data sources, private-sector truck data providers have formed innovative partnerships with freight carriers and other freight stakeholders to collect the GPS data and provide it to public agencies while protecting the confidentiality of the data. Notable among such efforts is a joint venture by the American Transportation Research Institute (ATRI) and the FHWA to develop and test a national system for monitoring freight performance measures (FPM) on key corridors in the nation. This FPM data system is built based on data obtained from trucking companies who use GPS-based AVM/AVL technologies to remotely monitor their trucks. ATRI's FPM database contains GPS traces of a large number of trucks as they traveled through the national highway system. This provides unprecedented amounts of data on freight truck movements throughout the nation (and in Florida). Such truck GPS data can potentially be used to support planning, operation, and management processes associated with freight movements. Further, the data can be put to better use when used in conjunction with other freight data obtained from other sources.

ATRI's truck GPS data has been used for freight performance measurement and planning applications in the recent past. The applications include identifying and prioritizing major freight bottlenecks on the nation's highways, analyzing truck route patterns after major highway incidents, truck parking issues, and regional and statewide truck flow modeling.

For the purpose of exploring ATRI data, as a project sponsored by Florida Department of Transportation (FDOT), the research team at the University of South Florida (USF) was provided with more than 145 million GPS records of truck movements. Each GPS record provided information on its spatial and temporal location along with a unique truck ID that does not change across all the GPS records of a truck for a certain time period varying from a day to over a month. The frequency (i.e., ping rate) of the GPS data streams varies considerably, ranging from a few seconds to over an hour of interval between consecutive GPS records. A part of the data contains information on the spot speed (or instantaneous speed) and the direction of heading for each GPS data point. To protect the confidentiality of the businesses associated with these trucks, information such as the businesses served, truck type/axle configuration, commodity being carried, and the purpose of travel are not available. ATRI estimates that a large share of trucks in ATRI's database falls under FHWA's vehicle type classification 9 or larger (i.e., tractor-trailers with 5 or more axles), some belong to class-8, and a small proportion belong to class-7 or below trucks that are less likely to carry freight. Figure 1.1 shows the FHWA's vehicle classification. The provided GPS records belong to Florida centric truck movements. In other words, all the trucks from ATRI's database that were in Florida at any time during 4 months of March, April, May and June 2010 were extracted. Then the GPS traces of those trucks were extracted for the range of two weeks to an entire month, as they traveled within Florida as well as in other states in the nation and Canada. This allowed the examination of truck movements within Florida as well as truck flows into (and out of) Florida from (to) other locations in the nation. Being provided with this large number of GPS records, a major task was to develop an algorithm to determine the truck trip ends. The GPS to Trip Conversion Algorithm (GTCA) was developed to identify the trips made by each truck by finding the origins and destinations of these trips. The algorithm was then coded in Java programming language which uses raw GPS records as the input and generates trips along with proper attributes and performance metrics as the output. The freight truck trips derived from ATRI's truck GPS data can be used to derive statewide origin-destination tables (or OD flow tables or OD matrices) of freight truck flows between various traffic analysis zones. However, it is important to note that while the trips derived from the ATRI data comprise a rather large sample, they represent only a sample of freight truck flows within, to, and from the state. Besides, while it is known that the ATRI data is predominantly comprised of tractor trailer trucks that fall under class-8 to class-13 of FHWA vehicle type classification (i.e., heavy trucks), it is not certain if the data represents a random sample of heavy truck flows in the state. Therefore, additional information and procedures must be employed to factor the sample of trips derived from the ATRI data to represent the population of heavy truck flows within, to, and from the state. The weighting process is required not only for inflating the sample of ATRI truck flows to the population truck flows but also for ensuring that the spatial distribution of the resulting truck flows is representative of the actual truck flows in the state. One approach to do this is Origin-Destination Matrix Estimation (ODME), which involves combining the sample OD truck flows derived from the ATRI data with other sources of information on truck flows observed at various links of the highway network to estimate a full OD flow matrix representing the population of truck flows in the state.

1.3 Objectives

The overarching goal of this thesis is to investigate the use of ATRI's truck-GPS data for statewide freight performance measurement, statewide freight truck flow analysis, and the use of this database in combination with other data sources for developing truck travel origin-destination flow patterns in the state of Florida (and to/from Florida). Specifically, the following goals were investigated in detail.

1.3.1 Develop Methods to Convert ATRI's Raw GPS Data Streams into Truck Trips

The raw GPS data streams from ATRI need to be converted into a truck trip format to realize the full potential of the data for freight planning applications. Therefore, the first objective of this thesis is to convert the raw GPS data into a data base of truck trips. Development of such a truck trip database involves the determination of truck starting and ending instances and locations, trip distance, total trip duration, and duration of intermediate stops (e.g., at traffic signals and rest stops) in the trip. In addition, the process involves resolution of potential anomalies in GPS data, such as data-discontinuities due to loss of satellite signals. This task involved the development of algorithms and a software code to convert the raw GPS data streams into a truck trip format. These algorithms were then applied to four months of raw GPS data from ATRI, comprising a total of 145 Million raw GPS data records, to develop a large database of truck trips traveling within, into, and out of the state. The resulting database comprises over 1.2 Million truck trips traveling within, into, and out of the state. Also ATRI's truck-GPS data was used to analyze truck travel characteristics in the state of Florida. To this end, this thesis involves an analysis of the truck trip data derived from the four months of ATRI's truck GPS data. The truck travel characteristics analyzed include trip duration, trip length, trip speed, time-of-day profiles, and origin-destination flows. Each of these characteristics are derived at a statewide level as well as for different regions in the state – Jacksonville, Tampa Bay, Orlando, Miami, and rest of Florida – defined based on the freight analysis framework (FAF) zoning system.

1.3.2 Assess ATRI's Truck GPS Data and Its Coverage of Truck Traffic in Florida

The second major objective of this thesis is to assess ATRI's truck-GPS data in Florida to gain an understanding of its coverage of truck traffic in the state of Florida. This includes deriving insights on: (a) the types of trucks (e.g., heavy trucks and light trucks) present in the data, and (b) the geographical coverage of the data in Florida, and (c) the proportion of the truck traffic flows in the state covered by the data.

1.3.3 Derive Statewide Truck Trip Flow Origin-Destination Tables for the Traffic Analysis Zone (TAZ)-Level Spatial Resolution in the Florida Statewide Model (FLSWM)

An important objective of the thesis is to use ATRI's truck-GPS data in combination with other available data sources to derive origin-destination (OD) tables of freight truck flows within, into, and out of the state of Florida. As part of this task, first, the truck trip database developed from four months of ATRI's GPS data is converted into OD tables at the traffic analysis zone (TAZ)-level spatial resolution used in the Florida Statewide Model (FLSWM). Such an OD table derived only from the ATRI data, however, is not necessarily representative of the freight truck flows in the state. This is because the ATRI data does not comprise the census of trucks in the state; the data comprises only a sample of trucks traveling in the state. Although it is a large sample, it is not necessarily a random sample and is likely to have spatial biases in its representation of truck flows in the state. To address these issues, the OD tables derived from the ATRI data need to be combined with observed truck traffic volumes at different locations in the state (and outside the state) to derive a more robust origin-destination table that is representative of the freight truck flows within, into, and out of the state. To achieve this, a mathematical procedure called origin-destination matrix estimation (ODME) method is employed to combine the seed OD flow matrix generated from the ATRI data with observed truck traffic volume information at different locations within and outside Florida.

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows. Chapter 2 provides a review of the literature and existing methods used for converting GPS records to trips as well as Origin-Destination Matrix Estimation (ODME). Chapter 3 presents the procedures and algorithms used to convert the raw GPS data streams into a large database of truck trips in Florida. This chapter also presents an analysis of some of the truck travel characteristics using the truck trip database developed from four months of raw GPS data. Chapter 4 presents an assessment of ATRI's truck-GPS data in Florida, specifically in terms of its coverage of truck traffic flows in the state of Florida as well as different geographical locations. Chapter 5 explains the methodology used for different steps in the ODME process including the highway assignment method as well as the estimation procedure. Chapter 6 describes the inputs and different assumption used for ODME. Chapter 7 explains the results of the ODME procedure and some suggestions to improve the ODME results and finally chapter 8 summarizes the findings in this study and identifies opportunities for future research and implementation.

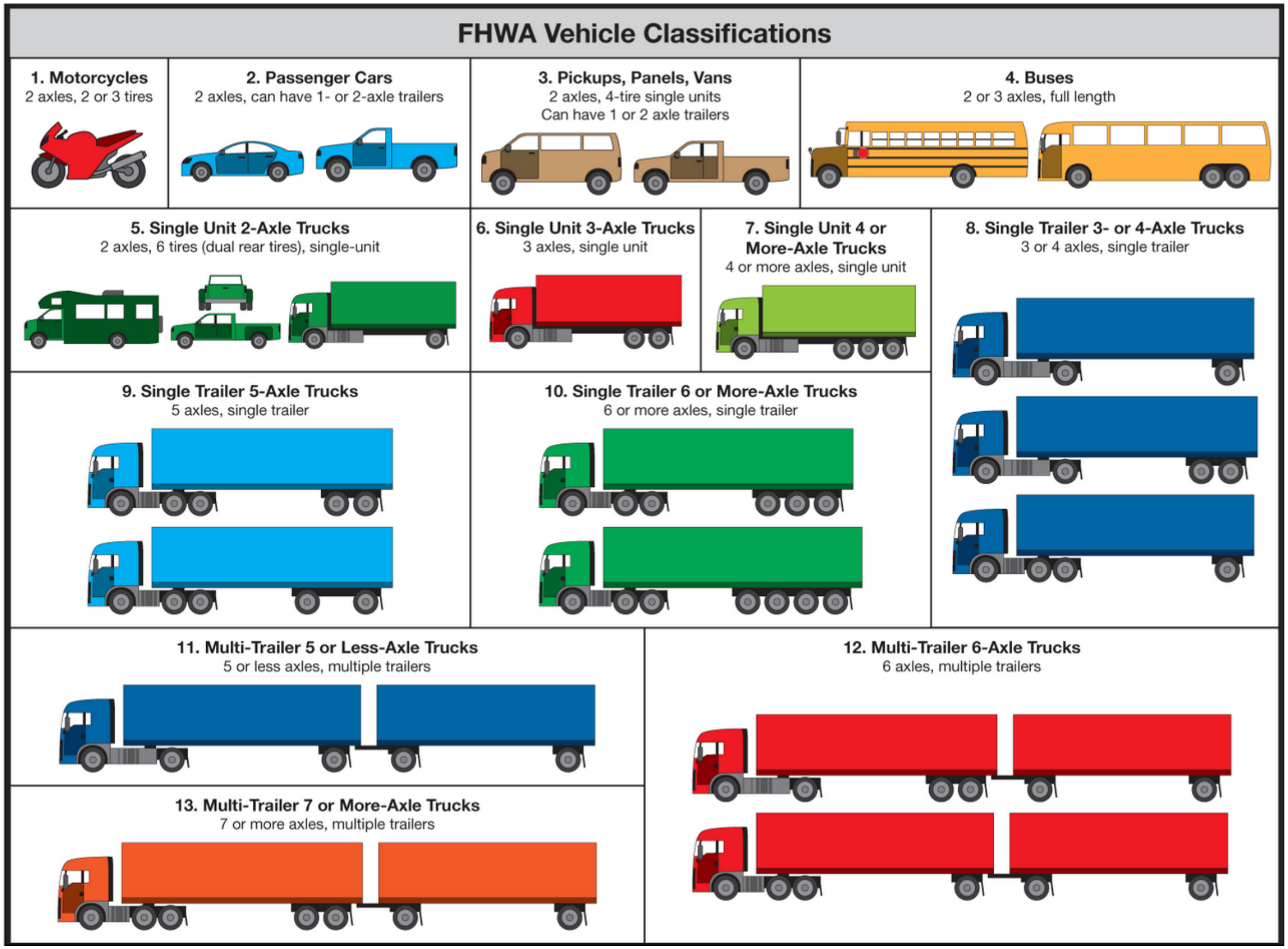


Figure 1.1 FHWA Vehicle Classification[1]

[1] Source: http://onlinemanuals.txdot.gov/txdotmanuals/tri/images/FHWA_Classification_Chart_FINAL.png accessed on 6-13-2014.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

As mentioned earlier, due to high cost of collecting truck traffic count data, this mode of transportation is not studied in deep detail to get the origin-destination matrices compared to passenger car origin-destination studies using GPS data. In this chapter we first discuss some of the previous studies performed on the conversion of GPS records to truck trips which, as mentioned in the previous chapter, is a required step to get the initial origin-destination flows. Later, we will review some of the previous studies on the origin-destination matrix estimation process (ODME).

2.2 Previous Studies on Conversion of GPS Data to Trips

Some studies in the literature as well as some government sponsored projects have explored the topic of conversion of GPS records to trips for different modes of transportation. During a project sponsored by Maricopa Association of Governments (MAG), analysts at Cambridge Systematics acquired and processed the ATRI data to gather sufficient details to support the development of truck trip and tour information. As mentioned earlier, the main data source used in their project was the GPS information gathered from trucking companies by ATRI. This database included information from different GPS providers on locations, random truck IDs, and time stamps. The data for the study area defined by MAG was collected for the period from April 1, 2011 to April 30, 2011. It contained more than 3 million GPS event records for the movements of more than 22,000 trucks. Some of the assumptions and findings in their project are listed as follows:

- “Air speed” (average speed between two consecutive GPS event records) of 5 MPH was considered as the threshold to determine starting and stopping points known as “transition events” in this project. Before any additional processing, close to 7 million GPS transition events were identified, which represent 349,913 potential truck trips.
- Furthermore, the traffic stops had to be removed. The transition events whose time duration at a potential trip end were less than 2 minutes were recognized as traffic stops and therefore not a valid trip end. Also some transition events were on roadways. Using the land use information these starting and stopping points were identified and eliminated.
- Final internal-internal truck model for the MAG model region consisted of 143,666 truck trips, 121,863 of which occur on weekdays, excluding Saturdays and Sundays. Average duration of a stop was 135 minutes. Average travel time between stops was 57.3 minutes.

During another similar project sponsored by Indiana Department of Transportation (INDOT) an effort was made to update Indiana Statewide Truck Travel Demand Model (ISTDM) based on ATRI data and the latest traffic count data as of November 2012 for Indiana and the adjacent states, i.e. Ohio, Kentucky, Illinois and Michigan. Some information on the data, procedure and findings include:

- Sample ATRI data set for eight one-week time periods from February, May, July and October 2010 were used. The study area covered Indiana in addition to the areas 50 miles beyond the state border.
- Data included approximately 6 million records for the entire 8 weeks. These records were later formatted into truck movements between origins and destinations. Large truck stops for refueling, rest and similar purposes were identified to separate these stops from valid pick up/delivery stops.

- After generating trip length distributions, some short distance trips were removed based on distance (less than quarter miles) and speed (less than 10 mph). Later, more than 2 million truck trips were extracted over the entire 8 weeks which were then factored to account for seasonal fluctuations and then scaled up 20% to get a better match for national freight trips. Also, during the process of scaling up, it was determined that the freight truck type has the best fit to the ATRI data. Overall, the factoring and scaling up process resulted in more than 312,000 freight truck trips per week.

Stopher et al. (2003) developed four rules to identify trip ends based on GPS data, which were effective and were able to minimize error. First, a potential trip end was marked if the time spent in the point was equal to or greater than 120 seconds, the difference in the latitude and longitude values were less than 0.000051^{o} and during that time the direction remained the same with zero speed. Secondly, if the engine turned off between 30 to 120 seconds. Thirdly, if the direction of the truck movement turned between 178^{o} and 182^{o} and finally, if the average speed during signal loss was lower than the average speed before and after the signal loss, then a trip end had occurred (Trip ends during signal loss were detected by comparing the average speed during the gap to the measured speed before and after the gap). Based on these rules, they could determine trip ends with an error about 5% in failure to detect real trip ends.

Ma, McCormack and Wang (2011) developed an algorithm to extract truck trips in order to quantify truck travel characteristics and performance metrics. GPS data were collected for almost 2,500 trucks (traveling in the Puget Sound region, Washington State) from three different vendors. Recorded information for each GPS event included: longitude, latitude, time and date stamp, truck ID and whether the engine status was “on” or “off”. An algorithm was generated to identify origins and destinations of truck movements. For the case of signal loss due to overhead obstructions such as tall buildings, a 5-mph speed limit was chosen, below which a trip had ended in that area. 3 minutes dwell time at the stop location was set as the minimum time for the duration of stop so that the stop would be flagged as a usual destination. Some “abnormal” trips were also eliminated. Among these were extremely short trips, trips with elapsed travel time of zero, trips with extremely high speed and trips which some portion of them occurred outside the study area. The algorithm was validated by using one month of one of the vendor’s data. It included 3 million GPS records resulting in 358,692 O-D trips with 6,443 abnormal trips.

2.3 Previous Studies on Freight Origin-Destination Matrix Estimation

Ogden (1978) took one of the first steps in modeling truck transportation demand. In their research, a single gravity model was used to get the trip distribution for urban truck trips and urban commodity flows using data for Melbourne, Australia. Commercial vehicle surveys were the source of acquisition of the data for this study which included a sample of 10% of all trucks registered in Melbourne metropolitan area. It was found that the gravity model is suitable for studying the urban truck trip distributions. Some analyses on the trip lengths also suggested that the shortest trips were mostly comprised of localized wholesale and retail delivery trips, as opposed to industrial delivery, construction and maintenance trip purposes which were identified as the longest trips.

Tamin & Willumsen (1989) used three types of travel demand models (Gravity, Opportunity and Gravity-Opportunity). These models were calibrated using observed traffic counts by developing three estimation methods (non-linear-least-squares, weighted- non-linear-least-squares and maximum likelihood). These methods were tested using the 1978 Ripon (urban vehicle movement) survey. All-or-nothing and "stochastic" Burrell's methods were used for the assignment and identification of routes taken in the network. Some of their conclusions include:

- The level of accuracy of the estimated O-D matrix depends on some factors such as : the estimation method, trip assignment techniques, errors in traffic counts.
- The gravity model gives the best fit with the traffic counts. Also non-linear-least-squares method results in the best estimation.
- More realistic assignment methods should be selected to obtain a better and more precise estimated O-D matrix. The proposed methods for future work are: capacity-restrained and equilibrium assignment techniques.
- With more traffic counts available, the estimation method converges faster and we will have a more accurate estimated O-D matrix.
- With a more detailed network and a more disaggregated zoning system, a better level of accuracy will be achieved.

Gedeon et al. (1993) dealt with a transshipment problem for freight flows over a multimodal network. The flows are computed to minimize cost and then a corresponding OD matrix is also computed. In this study, the problem is formulated and an algorithm for the solution is proposed. The model and the algorithm were applied to analyze the transportation of coal in Finland and it was concluded that the algorithm is practical and applicable to large scale networks as well.

Zargari and Hamedani (2006) studied a designated region in Iran to develop a realistic methodology to estimate a true OD matrix for freight movements from an initial OD matrix for that region and reproduce the observed traffic. After dealing with and removing the existing errors in the data sources, i.e. Waybill[2] data and traffic count information, an Entropy Maximization model was calibrated to estimate the true freight OD matrix. For the purpose of traffic assignment, 13 stochastic assignment models using Logit formulation were developed to reflect the special route choice patterns of truck drivers in intercity networks. At the end, the Waybill initial matrix for truck movements was updated using the calibrated assignment models and a calibrated entropy function.

During a study on synthesizing OD matrices by using secondary data sources such as traffic counts, Holguin-Veras and Patil (2007) used a gravity model combined with a commodity-based model to estimate loaded truck trips. They also developed a complementary model for empty trips. The model was later applied to a case study in Guatemala with available actual OD matrix and traffic counts. They considered two objective functions based on two scenarios, where in one scenario only the total link traffic was known, while in the other, only the split of loaded and empty link traffic was available. The results of the case study suggested that the proposed model performed notably better than when the empty trip model is not considered. Also it was found that the OD estimate results were improved by the additional information on observed empty trips.

Giuliano et al. (2010) presented a method for estimating intra-metropolitan freight flows on a highway network. Specifically, a model was developed for this purpose to address truck flows and later was run to estimate truck flows in the Los Angeles region. Actual truck count data collected by the California Departent of Transportation (CalTrans) and Southern California Association of Governments (SCAG) was compared to the estimated truck counts obtained from the model. The results show that the screenline estimates are reaonably close to the actual

[2] "A waybill is a document issued by a carrier giving details and instructions relating to the shipment of a consignment of goods. Typically it will show the names of the consignor and consignee, the point of origin of the consignment, its destination, and route." http://en.wikipedia.org/wiki/Waybill

observed truck counts. This approach was developed with the purpose of transferability to other metropolitan areas, since the data integration and computation is automated. By gathering sufficient data, one can apply this same model to derive truck flows.

Gonzalez-Caldero, Holguin-Veras and Ban (2012) performed a study to synthesize tour-based freight OD flows. This study includes some of the methodologies and modeling approaches used in OD matrix estimation problem. The methods for OD synthesis are categorized into two different groups: structured and unstructured. Structured approach uses traffic counts, while the unstructured approach uses methods such as entropy maximization, information minimization or maximum likelihood. Origin-destination synthesis can be looked into from the aspect of modeling approaches as well. Three major approaches in this aspect are considered:

- Traffic modeling based approaches which use minimum information to estimate OD matrices where the the parameters of the entropy measure are estimated.
- Statistical inference approaches which incorporate maximum likelihood, generalized least squares and Bayesian inference approaches. This approaches presume that some probability functions are used to generate traffic volumes and sample surveys are the source of acquiring the initial (target) OD matrix
- Gradient based solution techniques which are used to estimate OD matrices in a larger scale based on the equilibrium assignment. This method is used in this thesis.

As another step during the Indiana Statewide Truck Travel Demand Model mentioned in the previous section, several scenarios were tested to obtain the best fit of truck flows in such a way that the trips in the resulting estimated OD flow matrix, when assigned to the highway network, closely match the observed truck counts at various locations on the network. Some of the findings from examining these scenarios are:

- Including all the traffic flow (auto trips in addition to ATRI truck trips) during the assignment process results in better ODME results.
- RMSE values to compare estimated assigned flows with observed flows improve by increasing the assignment convergence criteria.

Multi-mode assignment (MMA) procedure was used for ODME in this project. The input trip table for trucks included the non-freight Quick Response Freight Manual (QRFM) single and multi-unit truck trips estimated by the ISTDM in addition to the ATRI data (as freight truck trips). The estimation resulted in almost 3 million freight and non-freight truck trips per week.

CHAPTER 3: CONVERSION OF ATRI TRUCK-GPS DATA TO TRUCK TRIPS[3]

3.1 Introduction

As discussed in chapter 1, ATRI data comprises large streams of GPS records of truck movements in North America. The raw GPS data streams from ATRI need to be converted into a truck trip format to realize the full potential of the data for freight flow analysis, modeling and planning applications. Development of such a truck trip database involves the determination of truck starting and ending instances and locations, trip distance, total trip duration, and duration of intermediate stops (e.g., at traffic signals and rest stops) in the trip. Doing so requires separation of valid pickup/delivery stops from congestion stops, stops at traffic signals, and stops to meet hours of service regulations[4], making use of land-use information and GIS analysis tools along with carefully considered assumptions. In addition, the process involves resolution of potential anomalies in GPS data, such as data-discontinuities due to loss of satellite signals. This chapter describes the algorithms and procedures developed in the project to convert the raw GPS streams provided by ATRI data into a database of truck trips. The next section provides a brief description of ATRI's truck GPS data used in the project. Section 3.3 describes the algorithms and procedures. Section 3.4 presents the results from the algorithms.

[3] The work described in this chapter is majorly done by Aayush Thakur, former graduate research assistant at the University of South Florida, currently working as travel demand modeler at Cambridge Systematics, Inc.

[4] Long-haul drivers make en-route stops for long durations to rest before resuming further driving. Hours of service regulations in 2010 by the Federal Motor Carrier Safety Administration (FMCSA) required a 11-hour daily driving limit. Specifically, Truck drivers were allowed to drive up to 11 consecutive hours driving only after10 hours of off-duty (or rest). Therefore not all stops of longer duration are valid pickup/delivery stops. See http://www.fmcsa.dot.gov/regulations/hours-of-service for more information on FMCSA's hours of service regulations for different years.

3.2 ATRI's Truck GPS Data

In this project, ATRI provided more than 145 million raw GPS records of truck movements from four months – March, April, May, and June – in 2010 for the state of Florida. Specifically, for each of these four months, all trucks from ATRI's database that were in Florida at any time during the month were extracted. Then the GPS records of those trucks were extracted for time periods ranging from two weeks to an entire month, as they traveled within Florida as well as in other parts of North America. This allows the examination of truck movements within Florida as well as truck flows into (and out of) Florida from (to) other locations in the nation.

Each GPS record contains information on its spatial and temporal location along with a unique truck ID that does not change across all the GPS records of the truck for a certain time period varying from a day to over a month (at least two weeks for most trucks in the data). In addition to this information, a part of the data, i.e., the GPS data for some trucks, contains spot speed information (i.e., the instantaneous speed of the truck for each GPS record), and another part of the data does not contain spot speed information. In the rest of this report, the former type of data is called data with spot speeds and the latter type is called data without spot speed. The data with spot speeds and the data without spot speeds were separately delivered to USF, presumably because they come from different fleets of trucks based on the type of GPS units/technology used to monitor the trucks. The frequency (i.e., ping rate) of the GPS data streams varies considerably, ranging from a few seconds to over an hour of interval between consecutive GPS records. Table 3.1 below shows the distribution of the time gap between consecutive GPS readings in one week of data during the month of May 2010. While a large proportion (79.7%) of data with spot speeds comprises GPS streams at less than 15-minute interval, a considerable proportion (29.5%) of data without spot speeds has GPS streams at greater than 1-hour interval.

Table 3.1 Distribution of Time Gap between GPS Readings in a Week of Truck-GPS Data

	% of consecutive GPS readings in the data	
	Data with spot speeds	Data without spot speeds
< 1 minute	12.4%	21.5%
1 to 5 minutes	25.5%	15.0%
5 to 10 minutes	17.8%	8.5%
10 to 15 minutes	24.0%	4.9%
15 to 30 minutes	12.4%	9.7%
30 minutes 1 hour	2.8%	10.9%
1 hour to 2 hours	1.7%	27.8%
> 2 hours	3.4%	1.6%

For each GPS record, ATRI extracted and provided to USF information on how far the location is from the nearest interstate highway. In addition, ATRI shared with USF a GIS shape files containing polygons of major truck stops (such as rest stop areas, weigh stations, welcome centers and wayside parking) within and outside Florida.

3.3 Algorithm Description

The overall procedure to convert ATRI's truck-GPS data into a database of truck trips can be described in the following five broad steps. Each of the broad steps is detailed in this section.

1) Clean, read and sort the GPS data for each truck ID into a time series, in the order of the date and time of the GPS records.
2) Identify stops (i.e., trip-ends or trip origins and destinations) based on spatial movement, time gap, and speed between consecutive GPS points.

a. Derive a preliminary set of trips based on a minimum stop dwell-time buffer value (i.e., eliminate stops of duration less than dwell-time buffer value). Use 30 minutes dwell-time buffer in the beginning.
b. Join insignificant movements (< 1 mile) to preceding trips or eliminate them.
c. Eliminate poor quality trips based on data quality issues such as large time gaps between GPS records and incomplete trips at study period edges.

3) Eliminate trip-ends in rest areas and other locations that are unlikely to involve a valid pickup/delivery.
    a. By overlaying trip ends on a shape file of rest areas, wayside stops, and similar locations.
    b. By eliminating stops in close proximity of interstate highways, which are most likely to be rest areas or wayside parking stops.
    c. Join consecutive trips ending and beginning at such stops.
4) Find circular (i.e., circuitous) trips based on the ratio between airfield distance to network distance. And use raw GPS data between the origin and destination of circular trips to split them into appropriate number of shorter, non-circular trips by allowing smaller dwell-time buffers at the destinations. To do this, implement step 2 with a smaller dwell-time buffer (15 minutes) and go through steps 3 and 4 to find any remaining circular trips. Repeat the process with a dwell-time buffer of 5 minutes to split remaining circular trips.
5) Conduct additional quality checks and eliminate trips that do not satisfy quality criteria.

3.3.1 Clean and Sort Data

The raw GPS data is first screened for basic quality checks such as the presence of spatial and temporal information and the presence of at least one day of data for each truck ID. Truck IDs that do not have GPS data for at least a span of one day or that have too few GPS records are removed. For such trucks, it is difficult to extract trips because most of the data is likely to be lost in the form of trips in progress without origin and/or destination in the data. The cleaned data is then sorted in a time series for each truck beginning from the GPS record with the earliest date and time stamp.

3.3.2 Identify Truck Stops (i.e., truck trip-ends) to Generate Truck Trips

This step comprises a major part of the procedure to convert raw GPS data into truck trips. The high level details of the algorithmic procedure in this step are presented in Figure 3.1. Below is a list of the terms used in the algorithm along with their definitions:

- Travel distance (td): Spatial (geodetic) distance between two consecutive GPS records.
- Travel time (trt): Time gap between the two consecutive GPS records.
- Average travel speed (trs): Average travel speed between consecutive GPS records (td/trt).
- Trip length (tl): Total distance traveled by the truck from origin of the trip to the current GPS point. This becomes equal to trip distance, when the destination is reached.
- Trip time (tpt): Total time taken to travel from origin of the trip to the current GPS point.

- Trip speed (tps): Average speed of the trip between the origin and the current GPS point.
- Origin dwell-time (odwt): Total time duration of stop at the origin; i.e. when the truck is not moving (the wait time for the truck before starting its trip)
- Destination dwell-time (ddwt): Total duration the truck stops at the destination of a trip
- Stop dwell-time (sdwt) – Duration of an intermediate stop (e.g., traffic stop).
- Total stop dwell-time (tsdwt): Total duration at all intermediate stops during the trip.

The first three terms – td, trt, and trs – are measures of movement between consecutive GPS data points. The next three terms – tl, tpt, and tps – are measures of total travel between the trip origin and the current GPS data point. When the truck destination is reached, these measures are for the entire trip beginning from its origin to the destination. The last four terms – odwt, ddwt, sdwt, and tsdwt – are dwell-times (i.e., stop durations) at different stages during the trip. odwt is the dwell-time at the origin of a trip, ddwt is the dwell-time at the destination of the trip, sdwt is the dwell-time at an intermediate stop (e.g., traffic stop) that is not the destination of the trip, and tsdwt is the sum of dwell-time at all intermediate stops during the trip.

For each truck ID, the algorithm begins with reading its first GPS record and initializing all the terms – td, trt, trs, tl, tpt, tps, odwt, ddwt, sdwt, and tswdt. Then the algorithm reads the next record and computes average travel speed between the two records to verify if the truck is moving or if it is at rest. The subsequent parts of the algorithm are described below.

3.3.2.1 Determining Truck Stops and Moving Instances

An important component of the algorithm involves determining whether a truck is at stop (i.e., in rest) or in motion. As can be observed from the flowchart, the primary condition used to

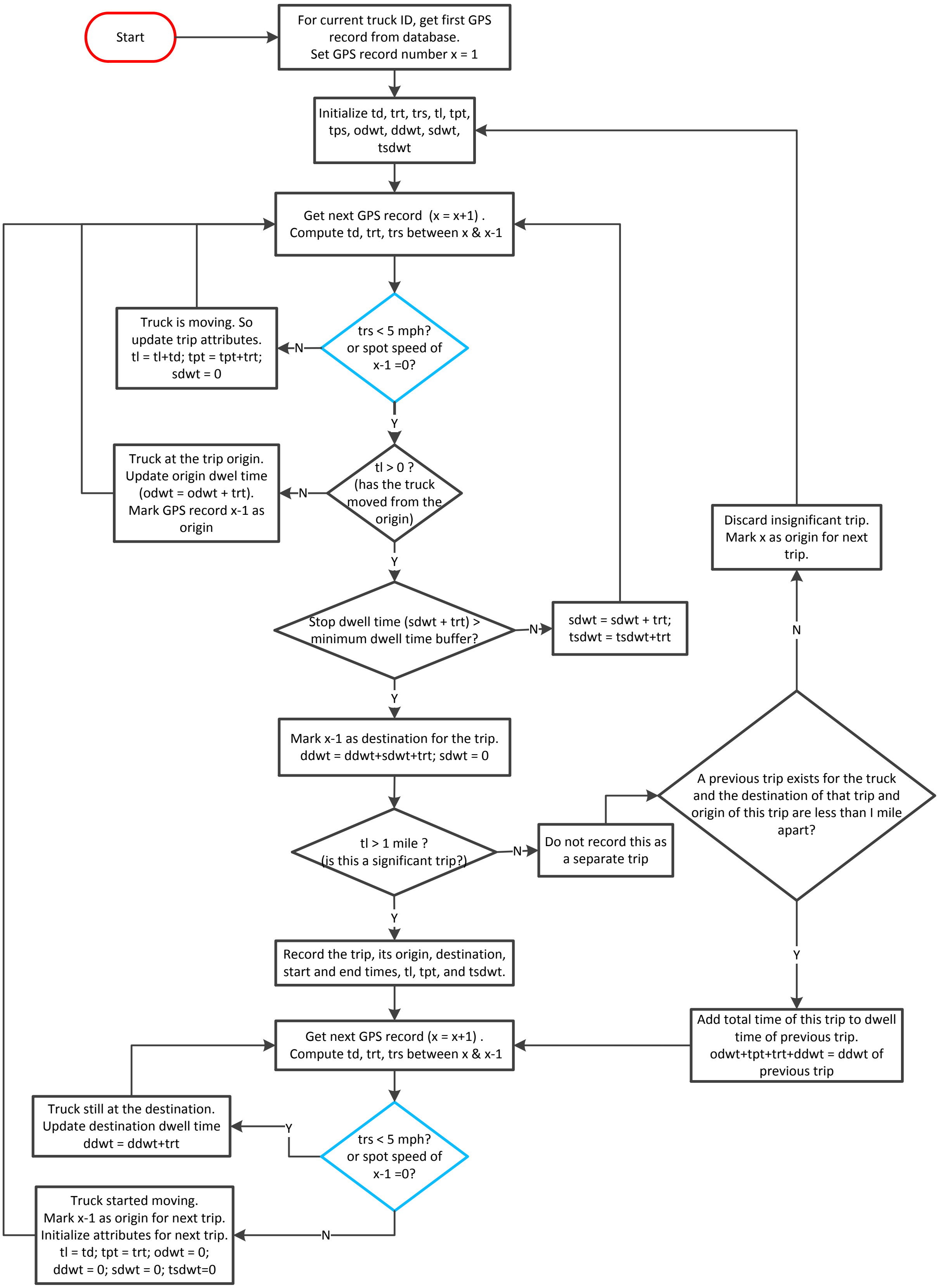


Figure 3.1 Algorithm for Identifying Truck Trip Ends from Raw GPS Data

determine whether a truck is at stop (which could be an origin, a destination, or simply an intermediate stop) or it is moving is based on the average travel speed between consecutive GPS data points. A cut-off speed of 5mph is used; if the average travel speed between consecutive GPS records is less than 5pmh (i.e., if trs < 5mph), then the truck is assumed to have stopped. As mentioned earlier, part of the data contains spot speeds (i.e., instantaneous speeds) and another part of the data does not contain information on spot speeds. We used the data with spot speeds to test different cut-off values on the travel speed between two consecutive GPS data points. Besides, other recent studies that converted ATRI's truck GPS data into trips (MAG Truck Model Update and Indiana Statewide Travel Demand Model Truck Model Update mentioned in chapter 2) also used a 5mph cut-off to determine whether a truck is moving or if it is at rest.

For data with spot speeds, we used the average travel speed criterion (i.e., if trs < 5mph) as well as checked if the spot speed was zero. If one of these two criteria is satisfied, the truck is assumed to be at rest. It is important to check travel speed between consecutive GPS data points (i.e., if trs < 5mph) even in data with spot speeds. When the time gap between two consecutive GPS data points that are moving (i.e., spot speed > 0) is large enough, the only way to determine if the truck stopped between two points is by checking its travel speed between the two points. Limited tests on data with spot speeds suggested that it was sufficient to use the average travel speed criterion (i.e., without checking for spot speeds), increasing our confidence in the quality of trips derived from the data without spot speeds.

An alternative approach to determine truck stops is to use a minimum distance criterion (i.e., assuming the truck to be at stop if it did not move beyond a distance cut-off). However, since the time gap between consecutive GPS records varies considerably – ranging from seconds to several hours – it is difficult to use a single distance cut-off for determining truck stops.

3.3.2.2 Separating Intermediate Stops from Trip Destinations

Checking for average travel speeds between consecutive GPS records helps in identifying if the truck is at rest or moving. However, truck being at rest doesn't necessarily indicate whether it is at a valid origin or destination or if it is at an intermediate stop. The intermediate stops could be due to a number of reasons, including stops at traffic signals and traffic congestion (these stops typically tend to be of a few minutes duration), stops at gas stations for refueling purposes, wayside stops for drivers' quick relaxation and other purposes such as restroom visits, and rest stops of long duration to comply with hours of service.

To identify and eliminate many such intermediate stops, once the algorithm detects a stop (i.e., trs < 5mph or spot speed =0) for a trip in progress (i.e., tl > 0), it starts updating the stop dwell-time (sdwt) based on the time elapsed between successive GPS data points (sdwt = sdwt + trt). If the truck starts moving again (i.e., if trs >5mph or spot speed > 0) before the stop dwell-time reaches a minimum dwell-time buffer value, then the stop is considered an intermediate stop and the algorithm proceeds to find another stop. On the other hand if the stop dwell-time exceeds the minimum dwell-time buffer value, then the stop is considered a candidate for valid destination, and the stop dwell-time (sdwt) is considered as part of the destination dwell-time (ddwt), by updating ddwt as sdwt + trt. Subsequently, if the length of this trip (tl) is greater than 1 mile (if not it is considered an insignificant trip and not recorded), the trip is recorded, along with its origin, destination, start/end times, trip length, and the total time the truck stopped at all intermediate locations between origin and destination (tsdwt).[5] The destination dwell-time of the trip is then updated using subsequent GPS records (i.e., ddwt = ddwt + trt) until the truck starts moving again (i.e., trs > 5mph or spot speed > 0). Once the truck starts moving (see the bottom

---

[5] Most previous studies do not record the total stop dwell-time (tsdwt) – a useful attribute for truck trips, especially for understanding how the duration of intermediate stops varies with trip length and other relevant factors.

part of the flowchart in Figure 3.1), the origin of a potential next trip is marked and the algorithm proceeds to find the next truck stop, as can be observed from the looping of the bottom part of the algorithm to the top part (on the left side of the figure).

A major determinant of the quality of outputs from the above procedure depends on the dwell-time buffer (i.e., the minimum stop duration required for a truck stop to be a destination instead of an intermediate stop). Some previous studies (Ma, McCormack and Wang, 2011) consider stops of less than 3 minute duration as intermediate stops while other studies (MAG Truck Model Update and Indiana Statewide Travel Demand Model Truck Model Update mentioned in chapter 2) considered stops of less than 5 minute duration as intermediate stops. In this study, however, based on discussions with ATRI and our own tests following individuals trucks on Google Earth (i.e., by observing the land-uses of stops made by trucks and the duration of those stops), we used larger dwell-time buffers. This is because the focus of this research was to extract tuck OD flows over long-haul distances of concern for Florida's statewide freight model. While most intermediate stops in urban areas tend to be of smaller duration (e.g., most traffic signal cycles tend to be of less than 3 minute duration), not all stops of larger duration tend to be for commodity pickups and/or deliveries. Intuition suggests that 5 minutes is not sufficient for picking up or delivering many types of goods. Besides, refueling stops, stops at weighing stations, and quick relation or restroom stops tend to be longer than 5 minutes. And of course, truck stops for the purpose of complying with hours of service regulations tend to be of several hours duration.

To arrive at an appropriate dwell-time buffer (i.e., the maximum stop duration within which all stops are considered as intermediate stops), we compared the trip outputs using different values of dwell-time buffer – (5 minutes, 10 minutes, 15 minutes, 30 minutes, 45 minutes, and an hour – with the land-uses of the trip ends in Google Earth. Dwell-time buffers of short durations such as 5 minutes or 10 minutes were resulting in false identification of too many intermediate stops as truck origins and/or destinations, while dwell-time buffers of too long durations led to missing valid origins/destinations at pickup/delivery locations such as distribution centers. Further, using small dwell-time buffers was leading to a large share of short-length trips (because a long trip between an origin and destination was broken down into several short trips). Besides, for low frequency data where the consecutive GPS records have large time gaps, small dwell-time buffers cannot be relevant (for example, testing for a 5-minute dwell-time would not give different results than testing for a 15-minute dwell-time buffer if the GPS records are spaced 15 minute intervals). After testing for different values of dwell-time buffers, it turned out that no dwell-time buffer value was perfect and a trade-off had to be made between minimizing false identification of unnecessary intermediate stops and skipping of valid origins and destinations. After extensive tests via following trucks on Google Earth, 30 minutes dwell-time buffer was used as a beginning point to separate intermediate stops from trip destinations.

The 30 minute dwell-time buffer helped in avoiding most intermediate stops, including traffic and congestion stops, wayside stops, and gas refueling stops, and short stops at rest areas. But it does not help eliminate longer duration stops at rest areas, including those made to comply for hours of service regulations. Another issue is that the 30 minute dwell-time buffer leads to skipping of some valid origins or destinations that involve smaller dwell-time buffers. These two issues are addressed latter.

3.3.2.3 Dealing with Insignificant Trips

Trips that were too short in length were not recorded as independent trips. The minimum acceptable trip length was assumed to be 1 mile. Therefore, if a trip is less than 1 mile, it was discarded unless the trip occurred in the same area of the previous trip's destination. In this case the insignificant trip's time was simply added to the previous trip's destination dwell-time (see the bottom right part of the algorithm in Figure 3.1). For example, if the destination of a trip is large in size (like a port), it might happen that the truck moves within the port for less than a mile leading to insignificant trips. Such movement was not considered a new trip but, since the truck would still be at the same destination as the previous trip (port), but incorporated into the previous trip's destination dwell-time.

3.3.2.4 Quality Control Checks in the Algorithm

Figure 3.1 does not present all details of the algorithm to make it easier for readers to understand the main components of the algorithm and for ease in presentation. These details include the following quality checks embedded into the algorithm.

Dealing with large time gaps between consecutive GPS records: The ping rate in the data (i.e., the frequency at which GPS positions are recorded) varies considerably, ranging from a few seconds to several hours. Data with large time gaps between consecutive records could be due to many reasons, including loss of GPS signals (e.g., in tunnels and mountains) and malfunctioning of the GPS device. In such cases, the extracted trips tend to be of lower quality because it is difficult to use only the spatial and temporal movement information to ascertain what happened in the time gap. On the other hand, it is also possible that some GPS units (depending on the type of equipment) may not record truck positions for an extended time period simply because the truck engine is switched off. In such situations, the truck is simply not making any movements for an extended time period. Therefore, if the time gap between two consecutive records was greater than 2 hours and if the trip was in progress (i.e., the travel speed between the two records was greater than 5mph), such a trip was discarded. However if the speed was less than 5mph, then the truck was assumed to be at rest (i.e., not moving) for the entire time gap.

Trips at the edges of the time periods for which data is available: It is not necessary that the GPS records of a truck begin with a trip origin and end with a trip destination. For many trucks the first several records indicate the truck is in motion (because the trip started before the first available GPS record for the truck) and/or the last records belonged to a trip that ended after the last GPS record. Such incomplete trips found at the edges of study periods were discarded.

Dealing with the last record of a truck ID: For each truck ID, the algorithm runs until all the GPS records of the truck ID are exhausted. When the algorithm reaches the last GPS record of the truck ID, the last trip of the truck is either retained or eliminated depending on whether the trip has completed (i.e., the trip is at its destination) or still in progress. This is determined by the average travel speed between the last record and its previous record. Subsequently, the algorithm moves on to the next truck ID.

The above quality checks were implemented in the algorithm every time a GPS record is read and the travel distance (td), travel time (trt), and average travel speed (trs) are computed between consecutive records. In addition to the above quality controls embedded within the algorithm, quality checks were conducted on the trips output at the end of the procedure. Specifically, trip speeds, trip time, and distance were examined for manifestations of any anomalies such as GPS jumps and jiggles. GPS jumps happen when GPS records show unrealistically large movements within short durations, which manifest as trips with unrealistically large speeds. Such trips were eliminated based on average trip speeds and travel time. Specifically, we retained only those trips within an average speed of 80 mph (between origin and destination). Further trips that were too short in time (i.e., trips of travel time less than a minute) were also removed.

3.3.3 Eliminate Trip Ends in Rest Areas

The above described algorithm eliminates unwanted trip ends (such as traffic stops and refueling stops) to some extent. However, the algorithm does not eliminate unwanted trip ends in rest areas and other locations (e.g., wayside stops) with dwell-times larger than the dwell-time buffer used in the algorithm. To address this issue, the trip ends derived from the above step are overlaid on a GIS shape file of rest stops provided by ATRI containing polygons of rest areas, commercial truck stops, weigh stations, wayside parking, etc. throughout the nation. All the trip ends falling in these polygons were eliminated by joining consecutive trips ending and beginning in those polygons.

Doing the above helped in eliminating a large number of unwanted trip ends in rest areas, wayside parking areas, and other such locations. However, further scrutiny suggested that a good portion of trip ends were still in rest areas and similar locations. This is because the data in the shape file of rest stops provided by ATRI is not necessarily exhaustive of all rest areas and other such stops (not for pickup/delivery) in the nation. To eliminate the remaining trip ends in rest areas, the research team used information on the distance of each trip end from the nearest interstate highway. When a random sample of 200 rest areas from the shape file provided by ATRI were examined vis-à-vis their distance from the highway network, a vast majority of the rest areas were found in very close proximity to interstate highways (45% of them were within 800 feet distance of an interstate highway). Therefore, we treated all trip ends within a buffer of 800 feet from interstate highways as stops at rest areas or wayside parking areas. Any consecutive trips ending and beginning in the same location were joined to form a single trip.

The natural next question is how we arrived at the 800 feet value. We examined if treating all trip ends within a close proximity of interstate highways as rest stops helps in removing additional unwanted stops while not removing true origin or destination stops. To examine this, we traced the GPS records of 40 trucks for at least two weeks on Google Earth and observed the land-uses of their stop locations. For each of these 40 trucks, the number of valid trip ends noticed in Google Earth were recorded and then compared with the number of trips output from the algorithm after eliminating trip ends that lied in the rest stops polygons of the shape file provided by ATRI and those that lied within a given proximity of interstate highways. We tested this for different buffers around interstate highways – half mile, quarter mile (1320 feet), 1000 feet, 800 feet, and 500 feet. As expected, no single buffer was ideal; using a large buffer led to elimination of too many valid origins/destinations while a small buffer led to the presence of too many invalid origins/destinations such as rest areas or wayside parking areas. However, using 800 feet provided a good trade-off between losing valid origins/destinations and counting invalid origins/destinations.

3.3.4 Find Circular Trips and Split Them into Shorter Valid Trips

Recall from Section 3.3.2.2 that a 30-minute dwell-time buffer was used to separate intermediate stops from valid trip destinations. Among the other dwell-time buffers examined, the 30-minute buffer helped strike a good balance between removing intermediate stops (such as traffic signal stops, congestion stops, and refueling stops) and skipping valid destinations (of less than 30 minutes dwell-time) en-route. However, it would be useful to recover such valid destinations of less than 30 minutes.

When we examined the trip outputs from the algorithm with 30-minute dwell-time buffer, some trips had origins and destinations that were too close to each other although the network distance of those trips measured in the algorithm (using GPS data) was large. Some trips, for example, had their origin and destination in the same location, although the distance traveled by the truck on the network was large. A major reason why this was happening was because the algorithm was skipping some valid destinations (of less than 30 minutes dwell-time) en-route.

One way to identify if a trip extracted from the algorithm has any valid destinations en-route that have been skipped is to check its circuity ratio. The circuity ratio of a trip is the ratio between the airfield distance between its origin and destination and the network distance between its origin and destination. The value of the circuity ratio can range from 0 to 1. If a truck travels in a straight line between its origin and destination, its circuity ratio would be 1. However, most trips on the highway transportation network tend to travel more than the airfield distance between the origin and destination. At the same time, if the ratio is too small, there is a high chance that the truck stopped en-route at valid destinations, albeit for shorter durations than 30 minutes. Intuitively, it is unlikely that trucks detour significantly between the origin and destination only for the sake of traveling to intermediate rest stops. Figure 3.2 shows an example of such a trip whose origin and destination locations (marked by blue circles) are too close to each other, although the distance traveled by the truck on the network along the route shown in red in the figure is more than 100 miles. In this example, the truck stopped at three other locations between the origin and destination (marked by yellow circles) for less than 30 minutes. The land-uses of these stops, when examined on Google Earth, were all valid destinations such as warehouses and large grocery stores. When we examined several other such examples, it became more apparent that trips with a small circuity ratio had skipped en-route stops of duration smaller than 30 minutes and most of these stops were valid destinations. After extensive testing, through tracing raw GPS data of trips with different circuity ratio values, we arrived at a cut-off value of 0.7. All trips extracted from the algorithm with a circuity ratio less than 0.7 were considered to be circular trips with a high chance of a skipped valid destination on-route.

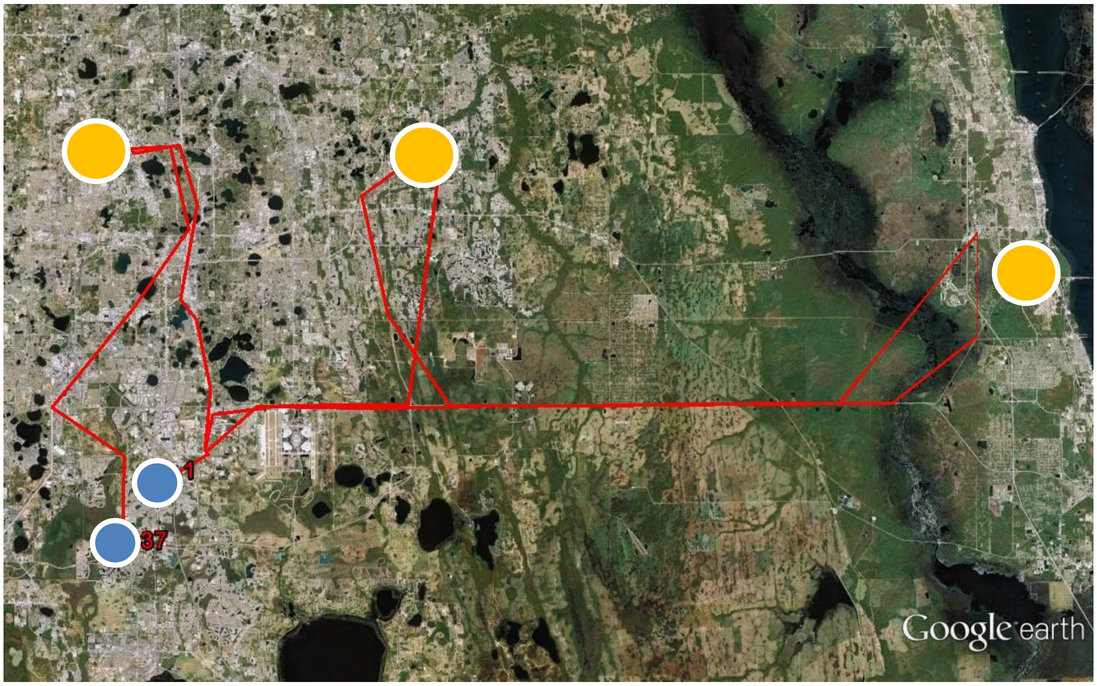


Figure 3.2 Example of a Circular Trip Extracted from the Algorithm with 30-Minute Dwell-time Buffer

The circular trips are then separated for further processing. Specifically, the procedure goes back to the raw GPS data of the trips with a circuity ratio less than 0.7 and re-applies the algorithm in Figure 3.1 with a smaller dwell-time buffer (15 minutes). This helps in splitting the circular trips into multiple potentially valid trips. Specifically, each circular trip is broken into an appropriate number of shorter, non-circular trips by allowing smaller stop dwell-time buffers at the destinations. The trip outputs from this process are again checked for circuity. For any remaining trips with a circuity ratio of less than 0.7, algorithm in Figure 3.1 is reapplied on the corresponding raw data, albeit with a smaller dwell-time buffer (5 minutes) this time. In the above example (in figure 3), this iterative process would result in four separate trips, each with a circuity ratio greater than 0.7, instead of a single trip with a small circuity ratio.

An example of the results from this procedure is in order here. Applying the algorithm in Figure 3.1 with a dwell-time buffer of 30 minutes and the subsequent step (of eliminating trip ends in rest areas and in close proximity of interstate highways) to one month of ATRI's GPS data – May 2010 – resulted in a total of 252,000 trips. Out of all these trips, 183,000 (72.6%) had circuity ratios greater than or equal to 0.7. After splitting the remaining 69,000 trips (with circuity ratio < 0.7) into smaller trips by reapplying the algorithm in Figure 3.1 with minimum dwell-time buffer of 15 minutes, we extracted about 123,000 trips. About half of these trips had a circuity ratio of at least 0.7. For the other half, we repeated the algorithm in Figure 3.1 with a minimum dwell-time buffer value of 5 minutes. This resulted in over 87,000 trips, of which 38,000 trips had a circuity ratio of at least 0.7. The remaining 49,000 trips, with a circuity ratio smaller than 0.7, were discarded. In all, the final number of trips extracted for the month of May 2010 is 183,000 + 62,000 + 38,000 = 283,000.

The above-described iterative procedure of checking for circuity and splitting circular trips into multiple valid trips using smaller dwell-time buffers helps in two different ways: (a) it helps capture trips with valid destinations of short dwell-times, and (b) it helps remove any remaining circular trips – some of which that are likely to have resulted from joining consecutive trips within a close proximity of interstate highways as described in the previous section.

3.4 Results

The above discussed procedure was applied to four months of raw data (March-June 2010) comprising over 145 Million raw GPS records. This resulted in over 2.7 Million truck trips. Of these, over 1.27 million trips had at least one end in Florida. Table 3.2 shows a summary of the raw data and the trips derived from the data. Summaries are provided for each month of the data, separately for data with spot speeds and data without spot speeds. The number of trips extracted, number of unique truck IDs to which these trips belong, and the average trip distance and trip speeds (without considering duration at rest stops) is presented for three different types of trips – (a) all trips including those outside Florida, (b) FL-link trips (trips with at least one end in Florida), and (c) FL-only trips (trips with both origin and destination in Florida).

Note from the table the trips extracted from data without spot speeds are longer than those from data with spot speeds. For a certain type of data (e.g., for data with spot speeds), the average trip distances and trip speeds are similar across the four months. Besides, the average trip speeds appear to be similar across different datasets and for different months. A detailed analysis of the characteristics of the trips extracted in this project is presented in the next chapter.

The trip outputs from the procedures discussed in this chapter were subject to a variety of quality checks, some of which are discussed here. The land-uses of the origin/destination locations of a random sample of 232 trips extracted from our algorithms were examined on Google Earth. Over 90% of the 464 trip ends were in locations that are highly likely to involve goods pickups/deliveries such as distribution centers, manufacturing companies, industrial areas, ports, retail stores, shopping centers, and agricultural lands. Of the remaining locations, 24 were on highways (that are not interstate highways) without nearby freight-related land-uses, 3 were in rest areas, and 9 were in gas stations. Most of the 24 trip ends on highways were truck stops of longer than 30 minutes. In future work, eliminating truck stops in close proximity of major highways (in addition to interstate highways) can potentially improve the results. For stops in gas stations, however, particularly those greater than 30-minute duration, it is difficult to decipher if they are made for refueling purposes or for fuel delivery services. In addition to trip end locations, we assessed the accuracy of temporal attributes (i.e., trip start and end times). The trip start times output from the algorithm were found to be accurate for over 95% of the trips while the trip end times were accurate for all trips. Overall, while scope exists for improving the algorithms in this chapter (e.g., by utilizing detailed land-use information), the quality of trips extracted suffices for the purpose of estimating statewide TAZ-to-TAZ truck flows.

3.5 Characteristics of the Derived Truck Trips

This section presents an analysis of the truck trip data derived from the four months of ATRI's truck GPS data described earlier. The truck travel characteristics analyzed include trip duration, trip length, trip speed, time-of-day profiles, and origin-destination flows. Each of these characteristics are derived at a statewide level as well as for different regions in the state – Jacksonville, Tampa Bay, Orlando, Miami, and rest of Florida – defined based on the freight analysis framework (FAF) zoning system. Furthermore, the next chapter presents an analysis of Origin-to-Destination (OD) travel distances, travel times, and travel routes between a selected set of TAZ-to-TAZ OD pairs.

Figure 3.3 shows the trip length distribution of over 2.7 Million trips derived from the data. As can be observed, a considerable proportion of the trips are within 50 miles length. Figure 3.4 shows the trip duration distribution of these trips. Two types of trip durations are reported: (1) Total trip time and (2) Trip time in motion. Total trip time is the time between trip start and trip end, including the time spent at rest stops. Trip time in motion excludes the time spent at rest stops and other long-duration stops. Note that trip time in motion includes time at smaller duration (< 5 minutes) stops such as traffic stops to reflect congestion effects. Figure 3.5 shows the trip speed distribution considering the two types of trip times discussed above. Specifically, the average trip speed considers all stops between trip start and trip end, while trip speed in motion excludes stops of longer duration (e.g., rest stops) but considers stops of smaller duration. Similar distributions were generated for trip length, duration, speed, and also time-of-day[6] profiles for different segments of the 2.7 Million trips discussed above. The different segments include trips starting and ending in different FAF zones in Florida – the Jacksonville FAF zone, Tampa FAF zone, Orlando FAF zone, and Miami FAF zone. Below are the specific counties in each of these FAF zones

- Jacksonville FAF-zone: Baker, Clay, Duval, Nassau, St. Johns
- Miami FAF-zone: Broward, Miami-Dade, Palm Beach
- Orlando FAF-zone: Flagler, Lake, Orange, Sumter, Osceola, Seminole, Volusia
- Tampa FAF-zone: Hernando, Hillsborough, Pasco, Pinellas

The distributions were provided separately for weekday and weekend trips. Such distributions can potentially be used for modeling heavy truck trip characteristics within the major regional models in the state. As an example, the truck trip characteristics for the Tampa FAF zone are provided below (Figures 3.6 through 3.9). It is interesting to note that the time-of-day profiles for all the four FAF zones in Florida show a single peak during the late morning period as opposed to a bi-modal peak typically observed for passenger travel for morning and evening peak periods.

[6] Note: Time-of-day of a trip is determined based on the hour in which the midpoint of the trip falls.

Table 3.2 Summary of Truck Trips Extracted from Four Months of ATRI's Truck-GPS Data

		Data with spot speeds			Data without spot speeds			All data		
		All trips	**FL-link trips**	**FL-only trips**	**All trips**	**FL-link trips**	**FL-only trips**	**All trips**	**FL-link trips**	**FL-only trips**
March 2010	Number of GPS Records	13,271,519			25,750,534			39,022,053		
	Number of trips extracted	284,092	145,245	119,602	449,074	195,298	128,178	733,166	340,543	247,780
	Number of unique truck IDs	7,406	6,594	4,815	47,523	39,277	25,979	54,929	45,871	30,794
	Average trip length (miles)	188	135	59	258	225	78	231	187	69
	Average trip time (minutes)	212	162	83	315	286	120	275	233	102
	Average trip speed (mph)	41	37	34	40	38	33	40	38	33
April 2010	Number of GPS Records	12,920,919			22,818,557			35,739,476		
	Number of trips extracted	283,673	144,526	118,288	397,098	175,717	116,647	680,771	320,243	234,935
	Number of unique truck IDs	7,434	6,645	4,848	42,493	35,337	23,786	49,927	41,982	28,634
	Average trip length (miles)	185	135	58	255	223	78	226	183	68
	Average trip time (minutes)	209	162	82	311	283	119	268	228	100
	Average trip speed (mph)	41	37	34	40	38	33	40	38	34
May 2010	Number of GPS Records	13,252,936			21,741,597			34,994,533		
	Number of trips extracted	283,017	145,946	119,359	360,734	159,992	104,148	643,751	305,938	223,507
	Number of unique truck IDs	7,327	6,527	4,676	36,888	30,046	19,287	44,215	36,573	23,963
	Average trip length (miles)	187	134	58	262	230	76	229	184	66
	Average trip time (minutes)	210	161	80	320	291	117	272	229	97
	Average trip speed (mph)	41	37	34	40	38	33	40	38	34
June 2010	Number of GPS Records	13,740,038			21,511,076			35,251,114		
	Number of trips extracted	293,266	148,895	120,950	356,727	156,227	101,513	649,993	305,122	222,463
	Number of unique truck IDs	7,525	6,736	4,882	36,438	29,731	19,113	43,963	36,467	23,995
	Average trip length (miles)	186	135	57	257	225	77	225	181	66
	Average trip time (minutes)	210	161	80	316	287	118	268	226	97
	Average trip speed (mph)	41	38	34	40	38	33	40	38	34
All four months	Number of GPS Records	53,185,412			91,821,764			145,007,176		
	Number of trips extracted	1,144,048	584,612	478,199	1,563,633	687,234	450,486	2,707,681	1,271,846	928,685
	Number of unique truck IDs	13,087	11,728	8,416	156,627	128,275	83,443	169,714	140,003	91,859
	Average trip length (miles)	186	135	58	258	226	77	227	184	67
	Average trip time (minutes)	210	162	81	315	287	119	271	229	99
	Average trip speed (mph)	41	37	34	40	38	33	40	38	33

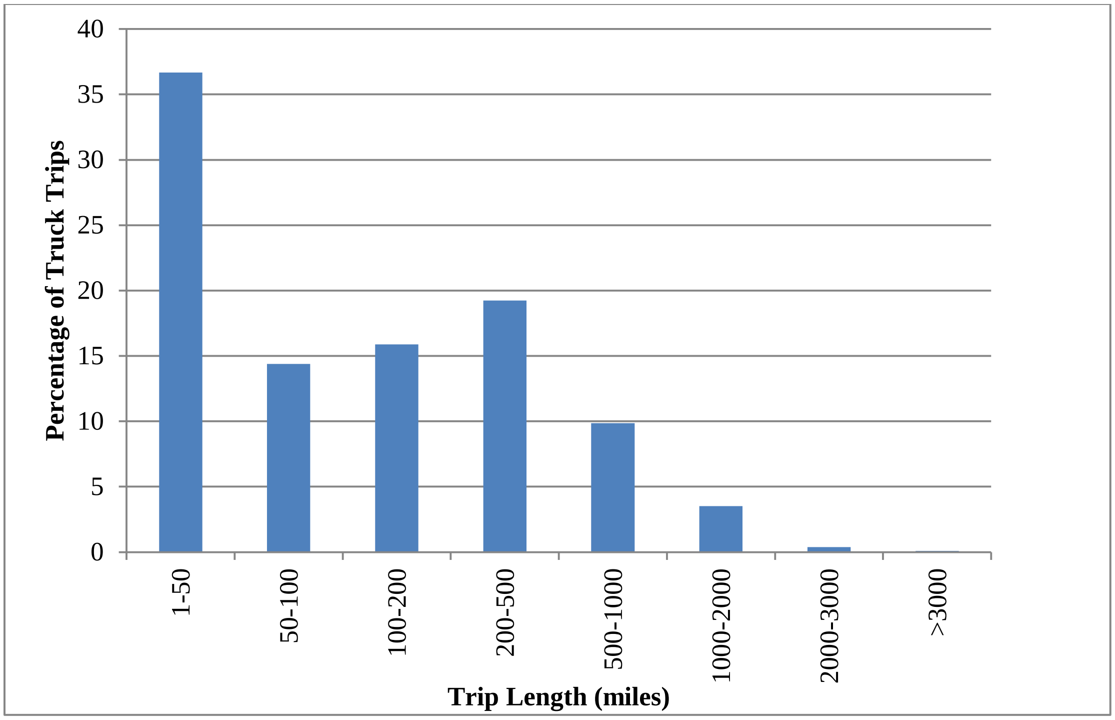


Figure 3.3 Trip Length Distribution of All Trips Derived from Four Months of ATRI Data

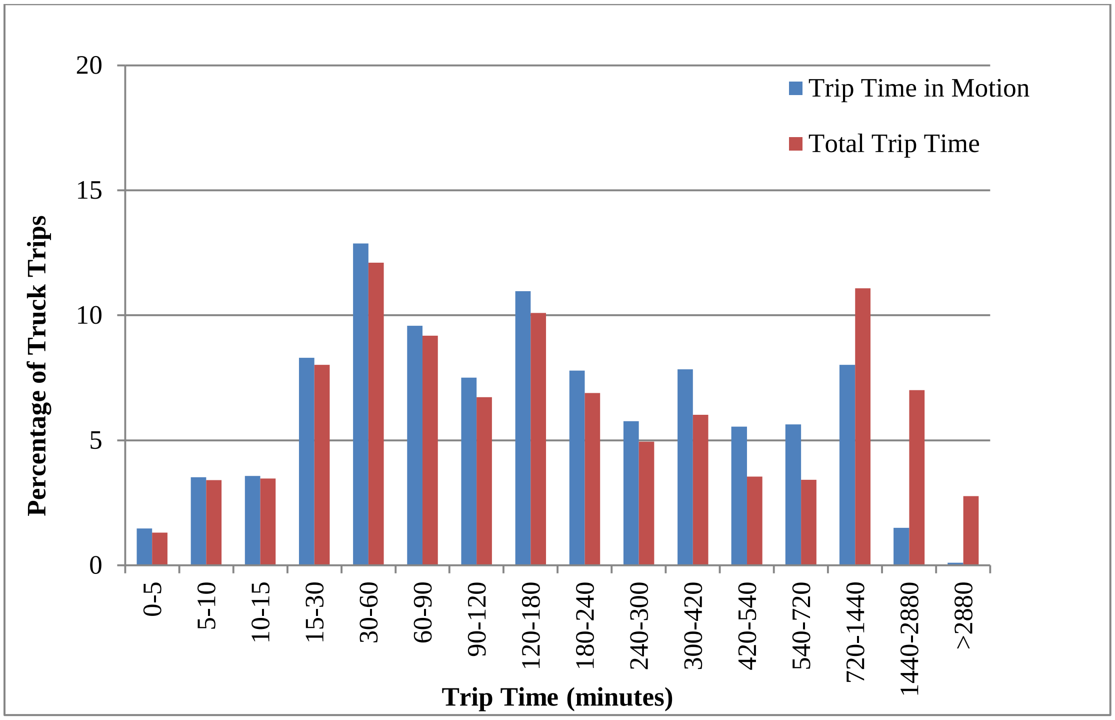


Figure 3.4 Trip Time Distribution of All Trips Derived from Four Months of ATRI Data

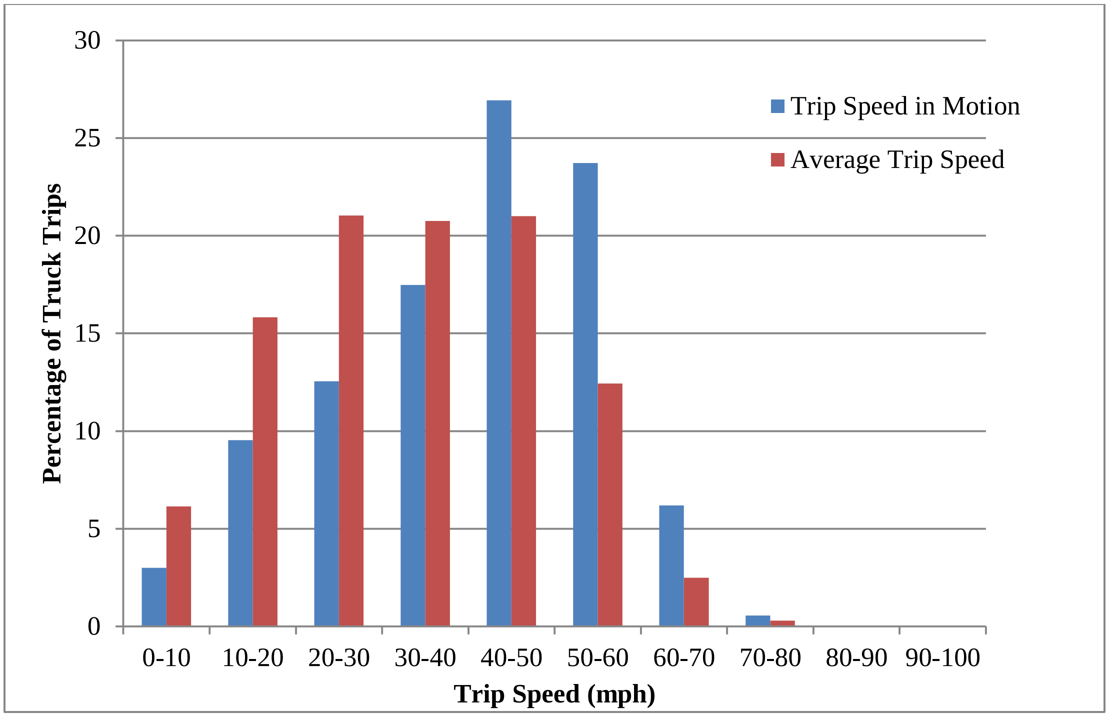
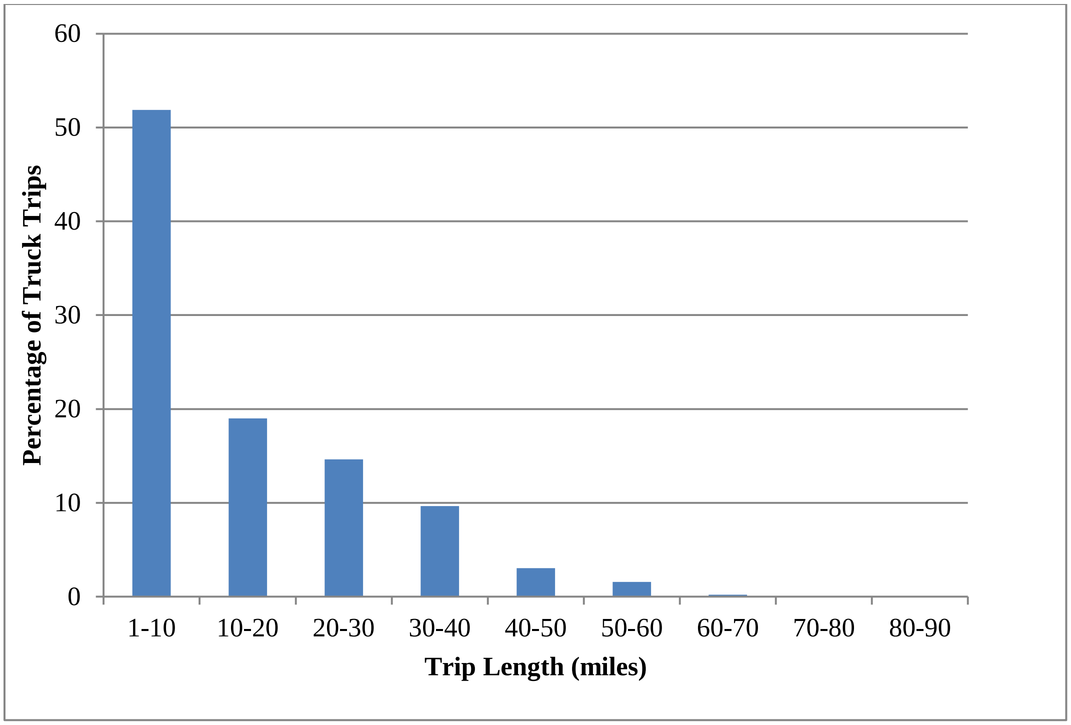
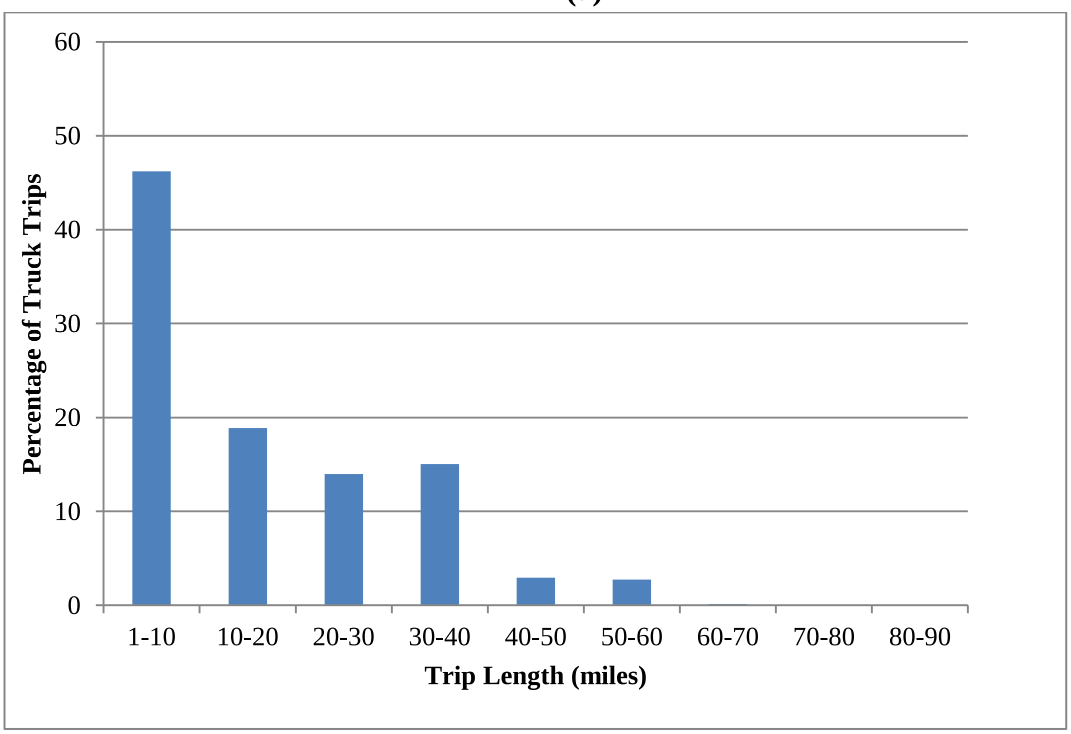


Figure 3.5 Trip Speed Distribution of All Trips Derived from Four Months of ATRI Data

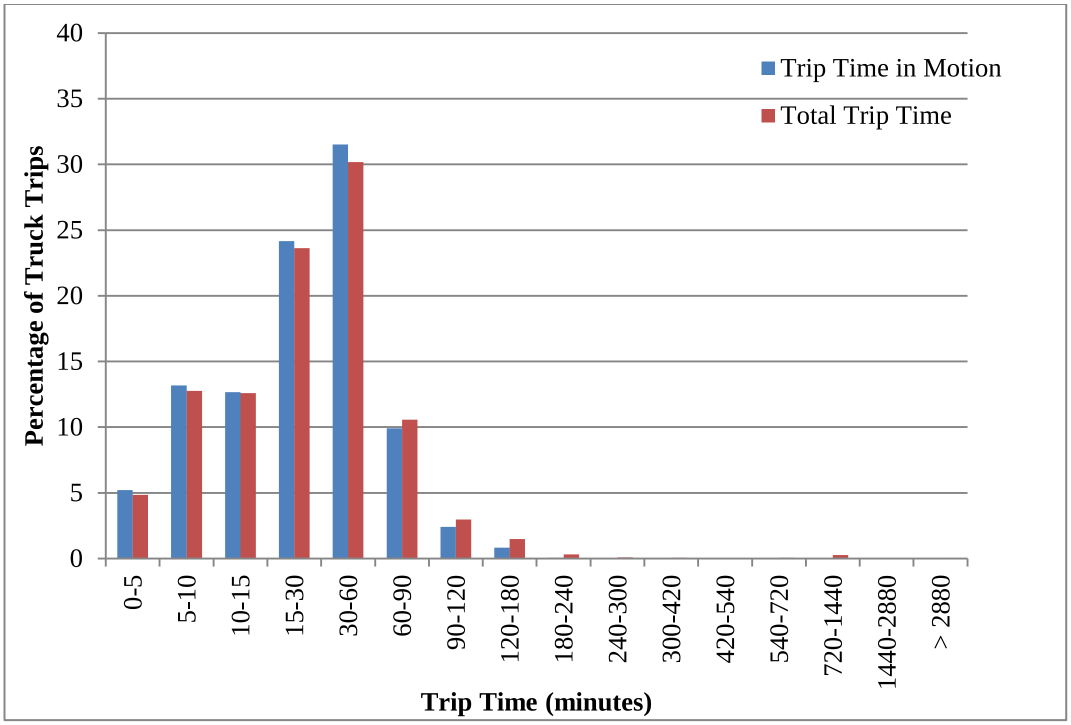


(a)

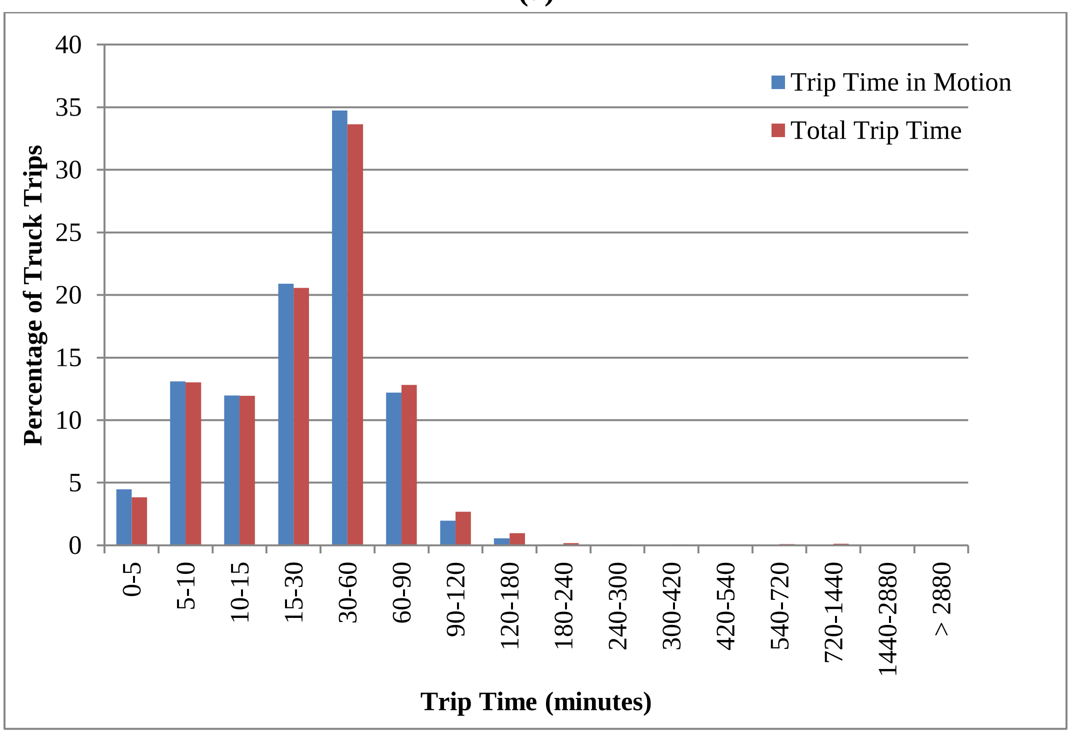


(b)

Figure 3.6 Trip Length Distribution of Trips Starting and Ending in Tampa FAF-Zone during: (a) Weekdays (61,465 trips), and (b) Weekend (7,780 trips)

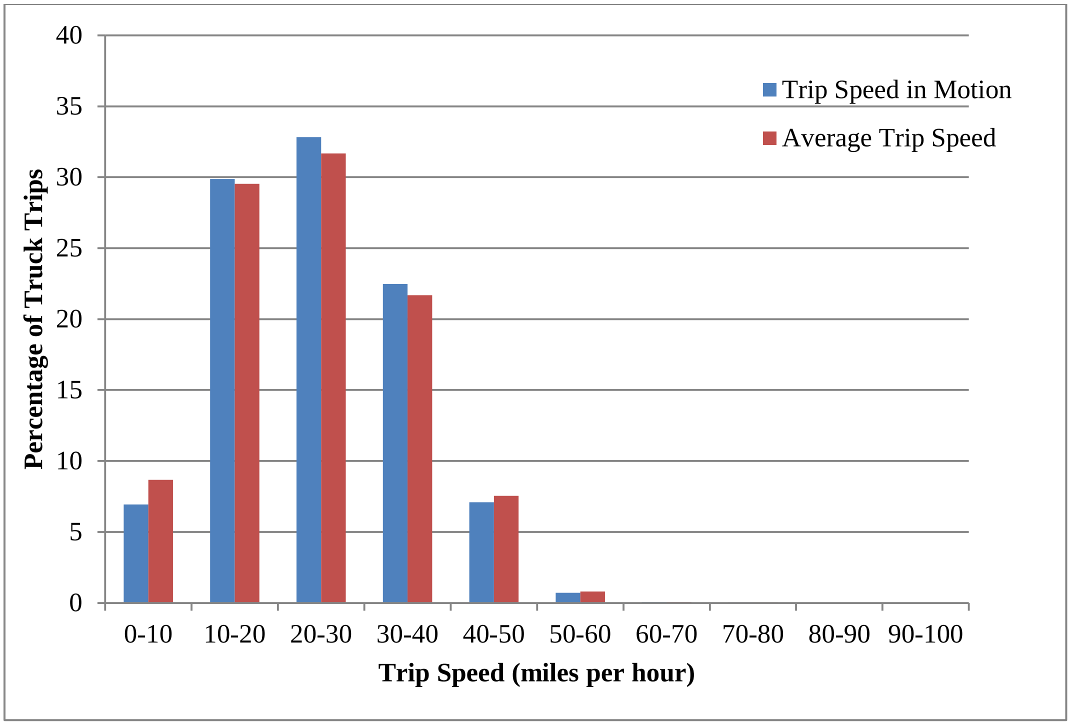


(a)

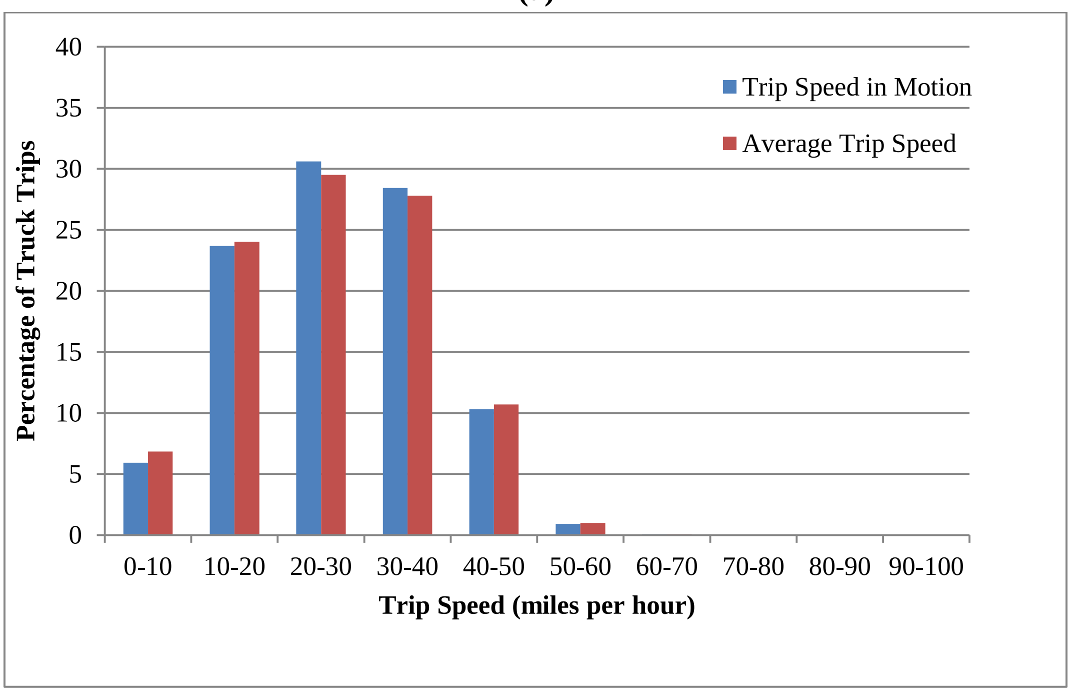


(b)

Figure 3.7 Trip Time Distribution of Trips Starting and Ending in Tampa FAF-Zone during: (a) Weekdays (61,465 trips), and (b) Weekend (7,780 trips)

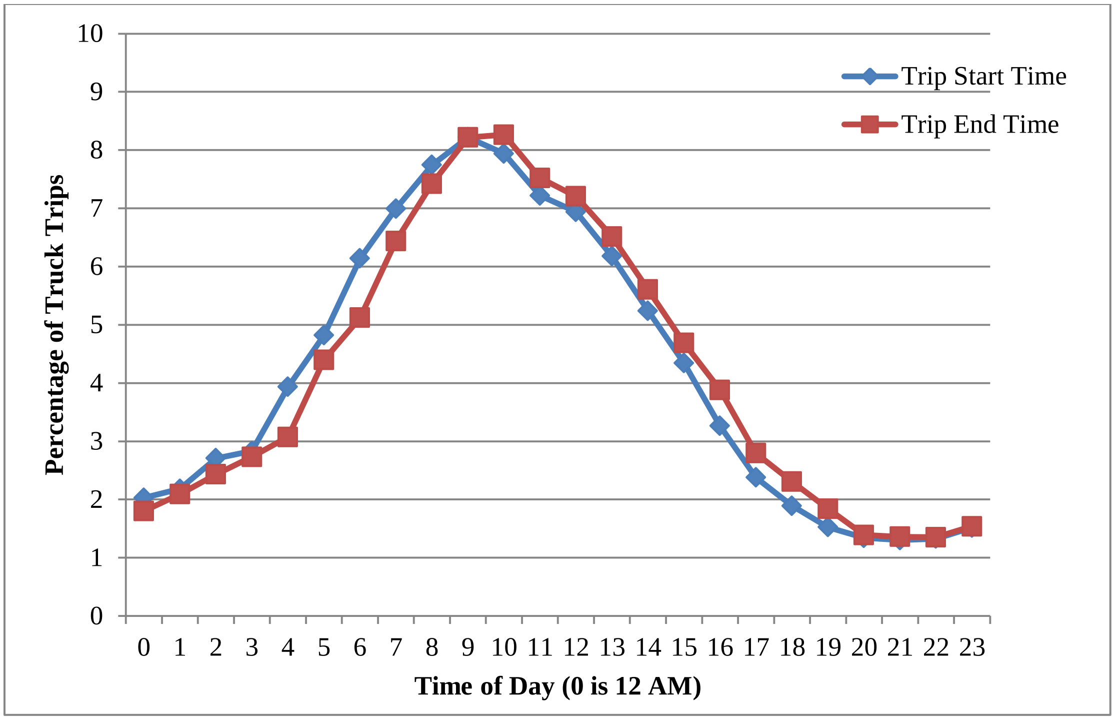


(a)

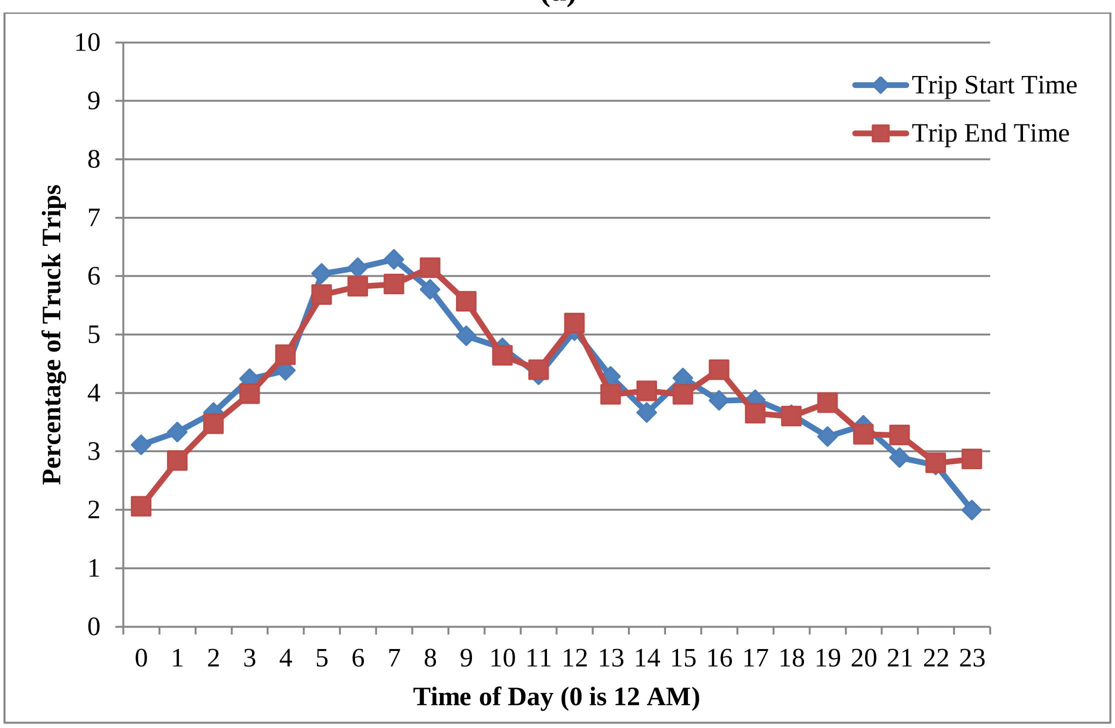


(b)

Figure 3.8 Trip Speed Distribution of Trips Starting and Ending in Tampa FAF-Zone during: (a) Weekdays (61,465 trips), and (b) Weekend (7,780 trips)



(a)



(b)

Figure 3.9 Time-of-Day Profile of Trips Starting and Ending in Tampa FAF-Zone during: (a) Weekdays (61,465 trips), and (b) Weekend (7,780 trips)

CHAPTER 4: AN ASSESSMENT OF THE COVERAGE OF TRUCK FLOWS IN FLORIDA BY THE ATRI DATA

4.1 Introduction

ATRI's truck-GPS dataset served as the source data for this analysis. While not a census of all truck movement within Florida, the substantial dataset proved valuable to understanding freight movement within the state. ATRI's truck-GPS dataset however, is not necessarily a random population of the trucks in the state. Therefore, this chapter provides an assessment of the type of trucks included in and the geographic coverage of the data.

4.2 Truck Types in ATRI Data

Based on the discussions with ATRI and anecdotal evidence, it is known that the major sources of ATRI data are large trucking fleets, which are typically comprised of tractor-trailer combinations. According to the FHWA vehicle type classification, these are tractor-trailer trucks of class 8 to 13 category.[7] However, a close observation of the data, through following the trucks on Google Earth and examining some travel characteristics of individual trucks, suggests that the data has a small proportion of trucks that do not necessarily haul freight over long distances. Since the data does not provide information on the vehicle classification of each individual truck, some heuristics were developed in the research to classify the trucks into heavy trucks and medium trucks. The heuristics, as explained below, are based on the travel characteristics of individual trucks over extended time periods (i.e., at least two weeks).

[7] Most freight in the US is carried by tractor trailer trucks of 5 axles or more (i.e., class 9 or above) and some on tractor trailer units of less than 5 axles (i.e., class 8 trucks). See page number ES-7 of http://www.fhwa.dot.gov/policy/vius97.pdf

As discussed in the earlier chapter, over 2.7 million truck trips were derived from about 4 months of ATRI's raw GPS data for the year 2010. This database had 169,714 unique truck IDs. Since the raw GPS data for each truck was available for at least two weeks (up to a month in most cases), trucks that did not make at least one trip of 100 miles in a two week period were removed from the data. In this step, a total of 88,869 trips made by 7,018 unique truck IDs were removed. The median length of such removed trips was 20 miles, suggesting the short-haul nature of these trucks. Subsequently, trucks that made more than 5 trips per day were removed assuming that these trucks are not freight carrying, tractor-trailer combination trucks. In this step, a total of 275,224 trips made by 918 unique truck IDs were removed. The median length of these trips was 16 miles. For the reader's information, a histogram of the distribution of the trucks in the database by their daily trip rates is provided in Figure 4.1 below.

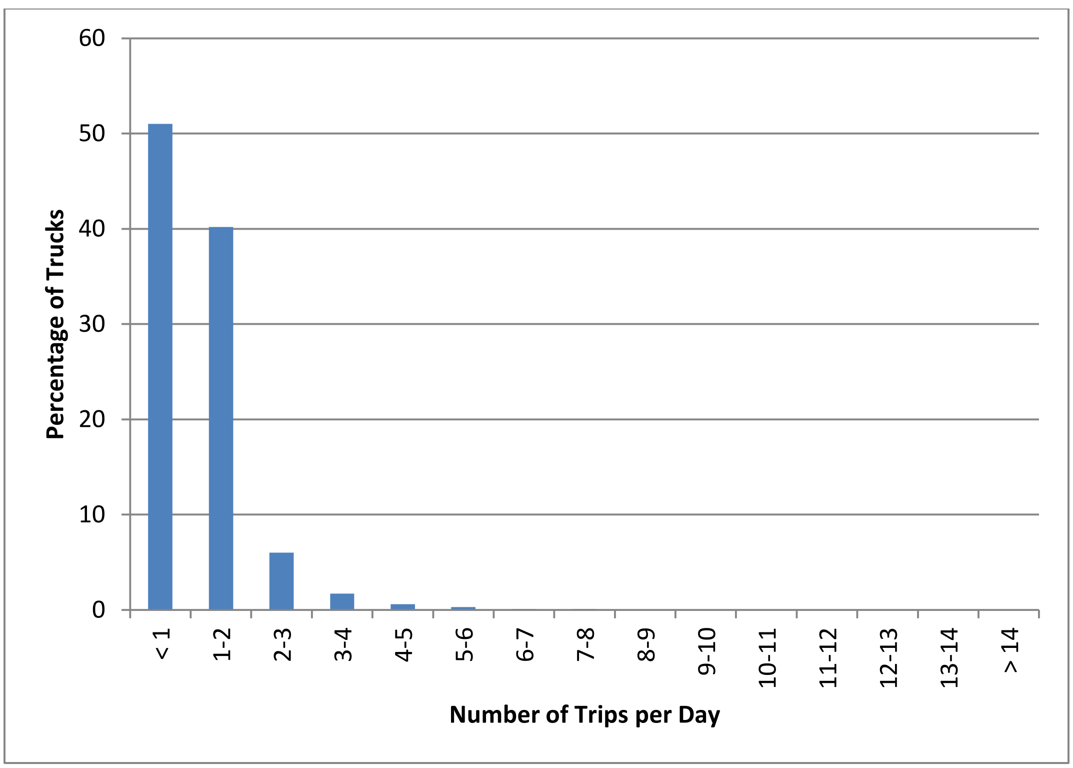


Figure 4.1 Distribution of Trucks by Daily Trip Rate

After the above discussed procedures, over 2.34 Million trips extracted from GPS data of over 161,776 unique truck IDs were considered as trips made by heavy trucks that carry freight. These trips were further used for OD matrix estimation later in the project. And the trucks making these trips are considered to be long-haul trucks or heavy trucks (assumed to be FHWA vehicle classification 8 to 13). The remaining trucks (i.e., 7936 truck IDs) whose trips were removed from further consideration in OD matrix estimation were assumed to be short-haul trucks or medium trucks that serve the purpose of local delivery and distribution.

It is worth noting here that the procedures used in this thesis to separate heavy trucks from other trucks are simplistic. Besides, it is not necessary that only heavy trucks carry freight over long distances while only medium trucks serve the purpose of short distance delivery and distribution services. Further research is needed to identify the composition of trucking fleet in the ATRI data and the purposes served by those trucks in the data.

4.3 What Proportion of Heavy Truck Traffic Flows in Florida is Captured in ATRI's Truck GPS Data?

ATRI's truck GPS data represents a large sample of truck flows within, coming into, and going out of the State. However, the sample does not necessarily comprise the entire population of truck flows. And it is unknown what proportion of truck flows in the state is represented by this data. To address this question, truck traffic flows in one-week of ATRI's truck GPS data was compared with truck counts data from Florida TTMS truck traffic counts for that same week. This section describes the procedure and results from this analysis.

One week of ATRI's truck GPS data – from May 9 to 15, 2010 – was used to derive weekly ATRI truck traffic volumes through the FDOT TTMS traffic locations. Specifically, all truck GPS records available with ATRI for that week within Florida as well as 60 miles beyond the Florida border into Alabama and Georgia were used. Including GPS data points 60 miles beyond Florida border helps account for truck trips coming into and going out of the state.

Generating data on weekly ATRI truck traffic volumes at each TTMS location requires counting the number of times the trucks in the ATRI data crossed the location in the week. To do so, we first attempted to run the raw GPS data through map-matching algorithms embedded in the network analyst tool of ArcGIS. However, given the sheer size of the raw GPS data, it turned out practically infeasible to run all the raw GPS data through the ArcGIS map-matching tool. To address the above issue, the raw GPS data records were reduced into a database of truck trip ends and intermediate GPS data points at a 5-minute interval (using the earlier discussed algorithms for converting the data into truck trips). Since the purpose was only to reduce the data to a manageable size and not to derive true pickup/delivery trip origins and destinations, any truck stop of more than 5 minutes dwell time was considered to be a trip end. A truck was considered to be "stopped" if either the spot speed was zero for at least 5 minutes or the average travel speed between consecutive GPS data points was less than 5mph for at least 5 minutes. For each truck trip derived in this fashion, given its origin and destination location coordinates, intermediate GPS data points in "motion" were sampled from the raw data at a time interval of 5 minutes. An intermediate GPS data point was considered to be in motion if the truck was moving at a speed of greater than 5mph. In all, the raw GPS data was reduced into a manageable GPS dataset comprising truck trip ends and intermediate GPS data points in motion at 5-minute time interval.

For each truck trip derived in the above-described fashion, the trip ends and intermediate GPS points were map-matched to the FLSWM highway network using the network analyst tool in ArcGIS. The map-matching algorithm snaps the GPS points to the nearest roadway links and also determines the shortest path between consecutive GPS data points. Since intermediate GPS data points between the trip ends were sampled at only a 5-minute interval, this procedure results in a sufficiently accurate route for the trip. The output from this process was an ArcGIS layer containing the travel routes for all trips generated from ATRI’s one-week truck GPS data. This ArcGIS layer was intersected with another layer of FLSWM network containing the TTMS traffic counting stations (specified in the form of network links on which the TTMS stations were located). This helped estimate the number of truck trips (or individual trip routes) crossing each TTMS count station, which is nothing but the volume of ATRI’ trucks crossing the count stations.

Data on weekly heavy truck traffic volumes (for May 9 to 15, 2010) was extracted from FDOT’s TTMS traffic counting data. Specifically, data on the total weekly volume of heavy trucks (i.e., Class 8 to 13 trucks) was extracted. Figure 4.2 shows the TTMS locations in Florida and the range of heavy truck traffic volumes (average daily traffic volumes) at those locations. Highest truck traffic volumes (i.e., around 5000 heavy trucks per day) can be observed in the northern part of I-75 near and above Ocala and on I-95 near Jacksonville. The section of I-75 between Ocala and Tampa, I-4 between Tampa and Orlando, and the section of I-95 in the southeast Florida has heavy truck volumes in the range of 3000 to 4000 truck per day.

Out of over 250 TTMS traffic counting locations in Florida, only 160 locations had traffic count data for all 7 days in the week. Therefore only these locations were selected for further analysis to compare the ATRI truck traffic volumes with observed TTMS truck traffic volumes at these locations. Figure 4.3 shows the results for each individual TTMS station. Specifically, the blue bars in the figure represent the observed heavy truck traffic volumes at those locations while the red bars represent the ATRI truck traffic volumes through those locations. Clearly, at no single location does the ATRI data provide 100% coverage of the observed truck traffic flows. However, the good news is that the data provides some coverage of the heavy truck traffic flows at all locations.

Table 4.1 shows these results aggregated by the facility type. The second column in the table shows the number of TTMS traffic counting locations on different types of highway facilities (along with the corresponding percentages). The third column shows the observed TTMS volumes counted at these sites (along with the corresponding percentages), again separately for each highway facility type. Notice that a bulk of heavy truck traffic (65.6%) is through locations on freeways and expressways that represent only 18.1% of the 160 TTMS sites considered in this analysis. The fourth column shows the truck traffic volumes counted in ATRI data using the afore-discussed map-matching procedure. As can be observed from the last row in this column, a total of 163,467 ATRI truck crossings were counted at the160 TTMS locations. It is worth noting that the distribution of these ATRI truck traffic counts across different facility types is very similar to the distribution of TTMS truck counts across facility types. This can be observed by comparing the percentage numbers in the third and fourth columns. This result suggests that the ATRI data provides a representative coverage of truck flows through different facility types in the state. The last column expresses the ATRI truck traffic counts as a percentage of observed heavy truck traffic counts at the TTMS locations. For example, a total of 111,608 ATRI truck crossings were counted at TTMS locations on freeways and expressways. These constitute 10.5% of over a Million observed heavy truck traffic counts at these locations. These percentages provide an aggregate picture of the coverage provided by ATRI data of the heavy truck traffic flows in Florida. Overall, as can be observed from the last row in the last column of the table, it can be concluded that the ATRI data provides 10.1% coverage of heavy truck flows observed in Florida. This result is useful in many ways. First, this provides an idea of

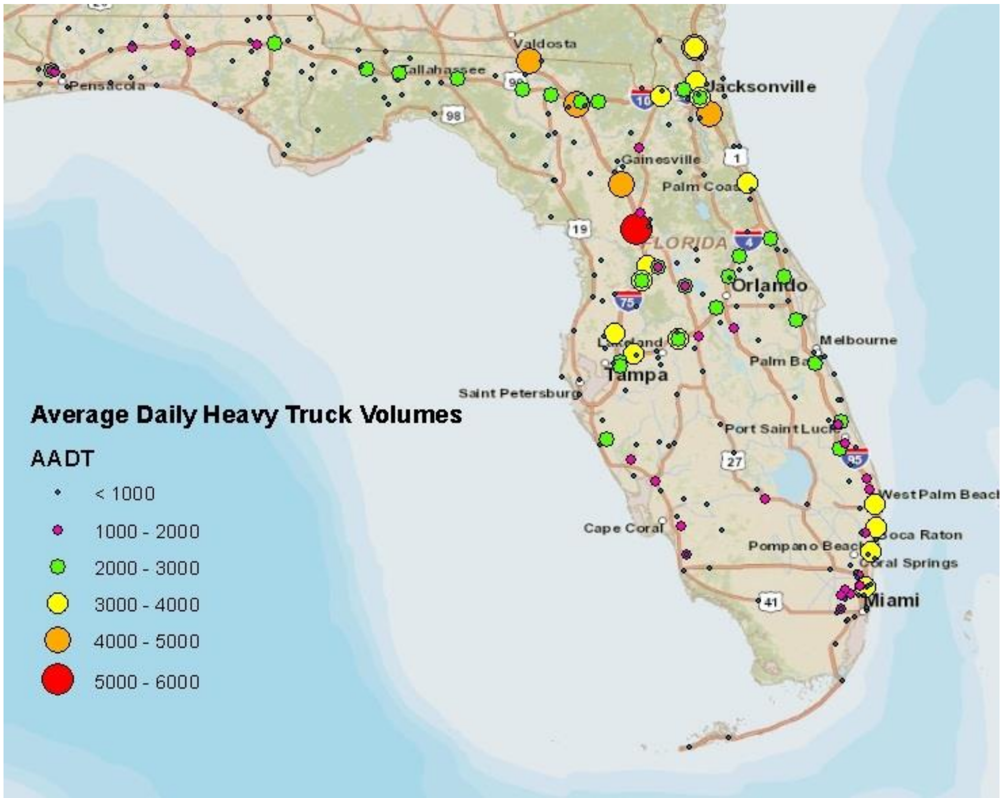


Figure 4.2 Observed Heavy Truck Traffic Flows at Different Telemetric Traffic Counting Sites (TTMS) in Florida

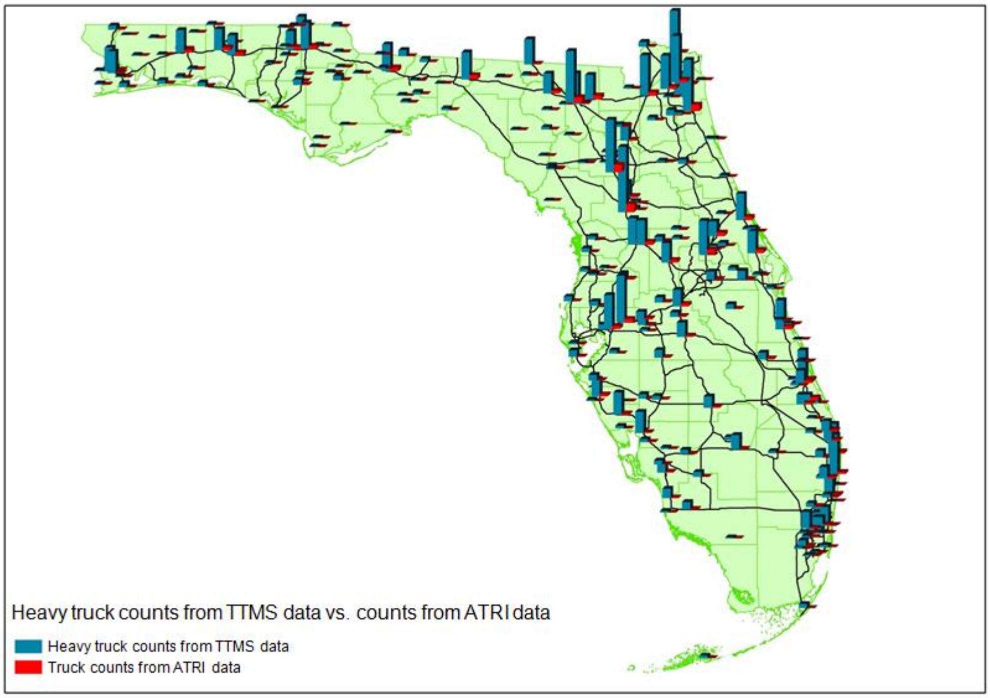


Figure 4.3 Heavy Truck (Class 8-13) Counts from TTMS Data vs Truck Counts from ATRI Data during May 9-15, 2010

the extent of coverage of Florida's heavy truck traffic flows in the ATRI data. Second, the result can be used to weigh the seed matrix of ATRI truck trip flows (by 10.1 times) to create a weighted seed matrix that can be used as an input for the ODME process.

Table 4.1 Aggregate Coverage of Heavy Truck Traffic Volumes in Florida by ATRI Data (for a week from May 9 to 15, 2010)

Facility Type	No. of TTMS Traffic Counting stations	Observed TTMS Truck Traffic Volumes (Class 8-13) during May 9-15, 2010	Truck Traffic Volumes in ATRI data during May 9-15, 2010	% Coverage assuming ATRI data comprises trucks of class 8-13
Freeways & Expressways	29 (18.1%)	1,063,765 (65.6%)	111,608 (68.3%)	10.5%
Divided Arterials	64 (40.0%)	333,791 (20.6%)	30,472 (18.6%)	9.1%
Undivided Arterials	52 (32.5%)	101,066 (6.2%)	6,969 (4.3%)	6.9%
Collectors	8 (5.0%)	42,164 (2.6%)	5,127 (3.1%)	12.2%
Toll Facilities	7 (4.4%)	80,493 (5.0%)	9,291 (5.7%)	11.5%
Total	**160**	**1,621,279**	**163,467**	**10.1%**

4.4 Geographical Coverage of ATRI's Data in Florida

To understand the geographical coverage of ATRI's data in Florida, we plotted the number of trips originating from (i.e., trip productions) and the number of trips destining to (i.e., trip attractions) each traffic analysis zone (TAZ) of the Florida Statewide Model (FLSWM). Figure 4.4 shows the TAZ-level trip productions and attractions while Figure 4.5 shows the County-level trip productions and attractions. Note that these trip productions and attractions were derived using the truck trips derived from four months of ATRI's truck-GPS data. Since the trips were derived from four months of data (i.e., 122 days) the total trip productions and attractions derived from the data were first divided by 122 to get average daily trip productions and attractions. However, since the data is found to represent 10% of observed heavy truck traffic volumes in the state, the average daily trip productions and attractions were weighted by 10. Such weighted average daily trip productions and attractions are shown in Figures 4.4 and 4.5.

In Figure 4.4, the TAZs shaded in yellow color have zero trip productions (in the left side figure) or zero trip attractions (in the right side figure) in ATRI's four-month GPS data. It can be observed that the Everglades region in the south and some TAZs in the northwest part of Florida have such TAZs with no trips extracted from the data. It is reasonable that none to limited heavy truck trips are produced from or attracted into the Everglades region. However, it is not clear if zero trip generation in some TAZs of north western Florida is a result of low penetration of data in those regions or if those TAZs indeed have no freight truck trip flows. To investigate this further, one can examine the observed heavy truck traffic flows in the TTMS data (Figure 4.2). Except on interstate 10, the northwestern region of the state does not have high truck traffic volumes. This suggests that zero trip generations in the ATRI data for several TAZs in the northwestern region is a reasonable representation of the truck flows in that region. Some TAZs in Duval County (Jacksonville area), Putnam, Polk, and Desoto counties have higher trip generation according to the ATRI data. When examined closely, all these TAZs have major freight activity centers such as distribution centers. However, major urban areas such as Miami, Tampa, and Orlando do not show TAZs with high trip generation. This is likely because the TAZs in these regions are smaller in size. Since the trip generations are not normalized by the area of the TAZ, it is difficult to make further inferences on the reasonableness of the TAZ-level trip generations.

Figure 4.5 shows the trip generations aggregated to a county-level. According to the ATRI data, Duval, Polk, Orange, Miami-Dade, and Hillsborough Counties, in that order, have the highest truck trip generation. It is expected that Counties with major metropolitan areas have the highest heavy truck trip generation. Further, Polk County is expected to have a high truck trip generation due to the presence of several freight distribution centers in the County. However, it is interesting that the truck trip generation in Polk County is higher than that in Hillsborough (Tampa), Orange (Orlando) and Dade (Miami) counties. Further, the truck trip generation in Southeast Florida (Miami, Broward and Palm Beach Counties) appears to be smaller than that in Polk County. These trends are not expected and are likely to be a manifestation of spatial biases in the data. To address such spatial biases, Chapter 6 and 7 combine the truck trip flows derived from the ATRI data with observed heavy truck traffic volumes at different locations in the state.

Figure 4.6 presents the percentage of heavy truck traffic covered by ATRI data at different locations. This information is similar to what was presented in Figure 4.3. However, for clarity and ease in interpretation the percentage covered is presented only for those locations with annual average daily heavy truck traffic greater than 1000 trucks per day. It can be observed that at most locations, at least 5 percent of the heavy truck traffic is captured in the ATRI data. Also, one can observe that the coverage in the southern part of Florida (within Miami) and the southern stretch of I-75 is slightly lower compared to the coverage in the northern and central Florida regions.

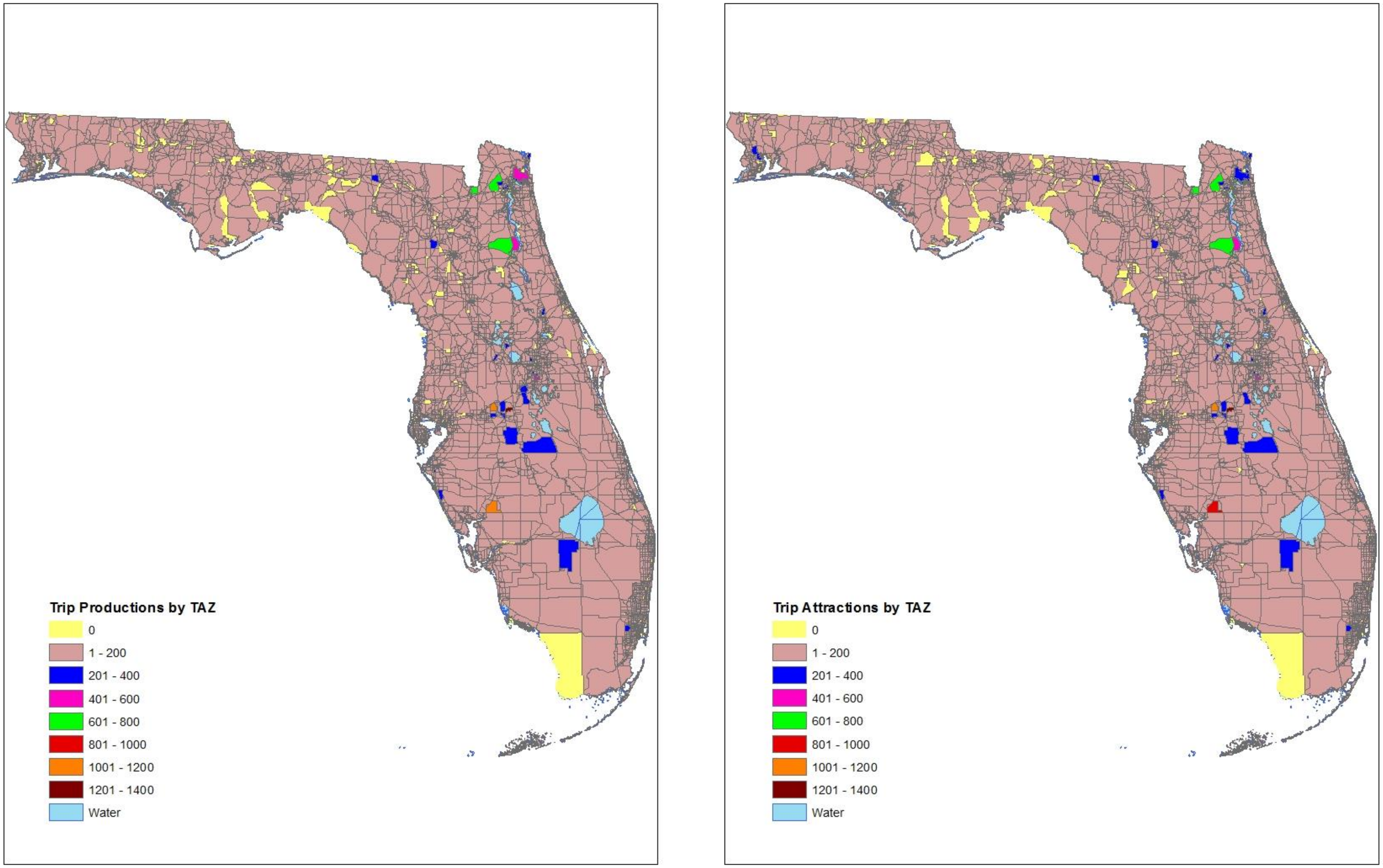


Figure 4.4 FLSWM TAZ-Level Trip Productions and Attractions in the ATRI Data (4 Months of Data Factored to One-day)

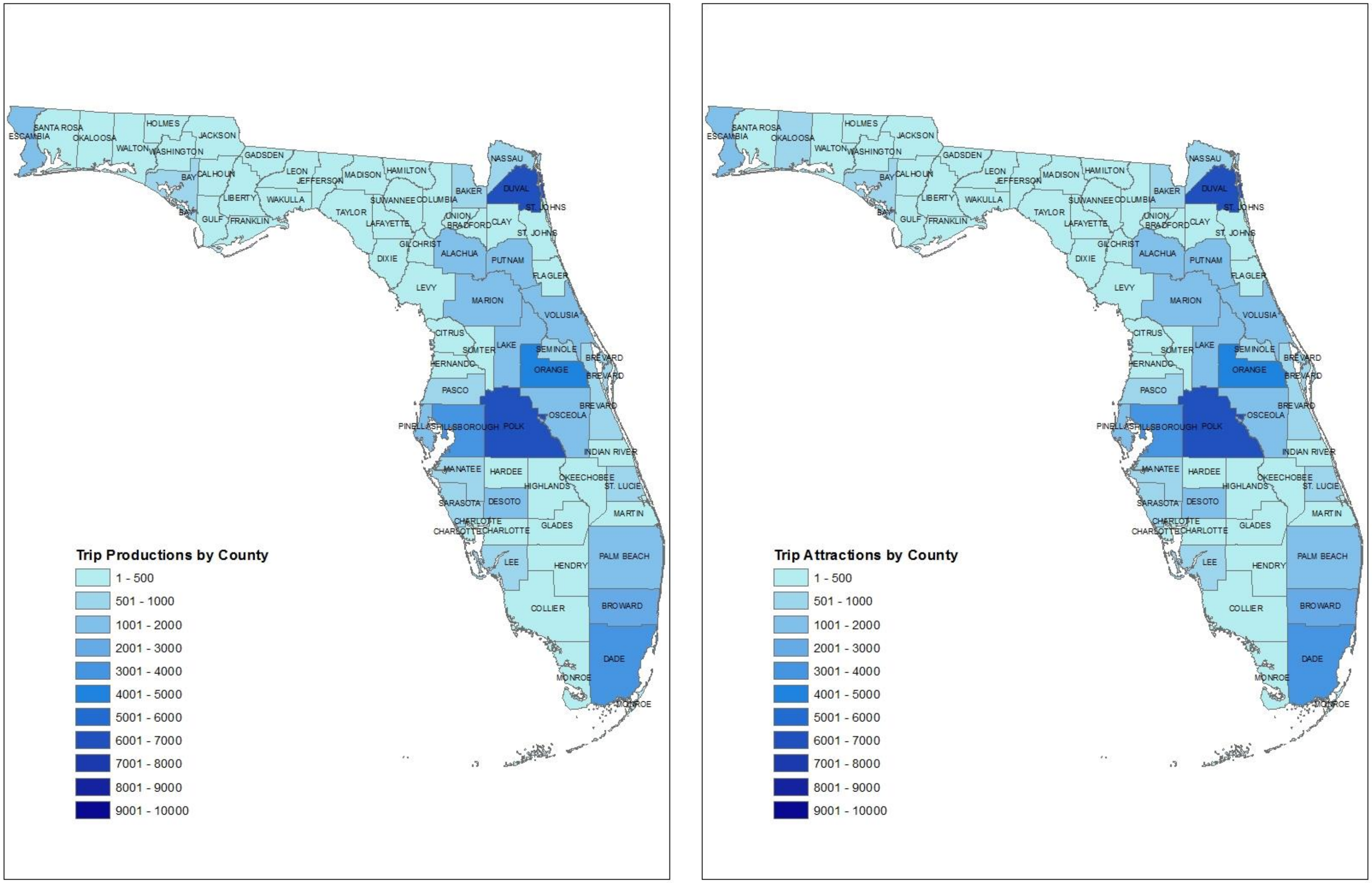


Figure 4.5 County-Level Trip Productions and Attractions in the ATRI Data (4 Months of Data Factored to One-day)

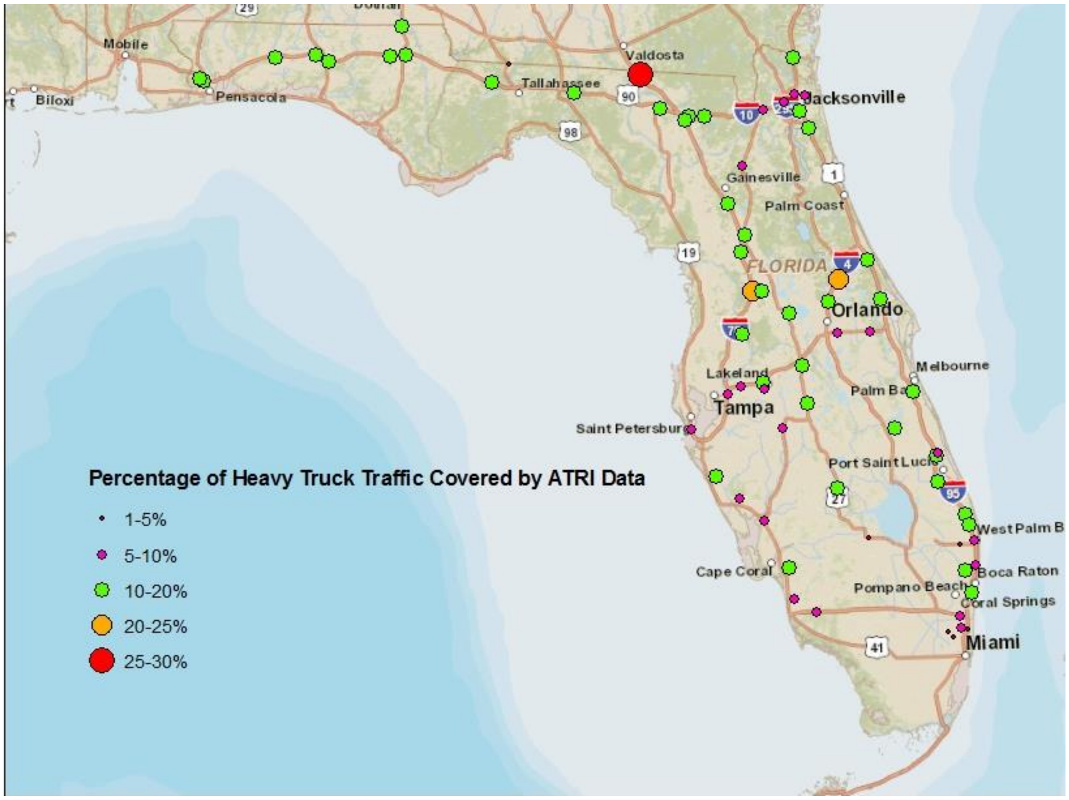


Figure 4.6 Percentage of Observed Heavy Truck (Class 8-13) Volumes Represented by ATRI Data at Telemetric Traffic Monitoring Sites in Florida, During May 9-15, 2010

CHAPTER 5: METHODOLOGY FOR HIGHWAY ASSIGNMENT AND ORIGIN-DESTINATION MATRIX ESTIMATION (ODME)

5.1 Introduction

ODME is a class of mathematical procedures used to update an existing matrix of origin-destination trip flows (i.e., number of trips between each origin-destination pair in a study area) using information on traffic flows at various locations in the transportation network. In the current project, the sample OD truck flows (also called the sample OD matrix or the seed matrix) extracted from ATRI's truck-GPS data can be updated using external information on truck traffic flows (or traffic counts or traffic volumes) observed on various links in the highway network within and outside Florida. Very broadly, the ODME procedures factor the ATRI data-derived truck trip flows in such a way that the trips in the resulting estimated OD flow matrix, when assigned to the highway network, closely match the observed heavy truck counts at various locations on the network. This chapter describes the methodologies used for highway assignment and ODME process.

5.2 Highway Assignment

ATRI freight trips, as the seed matrix, in addition to passenger cars flows as well as non-freight Quick Response Freight Manual (QRFM) trucks are used to reflect the overall traffic on the network. The QRFM truck and passenger car flows were obtained from Florida Statewide Legacy Model provided by Florida Department of Transportation. It is also notable that QRFM truck matrix includes trip rates by vehicle classifications for non-freight truck movements. Another input to this step is the network file with all the information such as speed, capacity, time and observed ground truck counts and so on. Traffic assignment is done based on the User Equilibrium Method using Bureau of Public Roads (BPR) function. John Glen Wardrop's first principle of equilibrium reads: "The journey times in all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route." [8] Basically the logic behind this principle is that people select routes that minimize their travel cost. In essence, travel time is the best representation of travel cost and therefore people seek to minimize their travel times. Wardrop's second principle of equilibrium also reads as: "At equilibrium the average journey time is minimum."[8] The second principle can also be justified based on first principle. When all people choose to minimize their travel times, equilibrium occurs in the whole traffic system and no user can decrease their costs by shifting to another route. The BPR function used in the equilibrium assignment is:

$$S_a\left(v_a\right)=t_a\left(1+\alpha\left(\frac{v_a}{c_a}\right)^{\beta}\right) \tag{1}$$

where, t_a is free flow travel time on link "a" per unit of time, v_a is the volume of traffic on link "a" per unit of time (more accurately: flow attempting to use link "a"), c_a is the capacity of link "a" per unit of time, $S_a\left(v_a\right)$ is the average travel time for a vehicle on link "a", α is the BPR coefficient and β is the BPR exponent.

5.3 Origin-Destination Matrix Estimation (ODME)

The mathematical procedure used for ODME in this thesis is based on the ODME procedure embedded in Cube Analyst Drive software from Citilabs. The procedure is essentially an optimization problem that tries to minimize a function of the difference between observed

[8] See pages 325–378 of "Some theoretical aspects of road traffic research (Wardrop, 1952)".

traffic counts and estimated traffic counts (from the estimated OD matrix) and the difference between the seed matrix and the estimated OD matrix, as below:

$$\begin{aligned} \arg\min_{X} \quad & J(X) = F(AX - b) + G(X - X_0) \\ \text{subject to} \quad & X \geq 0 \ \text{ and } \ \mathrm{X}_{lower} \leq X \leq X_{upper} \end{aligned} \tag{2}$$

where, X is the OD matrix to be estimated, X_0 is the initial (seed) OD matrix, G is a function measuring the distance between the estimated OD matrix and the initial matrix, b is a vector of observed counts at different locations in the study area, A is the route choice probability matrix obtained from the assignment of the OD flows in X on to the network (AX represents the estimated traffic counts at the same locations with available observed counts) and F is a function measuring the difference between estimated and observed traffic counts at different locations in the study area. As can be observed, the ODME procedure attempts to arrive at an OD flow matrix X in such a way that the resulting traffic volumes at differnet lcoations (AX) match closely with the observed traffic flows (b). At the sametime, the procedure avoids overfitting to the observed traffic flows by including the term $G(X - X_0)$ so that the estimated matrix is not too far from the seed matrix. X_{lower} and X_{upper} are boundaries (lower and upper bounds) within which the estimated matrix should fall. The Cube Analyst Drive has an option to use these boundary constratins to set lower and upper bounds on the estimated matrix, relative to the seed matrix.

Figure 5.1 shows a schematic of the ODME procedure used in the thesis. The primary inputs to the procedure are the seed matrix for freight truck trips (derived from the ATRI data); a highway transportation network for the study area, along with information on the travel times and capacity of each link in the network (extracted from the FLWSM); and observed heavy truck traffic volumes (or counts) at different locations added to corresponding links in the network. In addition to these OD flow matrices corresponding to travel other than freight truck flows – OD matrix for non-freight truck trips and OD matrix for passenger travel (both extracted from the FLSWM) – are required as inputs to generate realistic travel conditions in the network.

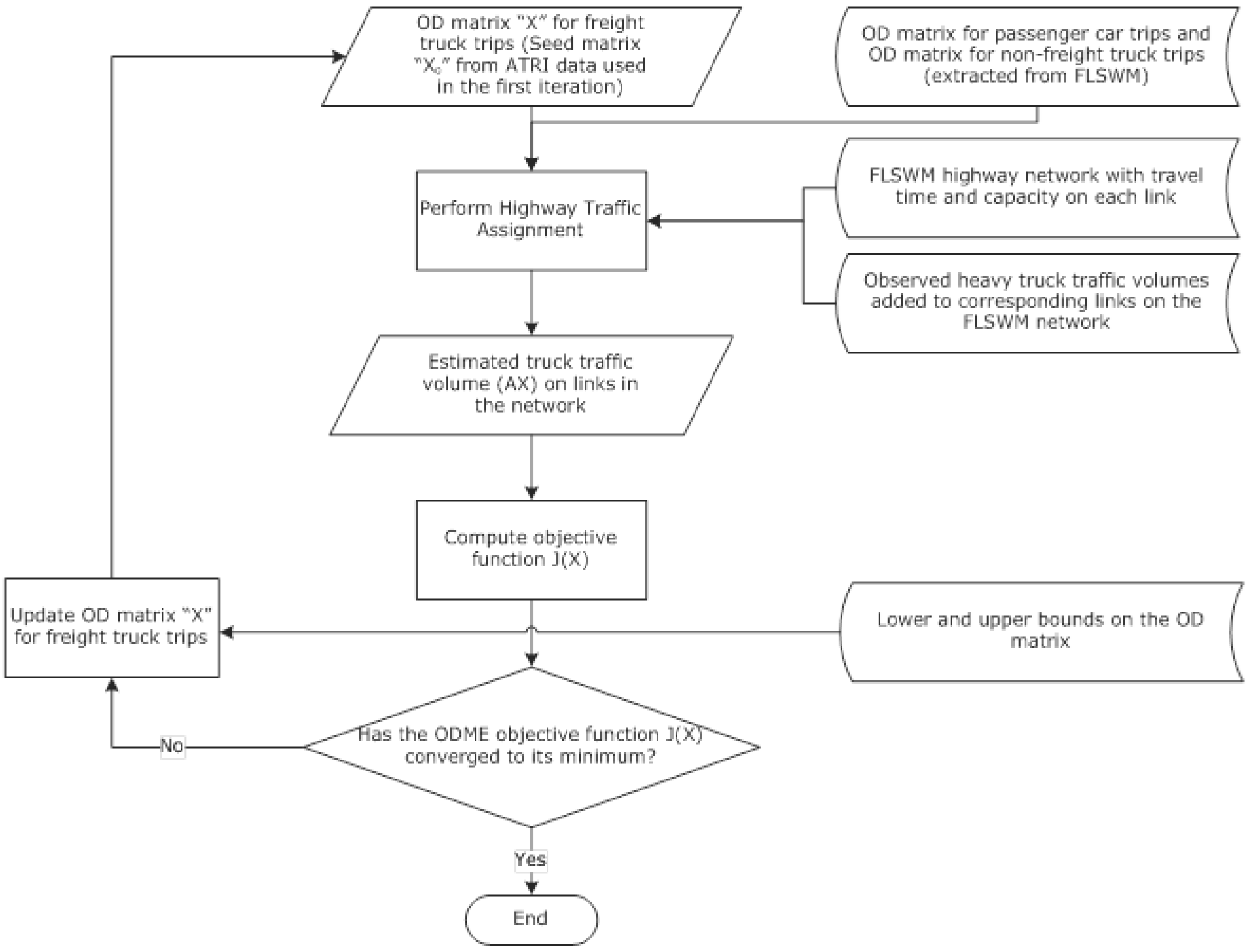


Figure 5.1 Schematic of the ODME Procedure Used in this Project

In the first step of the ODME procedure, the seed matrix of truck trips derived from the ATRI data (assumed to represent a sample of freight truck trips) and other OD matrices representing passenger travel and non-freight truck travel are loaded on to the highway network using user-equilibrium based traffic assignment procedures. The freight truck traffic volumes estimated from the traffic assignment procedure are then used in conjunction with the heavy truck traffic volumes observed at different locations in the network (along with the seed and estimated OD matrices for freight trucks, which are same in the first iteration) to compute the ODME objective function to be minimized. The seed matrix for freight truck trips is then updated toward minimizing the objective function while considering the lower and upper bounds on the matrix. This updated matrix is then used in conjunction with other OD matrices for passenger and non-freight travel (which are not updated in the procedure) as the seed matrix for the next iteration of the ODME procedure which begins with highway traffic assignment and follows with the computation of the ODME objective function. This process is repeated until the ODME objective function reaches its minimum, when the estimated freight truck traffic volumes are close enough to observed volumes and the estimated OD matrix is not too far from the initial seed matrix derived from the ATRI data.

The estimated OD matrix of freight truck trips can be evaluated using different evaluation metrics and procedures. One approach to evaluate the estimated OD matrix is based on comparison of estimated heavy truck traffic volumes and observed heavy truck traffic volumes at different locations within and outside Florida. Specifically, one can evaluate a root mean square error (RMSE) measure as below:

$$RMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(V_i - C_i\right)^2}{N}}}{C_{avg}} \tag{3}$$

where, V_i is the estimated heavy truck traffic volume corresponding to location *i*, C_i is the observed heavy truck traffic volume corresponding to location *i*, C_{avg} is the average heavy truck traffic count value of the entire set of observations, and N is the total number of truck counting locations in the set. One can compute a single RMSE value for the entire set of traffic counting locations and also separately for locations in Florida and for locations elsewhere. Similarly, one can compute the RMSE values separately for different ranges of observed heavy truck counts.

In addition to comparing the observed and estimated traffic volumes, it is important to assess the reasonableness of the estimated OD matrix in different ways. Aggregating the OD matrix to a coarser spatial resolution and examining the spatial distribution of flows, examining the total trips originating from (or trip productions) and total trips destined to (or trip attractions) each aggregate spatial zone, and examining the trip length distribution of the estimation OD matrix in comparison to the seed OD matrix are different ways of assessing the reasonableness of the estimated OD matrix.

CHAPTER 6: INPUTS AND ASSUMPTIONS FOR THE ODME PROCEDURE

6.1 The Seed Matrix

The seed matrix is essentially the matrix of OD truck trip flows derived from ATRI's truck-GPS data. Specifically, the truck trips derived from the GPS data were assigned to the traffic analysis zone (TAZ) system used in the Florida Statewide Model (FLSWM) to form the seed matrix. In FLSWM, Florida and the rest of the United States (and Canada) are divided into 6242 TAZs – 5403 of these zones are in Florida and the remaining zones outside Florida. Therefore, the seed matrix is a matrix of size 6242 x 6242, with each cell in it representing the number of trips extracted between the corresponding origin-destination (OD) pair.

In this thesis, the seed matrix was derived from 4 months of truck-GPS data – March, April, May, and June 2010. As described in an earlier chapter, while the total number of trips derived from 4 months of data was over 2.7 Million, for the purpose of OD matrix estimation only those trips deemed to be made by heavy trucks that haul freight (i.e., FHWA class 8 to 13 trucks, which are tractor trailers) were considered here. This is because most freight in the US is carried by tractor trailer trucks of 5 axles or more[9] (i.e., class 9 or above) and some on tractor trailer units of less than 5 axles (i.e., class 8 trucks). From discussions with ATRI, while most of the ATRI data comprises tractor trailer trucks, it is known that a small but non-negligible share of trucks in the data belong to FHWA classification 7, 6, or 5 (i.e., single unit trucks of 4 axles, 3 axles, or 2 axles). ATRI estimates that about 11% of the trucks in their data are trucks of class 7 or below that do not (for the most part) carry freight across regions. However, the raw data does

[9] http://www.fhwa.dot.gov/policy/vius97.pdf

not provide information on the classification of each truck. Therefore, some heuristics were developed to filter out trucks of class 7 or below. The heuristics used are discussed next.

Since the raw GPS data for each truck was available for at least two weeks (up to a month in most cases), trucks that did not make at least one trip of 100 miles in a two week period were removed from the data. In this step, a total of 88,869 trips made by 7,018 unique truck IDs were removed. The median length of such removed trips was 20 miles, suggesting the short-haul nature of these trucks. Subsequently, trucks that made more than 5 trips per day were removed assuming that these trucks are not freight carrying, tractor-trailer combination trucks. In this step, a total of 275,224 trips made by 918 unique truck IDs were removed. The median length of these trips was 16 miles. The remaining trucks in the data were considered to be tractor-trailer combination trucks that tend to make long-haul, freight carrying trips of interest to the Florida Statewide Model. Among the trips made by these tractor-trailer combination trucks, the following three scenarios were considered:

1) Only trips of greater than 10 miles made by tractor-trailer combination trucks.

2) Only trips of greater than 5 miles made by tractor-trailer combination trucks.

3) All trips made by tractor-trailer combination trucks.

After the above discussed procedures, the OD matrix of truck trips derived from 4 months of data comprised 2.07 Million trips. Since this was derived from 4 months of data (122 days), the OD matrix was scaled down to one day by dividing all the cells in the OD matrix by 122. The resulting OD matrix then represents one day of trips extracted from the ATRI data. We call this the one-day seed matrix. However, as discussed in Chapter 4, the research team found that at an aggregate level, the trucks flows in the ATRI data represent 10% of the heavy truck flows in Florida. Therefore, the one-day seed matrix was inflated by multiplying all the cells in the matrix by 10. We call this the weighted one-day seed matrix (or seed matrix in short). This weighted one-day seed matrix is the input as a seed matrix for the ODME process.

6.1.1 Geographical Coverage of the Seed Matrix

We will now examine the spatial structure of the seed matrix, focusing on its geographical coverage. To do so, the seed matrix was aggregated to county-level within Florida and state level outside Florida. The aggregation enables meaningful analysis of the spatial coverage of the seed matrix.

Within the state of Florida, there are a total of 67x67 = 4489 county-to-county OD pairs. Out of these, the seed matrix derived from the ATRI data contains trips for 3564 OD pairs (i.e., 79.4% coverage). The remaining 925 (20.6%) of the county-to-county OD pairs in Florida do not have trips in the seed matrix. From this, one can conclude that the heavy truck trips derived from the ATRI data cover close to 80% of the OD pairs in Florida. To examine the remaining 20% OD pairs for which the seed matrix does not contain any trips, Table 6.1 separates those OD pairs by county of origin (in the second column) and county of destination (in the third column). For example, it can be observed from the row for the Baker County that 9 counties in Florida did not have trips coming from the county while 7 counties did not have trips going into the county. That is, the seed matrix contains trips coming from all other 58 (= 67-9) counties to Baker and trips going from Baker to 60 (= 67-7) Counties in Florida. A close examination of this table suggests that counties associated with major urban regions (Miami Dade, Hillsborough, Orange, and Dual) and other counties with large freight activity (e.g., Polk) have a small number of counties to or from which there are no trips in the seed matrix. counties in the northwest Florida such as Franklin, Gulf, Calhoun, Holmes, Lafayette, Jefferson, and Hamilton; some rural counties in the south such as Glades, Hardee, and Monroe have higher number of zero trip flows coming into and going out of other counties in Florida.

Table 6.1 County-to-County OD Pairs in Florida with No Trips in the Seed OD Flow Matrix Derived from ATRI's Truck GPS Data

County	**No. of Counties to which there are no trips in the seed matrix**	**No. of Counties from which there are no trips in the seed matrix**
Alachua	0	1
Baker	9	7
Bay	10	4
Bradford	11	10
Brevard	8	9
Broward	5	4
Calhoun	33	37
Charlotte	19	17
Citrus	16	15
Clay	9	8
Collier	14	18
Columbia	5	5
De Soto	11	8
Dixie	18	14
Duval	0	2
Escambia	13	11
Flagler	18	20
Franklin	**45**	**43**
Gadsden	8	8
Gilchrist	19	20
Glades	**30**	**32**
Gulf	**36**	**39**
Hamilton	**24**	**22**
Hardee	**27**	**26**
Hendry	13	20
Hernando	9	8
Highlands	15	15
Hillsborough	1	2
Holmes	**30**	**33**
Indian River	19	24
Jackson	6	10
Jefferson	**33**	**35**

Table 6.1 (Continued)

Lafayette	**35**	**34**
Lake	3	3
Lee	7	8
Leon	4	7
Levy	19	13
Liberty	21	7
Madison	8	5
Manatee	5	7
Marion	4	3
Martin	16	16
Miami-Dade	3	3
Monroe	**31**	**38**
Nassau	6	2
Okaloosa	19	10
Okeechobee	20	24
Orange	2	1
Osceola	2	5
Palm Beach	6	7
Pasco	5	9
Pinellas	5	7
Polk	1	0
Putnam	4	4
Santa Rosa	23	20
Sarasota	9	14
Seminole	8	16
St. Johns	11	13
St. Lucie	7	8
Sumter	4	8
Suwannee	9	8
Taylor	12	12
Union	20	19
Volusia	6	4
Wakulla	22	23
Walton	24	25
Washington	**30**	**15**
Total	**925**	**925**

For OD pairs with at least one end in Florida, Table 6.2 presents the number of OD pairs with no trips for each origin and destination county in Florida. Specifically, the second column shows the number of states outside Florida to which no single trip was extracted while the third column shows the number of states from which no single trip was extracted. For example, it can be observed from the row for the Alachua County that 11 states outside Florida did not have trips coming from the county while 10 states did not have trips going into the county. Similar to the county-to-county Flows, counties in the northwest such as Franklin, Gulf, Calhoun, Hamilton, Walton, Holmes, Lafayette, Jefferson and rural counties in the south such as Glades and Monroe have no trips coming into or going out of other states.

Table 6.2 Florida County to Non-Florida State OD Pairs with No Trips in the Seed OD Flow Matrix Derived from ATRI's Truck GPS Data

County	No. of Counties to which there are no trips in the seed matrix	No. of Counties from which there are no trips in the seed matrix
Alachua	11	10
Baker	20	11
Bay	9	11
Bradford	22	23
Brevard	15	11
Broward	8	5
Calhoun	**41**	**37**
Charlotte	26	21
Citrus	26	24
Clay	20	13
Collier	21	20
Columbia	22	19
De Soto	19	10
Dixie	26	34
Duval	4	2
Escambia	11	12
Flagler	25	23
Franklin	**42**	**41**
Gadsden	16	20
Gilchrist	23	28
Glades	**33**	**33**
Gulf	**38**	**39**
Hamilton	**35**	**36**
Hardee	26	27

Table 6.2 (Continued)

Hendry	13	25
Hernando	17	18
Highlands	18	23
Hillsborough	4	4
Holmes	**36**	**32**
Indian River	18	24
Jackson	21	23
Jefferson	**34**	**34**
Lafayette	**39**	**38**
Lake	4	5
Lee	23	20
Leon	25	20
Levy	23	26
Liberty	25	32
Madison	23	19
Manatee	7	14
Marion	13	9
Martin	21	28
Miami Dade	5	2
Monroe	**36**	**37**
Nassau	16	18
Okaloosa	18	18
Okeechobee	24	26
Orange	4	3
Osceola	12	10
Palm Beach	10	9
Pasco	17	14
Pinellas	6	10
Polk	3	2
Putnam	8	12
Santa Rosa	22	20
Sarasota	22	17
Seminole	12	13
St. Johns	18	17
St. Lucie	18	15
Sumter	16	17
Suwannee	21	26
Taylor	24	29
Union	28	34
Volusia	8	12

Table 6.2 (Continued)

Wakulla	25	27
Walton	**32**	**32**
Washington	26	29
Total	**1334**	**1353**

Overall, it can be concluded that the ATRI data provides a sound geographic coverage of trip flows within, to, and from Florida. While several Counties in northwest Florida and a few rural counties in the South (e.g., Glades and Monroe) show no trips to and from several other counties and states, it is likely because these counties may not actually have truck flows to/from a large number of locations. Considering that the seed matrix was derived from 4 months of raw GPS data (which is a large amount of data), if some OD pairs at a county-level resolution do not have any trip exchanges, it is reasonable to expect that those OD pairs may not indeed have truck flows in reality. On the other hand, for OD pairs with both ends outside the state of Florida, 350 out of the 2500 (=50x50) state-to-state OD pairs did not have any trips in the seed matrix. Since the data is Florida centric, it is likely that the seed matrix is not necessarily a good representation of OD flows outside Florida.

6.1.2 Zero Cells in the Seed Matrix

When the seed OD matrix was examined at its actual spatial resolution (i.e., the FLSWM TAZ level), only 0.41 Million of the 39 Million TAZ-to-TAZ OD pairs had trips. That is, the 2 Million heavy truck trips extracted from ATRI's truck GPS data could fill only 0.41 Million OD pairs. The remaining 38.5 Million OD pairs had no trips. This is relevant here because most ODME methods used in practice operate only with OD pairs that have non-zero trips in the seed matrix. Consequently the final OD matrix output from ODME methods will have zero trips for OD pairs that began with zero in the seed matrix. To address this issue, a common practice is to introduce a small positive number (e.g., 0.01) for zero-cells (i.e., OD pairs with zero trips) in the seed matrix that the analyst believes should have trip flows. The question then becomes which OD pairs with zero trips can be expected to have trip flows in reality. The earlier discussion on the spatial coverage of the seed matrix, albeit at an aggregate spatial resolution of counties and states, sheds light on this issue. Recall from the earlier discussion that the OD pairs with at least one end in Florida have sufficient coverage at the county level in Florida and at the state level outside Florida. While there may be gaps at the disaggregate TAZ level, it was considered unnecessary to alter zero cells for such OD pairs. For OD pairs outside Florida, it may be reasonable to explore altering the zero-cells to include a small number (0.01) and examine if the ODME procedure provides better results.

The following scenarios were considered for altering the zero cells in the seed matrix:

1) None of the zero-cells were altered (assuming the zero cells in the seed matrix are truly representative of zero truck flows)
2) Only the zero-cells for OD pairs outside Florida were altered to 0.01, to allow for the possibility of truck flows between those OD pairs. This scenario assumes that zero cells for OD pairs within, to, and from Florida are truly representative of zero truck flows.
3) All zero-cells were altered to 0.01. This scenario, that all OD pairs will have truck flows, is very unlikely in reality. Nevertheless it was considered to be sure.

6.2 Truck Traffic Counts

Observed volumes of truck traffic at different locations on the network is an important input into the ODME process. For the current study, data on heavy truck traffic counts were gathered for several locations within Florida as well as outside Florida. Since the OD matrix to be estimated includes truck traffic flows going into (out of) Florida from (to) other states, it was considered important to include truck traffic counts outside Florida as well.

6.2.1 Truck Traffic Counts in Florida

Data on truck traffic counts in Florida was obtained from FDOT's Telemetered Traffic Monitoring Sites (TTMS) traffic counting program. FDOT collects daily data on traffic volumes (by direction), speed, vehicle type, and weight from over 250 TTMS locations on Florida's highway network. From such TTMS data for the year 2010, the daily traffic volume information for different vehicle classes was extracted for the months of March, April, May and June 2010 (the same months for which the seed matrix is available). The vehicle classifications range from 1 to 15, with class 8 through 13 representing heavy trucks (i.e., tractor-trailer combinations) and class 15 representing 'unknown' category. For each TTMS location, the average daily traffic (ADT) was computed for heavy trucks along with the number of days for which the traffic count data was available. Subsequently, the data was examined for any anomalies as discussed below.

First, 241 locations whose coordinates fell on the FLSWM highway network locations were selected. The other TTMS locations that were on highway links not in the FLSWM network were removed from consideration. For 237 of these locations, the TTMS traffic count data was available for both directions (141 sites with traffic counts in north-south directions and 96 sites in east/west directions). The remaining 4 sites had counts for one direction. This makes it a total of 478 TTMS traffic counts distinguished by location and direction. Out of these, we considered only 460 locations with TTMS data for more than 30 days of the 4 months. Subsequently, sites with the following types of anomalies were removed – those with abnormally high percentage of traffic counts, those with abnormally high difference in directional counts, and those with a high percentage of unclassified trucks (i.e., class 15). After all these screening procedures, TTMS heavy truck counts (i.e., ADT for heavy trucks) for 413 different locations were retained for use in the ODME process. Figure 6.1 shows the spatial distribution of those locations in Florida (Percentages in parentheses show the distribution of heavy truck ADT at these 413 locations). It is worth noting that 22.5% of these locations are on freeways and expressways, 36% are on divided arterials, 30% are on undivided arterials, 6.5% are on toll facilities, and the remaining are on collector roads, ramps, one-way facilities, and centroid connectors. Out of all the 413 different heavy truck counts at different locations in Florida, data from 365 locations were used in the ODME process while data from the remaining 48 locations were kept aside for validation purposes.

6.2.2 Truck Traffic Counts outside Florida

The FHWA's vehicle travel information system (VTRIS) database was utilized to obtain truck traffic counts on highway network locations in all states other than Florida and Georgia. For Georgia, truck traffic counts from Georgia Automated Traffic Recorder (ATR) locations were obtained from Georgia Department of Transportation (GDOT). While the VTRIS database provides traffic count data for a large number of locations outside Florida and the Georgia ATR data provides so similar data in Georgia, only 635 of these locations fell on the FLSWM highway network links outside Florida. This is because the FLSWM network outside Florida is not very detailed. Figure 6.3 shows the locations of all these 635 counting sites along with the 413 counting sites in Florida. As can be observed, while Florida and Georgia have good coverage of traffic counting stations, other states in the southeast such as Alabama, Mississippi, Louisiana, and South Carolina have very few traffic counting locations. Tennessee, Kentucky, and North Carolina do not have any traffic counting locations. This will likely have a bearing on ODME results. In future, the ODME results can potentially be improved by increasing the spatial coverage of the traffic counting stations in the southeastern states. Finally, Out of all the 635 different heavy truck counts at different locations outside Florida, data from 598 locations were used in the ODME process while data from the remaining 37 locations were kept aside for validation purposes.

6.3 Network

The highway network from the FLSWM was used as an input for traffic assignment purposes in ODME. The network is detailed within Florida and just outside the border of Florida and less detailed as it extends into other states farther from Florida. As can be observed from figure 6.2, the network extends in Canada and Mexico as well. The inputs associated with the network include the following:

- Free flow speed, travel time (including any delays due to toll plazas) and capacity of each link in the network, and
- Free flow speed, travel time (including any delays due to toll plazas) and capacity of each link in the network, and
- Observed average daily heavy truck counts by direction on over 200 links obtained from the TTMS data in Florida and the VTRIS data in other states (figures 6.2 and 6.3 show the locations of those traffic counting locations).

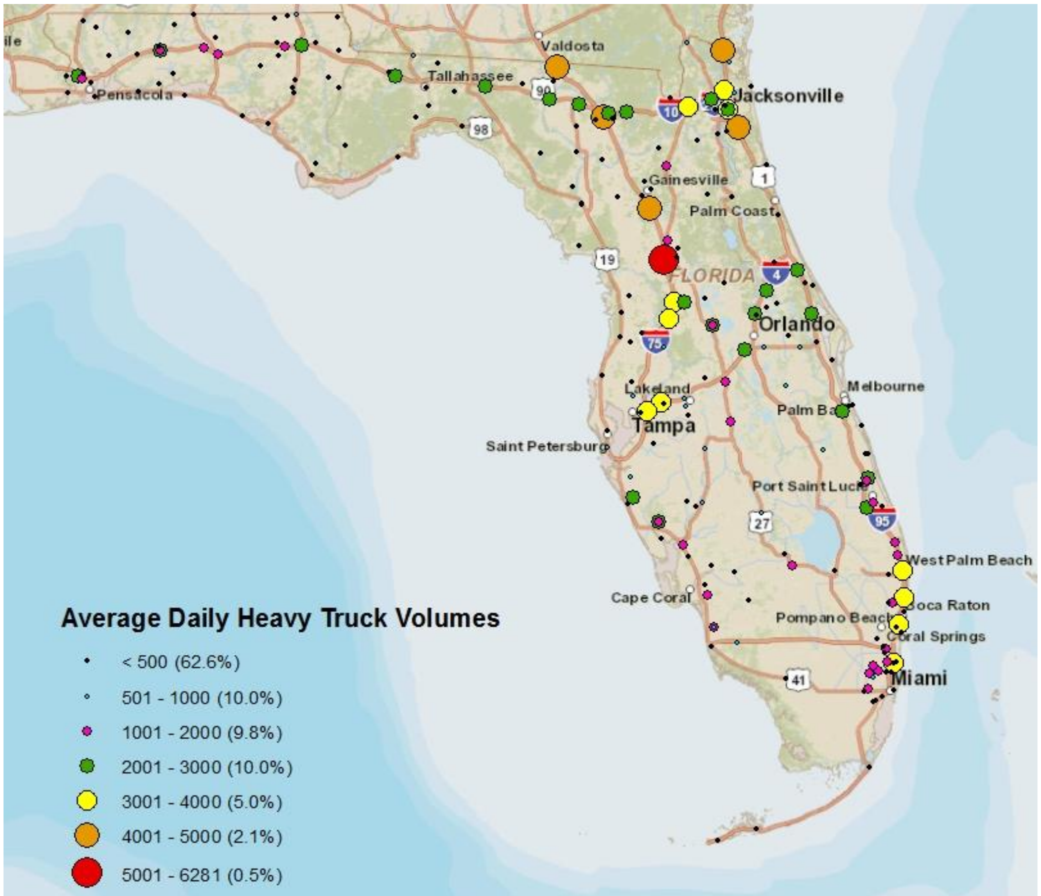


Figure 6.1 Spatial Distribution of Telemetered Traffic Monitoring Sites (TTMS) in Florida Used for ODME

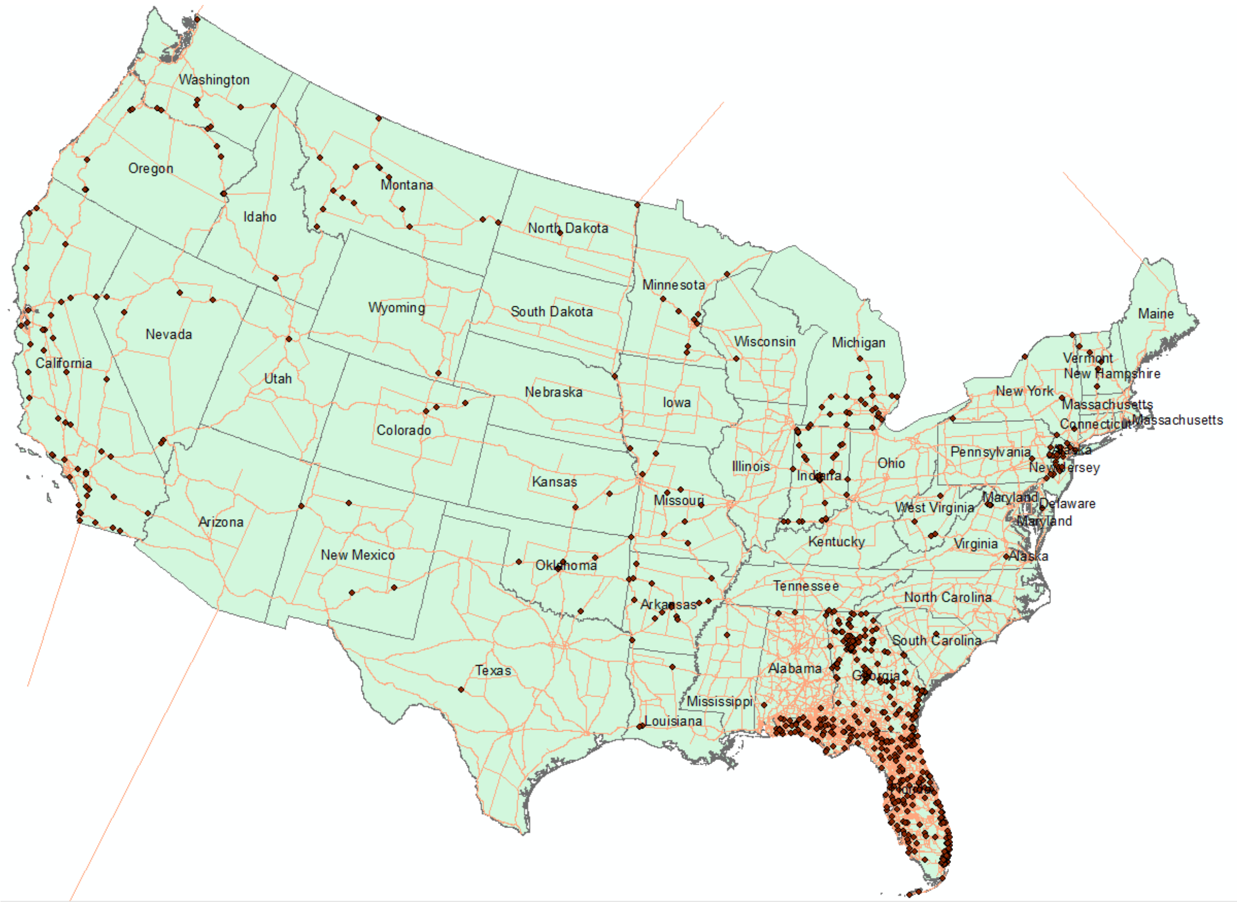


Figure 6.2 Spatial Distribution of Traffic Counting Stations in the Nation Used for ODME

CHAPTER 7: RESULTS FROM THE ODME PROCEDURE

7.1 Evaluation of Different Assumptions and Scenarios

The ODME procedure in CUBE's Analyst Drive was run several times to evaluate different assumptions on the OD matrix. These assumptions include assumptions (or constraints) of lower and upper bounds on the seed matrix, assumptions on zero cells in the seed matrix, and assumptions on the minimum trip length to be considered in the seed matrix.

Assumptions on the bounds for trips between different OD pairs in the seed matrix include:

1) Lower bound equal to the seed matrix; i.e., none of the estimated number of trips between any OD pairs should be less than those in the seed matrix,
2) No lower bound on the seed matrix; i.e., the estimated matrix can have zero trips between OD pairs even if there were trip flows observed in the seed matrix,
3) Lower bound equal to 0.7 of the seed matrix; i.e., none of the estimated trips between any OD pair should be less than 0.7 times those in the seed matrix,
4) An upper bound of 50 times the seed matrix; i.e., none of the estimated trips between any OD pair should be more than 50 times those in the seed matrix,
5) An upper bound of 100 times the seed matrix; i.e., none of the estimated trips between any OD pair should be more than 100 times those in the seed matrix, and
6) No upper bound on the seed matrix.

It is worth noting here that Cube's Analyst Drive software does not allow the bounds to be different across different cells in the matrix. The bounds have to be uniform across all cells.

Assumptions on zero-cells in the seed matrix include:

1) Assume OD pairs with zero-cells in the seed matrix do not have truck flows in reality (i.e., all zero-cells were retained as zeroes),
2) Alter only the zero-cells for OD pairs outside Florida to 0.01 to allow the possibility of truck flows between those OD pairs,
3) Alter all zero-cells to 0.01, assuming that each zero-cell in the seed matrix is likely to have truck trips in reality.

Assumptions on minimum trip length in the seed matrix include:

1) Minimum trip length of 10 miles
2) Minimum trip length of 5 miles
3) Minimum trip length of 1 mile

Among the assumptions on lower/upper bounds on the OD matrix the extent of upper bound did not influence the results (i.e., the estimated OD matrix) as long as the bound was large enough. Therefore, we removed the upper bound on the OD matrix. The extent of lower bounds had a significant influence on the estimated OD matrix. When lower bounds were removed on all cells in the OD matrix, the estimated truck traffic volumes matched better than the scenarios that imposed lower bounds on the OD matrix. This can be observed from Table 7.1. Specifically, the RMSE values between estimated and observed truck traffic volumes are smallest when no lower bounds are imposed on the OD matrix. However, in this scenario, the trip length distribution of the estimated OD matrix was changing considerably toward a greater share of shorter trips than those in the seed matrix. There are two possible reasons for such a change in the trip length distribution from the seed matrix to the estimated OD matrix. One possible reason is that the seed matrix is biased toward long-distance trips and that combining the seed matrix with the observed traffic counts helps reduce the bias by estimating more short-length trips. The other possible reason is that the estimated OD matrix from the ODME procedure is over fitting to the observed traffic counts without necessarily correcting for biases in the seed matrix. To investigate this further, we closely examined for any possible anomalies in the estimated OD matrix.

When closely examined, the estimated OD matrix (when the lower bounds were removed) had zero trips between many OD pairs that originally had some observed trips in the seed matrix from the ATRI data. While this is not necessarily a problem in itself, the estimated OD matrix had zero trips between Florida and some southeastern states that had no observed traffic counts from the VTRIS data (recall that we could not use any observed truck traffic counts from Tennessee and North Carolina). For example, no OD pair between North Carolina and Florida and between Tennessee and Florida had any trips in the estimated OD matrix, although at least 100 trips were observed between those states and Florida in the ATRI data. This suggests that the estimated OD matrix is an artifact of over-fitting to the observed traffic volumes rather than a realistic representation of OD flows within, to, and from Florida. A closer examination of the RMSE values for this scenario also suggests that the ODME procedure in this scenario is over-fitting to the observed traffic counts. Specifically, the RMSE value between the estimated and observed heavy truck volumes for input stations (i.e., the TTMS locations from which the data was used for ODME procedure) was only 4%. Such an excellent fit to the observed data did not translate to the validation data; i.e., the RMSE between estimated and observed heavy truck volumes is 37% for TTMS locations from which the truck traffic count data was kept aside for validation. Therefore, the research team believes that the estimated OD matrix in this scenario is an artifact of over fitting to the observed truck traffic volumes. In future work, this issue can be resolved by obtaining better observed truck traffic count information from all southeastern states, especially those that do not have any or very few traffic counts in the inputs used in the project.

When the lower bound was set to be equal to the seed matrix, the estimated OD matrix was very close in its trip length distribution to the seed matrix. However, the heavy truck traffic volumes implied by the estimated OD matrix (obtained from traffic assignment) were not close enough to the observed heavy truck traffic counts. The root mean squared value between the estimated traffic volumes and the observed traffic volumes was close to 60%.

Table 7.1 RMSE Values between Estimated and Observed Heavy Truck Traffic Volumes at Different Locations in Florida for Different Assumptions in the ODME Procedure

	No lower bounds assumed on the OD matrix		Lower bound equal to the no of trips in seed matrix		Lower bound equal to 0.7 times number of trips in seed matrix	
Observed daily heavy truck counts	RMSE for input stations	RMSE for validation stations	RMSE for input stations	RMSE for validation stations	RMSE for input stations	RMSE for validation stations
20-100	8%	92%	83%	104%	70%	105%
100-500	6%	91%	77%	93%	47%	93%
500-1000	2%	38%	46%	47%	33%	49%
1000-7000	2%	25%	37%	40%	11%	24%
All	**4%**	**37%**	**59%**	**60%**	**20%**	**38%**

As a middle ground between the above two scenarios, we explored a scenario where the lower bounds were set to be 0.7 times the seed matrix (i.e., none of the estimated trips between any OD pairs should be less than 0.7 times those in the seed matrix). Note that the seed matrix used as input for the ODME procedure is a 10-times inflated version of the one-day seed matrix extracted from the ATRI data. This was done to recognize that the ATRI data represents about 10% of the observed heavy truck flows in the state (at an aggregate level). However, it is not necessary that the data represents 10% of heavy truck flows at every location. In some locations, the data might represent more or less than 10% of the observed heavy truck flows. Therefore, setting a lower bound of 0.7 allows for the possibility that the actual heavy truck trip flows might be less than the 10-time inflated number of heavy truck trips in the ATRI data. This scenario provided reasonable results, with RMSE value of 20% for input stations and 38% validation stations while also allowing trips from (and to) all states to (and from) Florida.

Among the assumptions on zero-cells, keeping the zero-cells as is provided better results both in terms of validation measures against observed heavy truck counts as well as reasonableness of the spatial distribution of truck flows. For instance, altering all zero-cells to 0.01 provided high RMSE values results unless the lower bounds were removed on all cells. However, removing the lower bounds on all cells, as discussed earlier, was leading to over-fitting of the estimated heavy truck traffic volumes to observed truck traffic volumes. Since there was no easy mechanism in Cube's Analyst Drive software to incorporate different bounds for different OD pairs, we could not impose lower bounds on only non-zero cells in the OD matrix and allow the altered zero-cells to become zero. Besides, since the seed matrix was derived using a large database from 4 months of ATRI data, zero-cells can be reasonably assumed to represent no truck flows between the corresponding OD pairs. Recall from the discussion in Section 7.1.1 that, when the seed matrix was aggregated to the county-level, not too many OD pairs in the state had zero trips.

Assumptions on minimum trip length in the seed matrix did not significantly alter the estimated OD matrix except that assumptions with smaller trip length cutoffs led to a higher share of intra-county trips. Since the purpose of this effort is toward statewide freight truck flow modeling, we retained the assumption that valid pickup/delivery trips of heavy trucks should be of at least 10 mile length.

7.2 ODME Results for One Set of Assumptions

This section presents and discusses results from the following set of assumptions in the ODME procedure: (1) No upper bounds but a lower bound of 0.7 times the seed matrix on the estimated OD matrix, (2) Trips of at least 10 miles length, (3) Zero-cells in the seed matrix assumed to truly represent zero truck flows. The results based on these assumptions are considered to be the final results for ODME in this thesis. However, there is scope for improving the results, which will be discussed toward the end of this chapter.

Table 7.2 presents a summary of the truck trips in the seed matrix and those in the estimated OD matrix. As mentioned before, the seed matrix has trips between nearly 0.41 Million OD pairs. Of these, close to 0.3 Million OD pairs have at least one end in Florida, while 0.18 Million OD pairs have both ends in Florida. As can be observed from the table, the same OD pairs have trips in the estimated OD matrix. The seed matrix contained a total 69,025 trips that started and/or ended in Florida while the estimated OD matrix contains a total of 104,587 trips. Close to 70% (i.e., 73,202) of the estimated trips with at least one end in Florida were within Florida. The daily mileage of estimated trips with at least one end in Florida was over 27 Million miles. 26.6% of these miles (i.e., over 7 million miles) were due to trips within Florida.

Table 7.2 Summary of Truck Trips in the Seed and Estimated OD Matrices

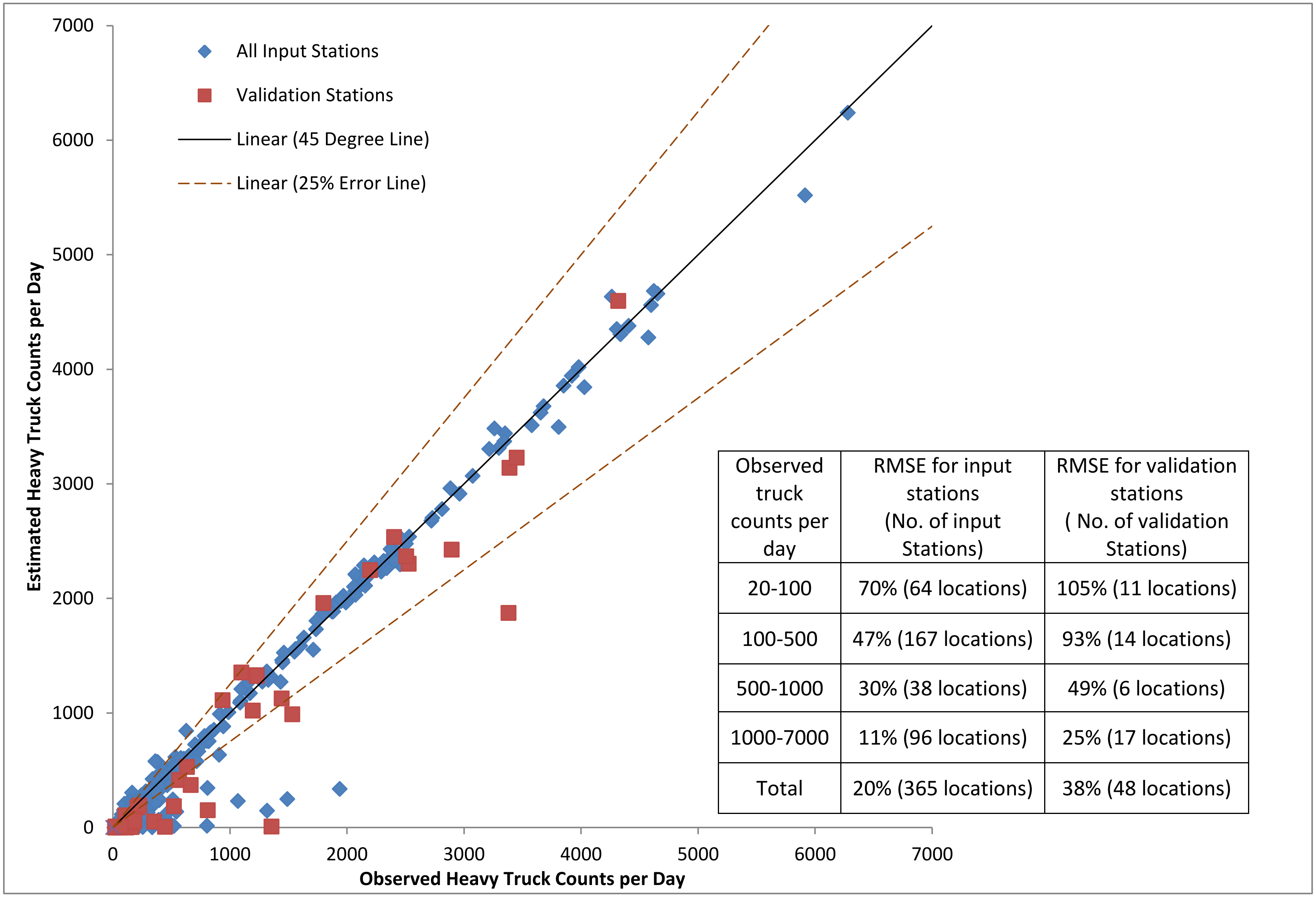
	Seed OD Matrix	**Estimated OD Matrix**
Trips between all OD pairs within and outside Florida		
No. of OD pairs with trips	410,559	410,559
No. of trips per day	169,859	343,071
Miles traveled per day	54,642,638	206,760,073
Trips with at least one end in Florida		
No. of OD pairs with trips	304,730	304,730
No. of trips per day	69,025	104,587
Miles traveled per day	18,877,950	27,207,442
Trips with both ends in Florida		
No. of OD pairs with trips	183,050	183,050
No. of trips per day	42,434	73,202
Miles traveled per day	4,662,039	7,246,461

Figure 7.1 shows a comparison of estimated truck traffic volumes (using the estimated OD matrix) and observed truck traffic volumes in the TTMS data. The blue dots in the figure are for TTMS stations from which the observed traffic volume data was used on the ODME process,

while the red dots are for TTMS stations from which the observed traffic volume data was kept aside for validation. The solid straight line is the 45 degree line. All dots that fall on this line indicate a perfect fit between the estimated truck volume and the observed truck volume. The dots that fall between the two dotted lines correspond to those locations where the estimated truck traffic volumes are within 25% deviation from the observed truck traffic volumes. A table embedded within the figure shows the aggregate RMSE values for different ranges of observed truck traffic volumes. It can be observed that the estimated truck traffic volumes are matching reasonably well with the observed volumes, especially at locations with truck volumes higher than 1000 trucks per day.

Figure 7.2 shows the trip length distributions of the trips in the seed and estimated OD matrices (the trip lengths are based on TAZ-to-TAZ distances in the FLSWM). The top graph in the figure shows the distribution for trips with at least one end in Florida (this include trips between other states and Florida) while the bottom graph shows the distribution for trips with both ends in Florida. It can be observed that the distribution of the trips in the estimated OD matrix is closely following those from the seed matrix derived from the ATRI data, albeit the estimated OD matrix has a slightly greater proportion of shorter length trips than the seed matrix. Notice from the top graph that the estimated trips show a spike in the trips of length greater than 2000 miles when compared to those in the seed matrix. These are likely trips between Florida and states from the northwestern including California.

Figures 7.3 and 7.4 show the county-level trip productions and attractions, respectively, for both the seed and estimated OD matrices. As discussed in Chapter 4, the seed matrix shows lower than expected trip generation in the south Florida region (especially in and around Miami) and the southern stretch of I-75 beginning from the Tampa region (when compared to those in



Observed truck counts per day	RMSE for input stations (No. of input Stations)	RMSE for validation stations (No. of validation Stations)
20-100	70% (64 locations)	105% (11 locations)
100-500	47% (167 locations)	93% (14 locations)
500-1000	30% (38 locations)	49% (6 locations)
1000-7000	11% (96 locations)	25% (17 locations)
Total	20% (365 locations)	38% (48 locations)

Figure 7.1 Observed Vs. Estimated Heavy Truck Counts per Day at Different Locations in Florida

the Polk County). These trends were observed in discussions related to the coverage of heavy truck flows in Florida by the ATRI data. The estimated OD matrix, due to its use of additional information on the observed heavy truck traffic flows, addresses this issue to a certain extent. This can be observed in Figures 7.3 and 7.4, where counties in the southeast Florida and Hillsborough County have higher trip generation in the estimated OD matrix than in the seed OD matrix. Also note that the trip generation in the Duval County has increased as well. This is perhaps due to a high volume of heavy truck traffic in the Jacksonville region.

Tables 7.3 and 7.4 show the state-to-state trip flows in the seed and estimated OD matrices for a selected set of states. Specifically, Table 7.3 shows the distribution of the trips starting in Florida while Table 7.4 shows the distribution of trips ending in Florida. The seed matrix shows that around 75% of the trips staring (ending) in Florida stay within (are from within) Florida, while the estimated matrix adjusts this distribution to contain about 82% of those trips within Florida. The next top destinations (origins) for trips starting (ending) in Florida include Georgia, Alabama, and California. It is interesting that California is one of the top destinations (origins) for trips starting in Florida. The reasons for this are not clear and need further investigation. One possibility is that such heightened flows between California and Florida may be artifacts of the ODME procedure given the observed heavy truck traffic volumes in the two states.

Tables 7.5 and 7.6 show the county-to-county trip flows in the seed and estimated OD matrices for counties with the highest truck trip productions and attractions in the data. In both the tables, cells with greater than 10% value are shaded in red while those with 5-10% value are shaded in brown. Table 7.5 shows the destinations of heavy truck flows from Counties with highest trip production in Florida, while Table 7.6 shows the origins of heavy truck flows to counties with highest trip attraction. Several observations can be made from these tables. First, as expected, a good proportion of trips to/from each county are from/to within the county. Second, the seed matrix shows Polk County as one of the major origins/destinations for trips from/to other counties. The estimated OD matrix makes adjustments to this trend for Miami-Dade, Palm Beach, and Broward Counties. Specifically, the estimated OD matrix shows greater flows between these three counties. Third, the estimated OD matrix shows smaller proportion of flows between Hillsborough and Miami Dade Counties than that in the seed OD matrix. While one would expect greater amount of flows between these two counties, the observed heavy truck traffic volumes on major highways between these two counties are not high enough to support this notion.

Finally, Figure 7.6 shows a comparison of the truck trip flows between seed and estimated OD matrices. It can be observed that only those OD pairs with less than 100 trips in the seed matrix have been modified in the estimated seed matrix.

7.3 Scope for Improvements to ODME Results

Even though different assumptions were made during ODME process, there are a few improvements that can be implemented in order to obtain more reliable estimation results. As mentioned earlier, heavy truck volume counts in different count stations play an important role as a major input in the ODME process. Even though the research team made several efforts to obtain heavy truck counts for more locations, especially within adjacent states to Florida, only the counts for more locations in Georgia were acquired successfully. Therefore, the ODME results can potentially be improved by increasing the spatial coverage of the traffic counting stations in the southeastern states (Alabama, Louisiana, South Carolina, North Carolina, Mississippi and Tennessee). This will also help us to capture the actual heavy truck flows more accurately, especially flows from other states to Florida and from Florida to other states. Furthermore, the issue of over fitting the estimated OD matrix to the observed truck traffic volumes can be resolved by obtaining better observed truck traffic count information from all southeastern states, especially those that do not have any or very few traffic counts in the inputs used in the project. Hence, the estimated OD matrix will be a more proper representation of heavy truck flows.

Also as mentioned earlier, Cube's Analyst Drive software does not allow the bounds to be different across different cells in the matrix and the bounds have to be uniform across all cells. In addition to the aforementioned modifications, one can try different means of OD estimation to incorporate different boundary conditions to different cells. This way different limits can be applied to different OD pairs based on the reliability of number of trips between each OD pair. For instance, since the truck trips extracted from ATRI data in the seed matrix are Florida centric, one can allow for more changes in the seed matrix for OD pairs outside Florida, while keeping some constraints for the OD pairs with at least one end in Florida.

Table 7.3 Heavy Truck Trip Flows from Florida to Other States (Greater than 0.5 % in Estimated Matrix)

Destination State	AL	CA	FL	GA	IL	MI	NJ	Other	Total
Seed OD matrix from ATRI data	4.6%	0.1%	75.8%	10.3%	0.3%	0.2%	0.3%	8.4%	100.0%
Estimated OD matrix	2.8%	1.2%	82.1%	9.3%	1.0%	0.9%	1.1%	1.6%	100.0%

Table 7.4 Heavy Truck Trip Flows from Other States to Florida (Greater than 0.5 % in Estimated Matrix)

Origin State	Seed OD matrix from ATRI data	Estimated OD matrix
AL	3.9%	2.8%
CA	0.1%	1.4%
FL	76.4%	82.6%
GA	9.4%	8.4%
IL	0.3%	1.1%
MI	0.1%	0.6%
NJ	0.3%	0.8%
Other	9.5%	2.3%
Total	100%	100%

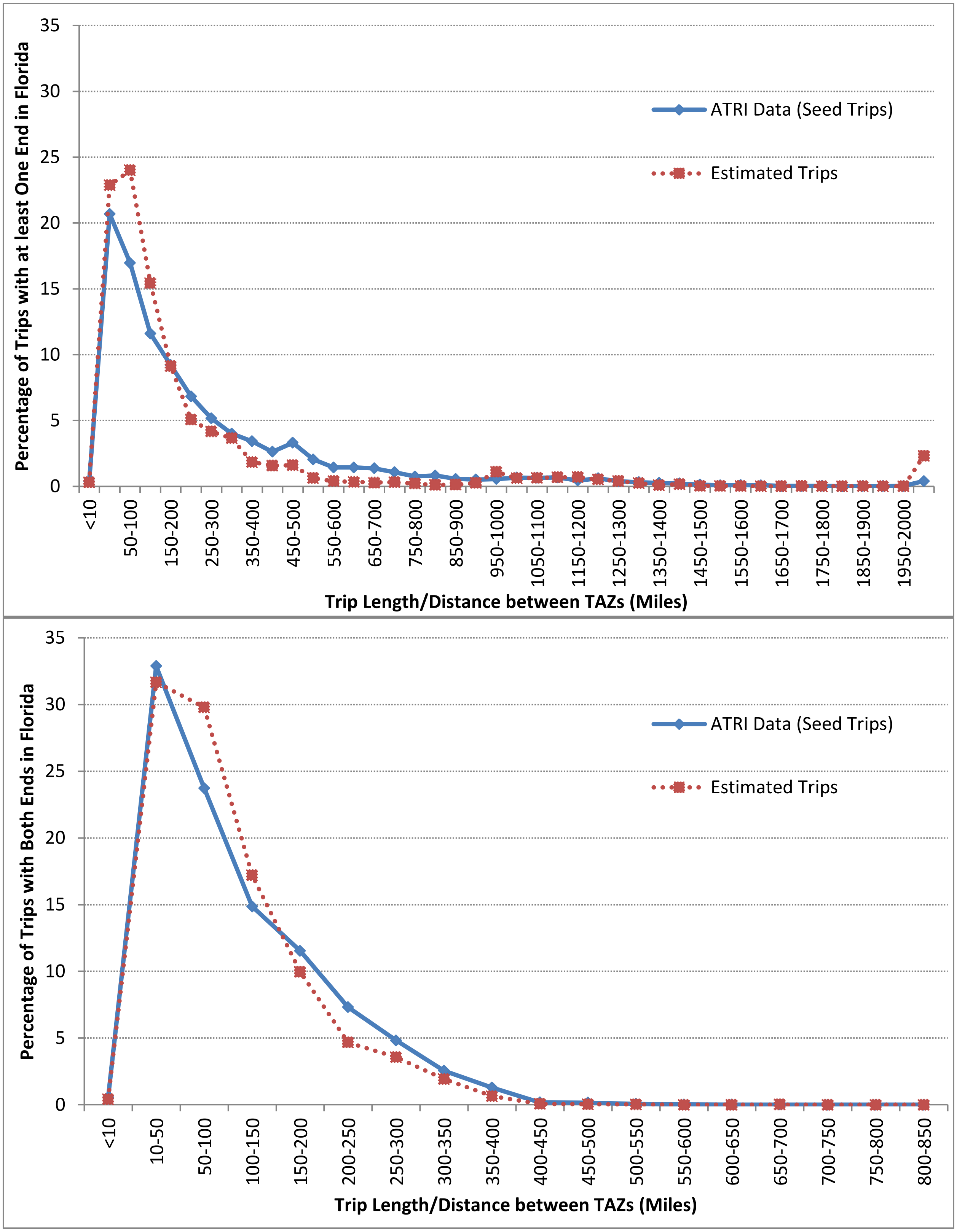


Figure 7.2 Trip Length Distributions of the Trips in the Estimated and Seed OD Matrices

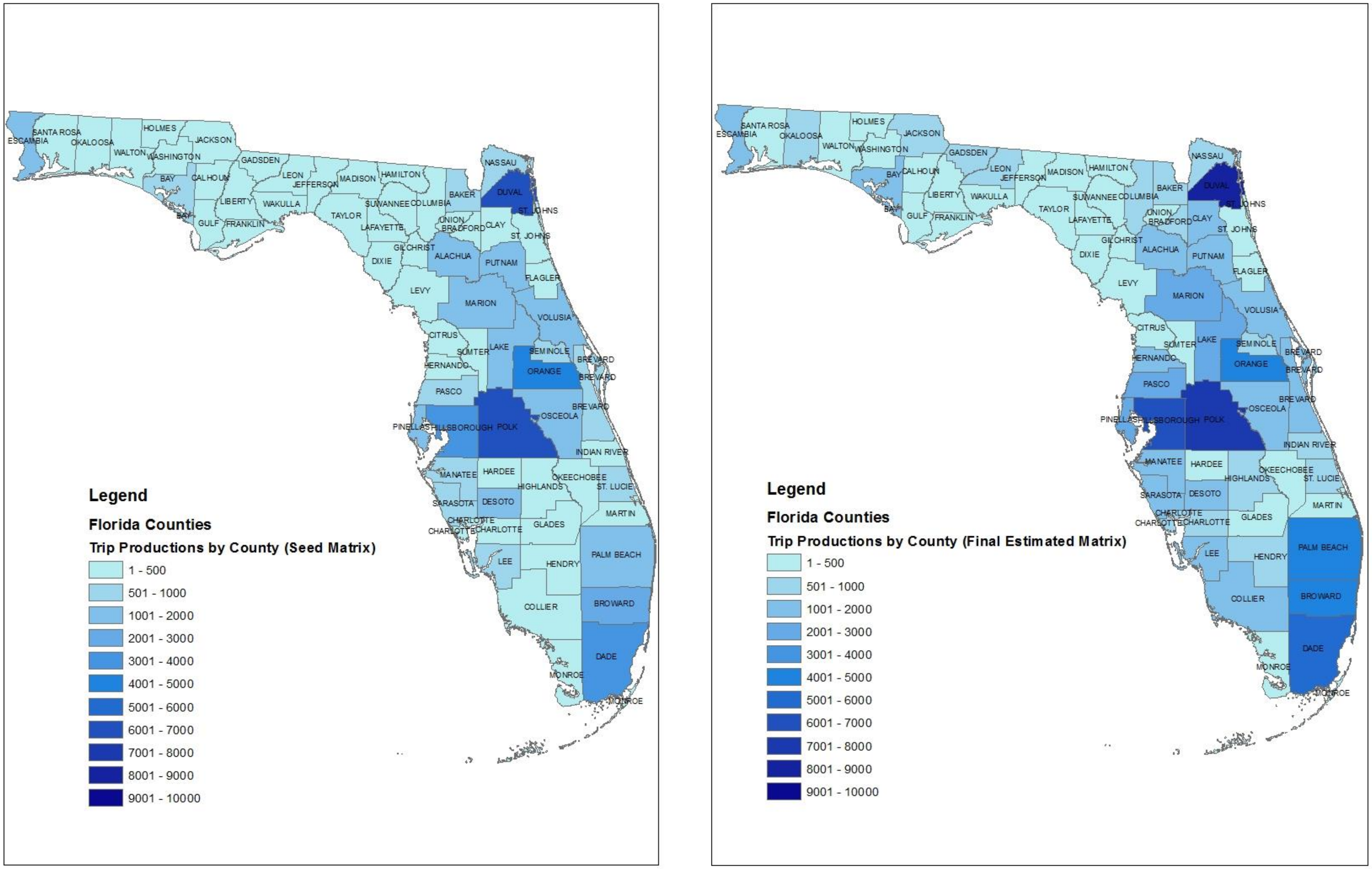


Figure 7.3 Comparison of Trip Productions by County between Seed OD Matrix and Estimated OD Matrix

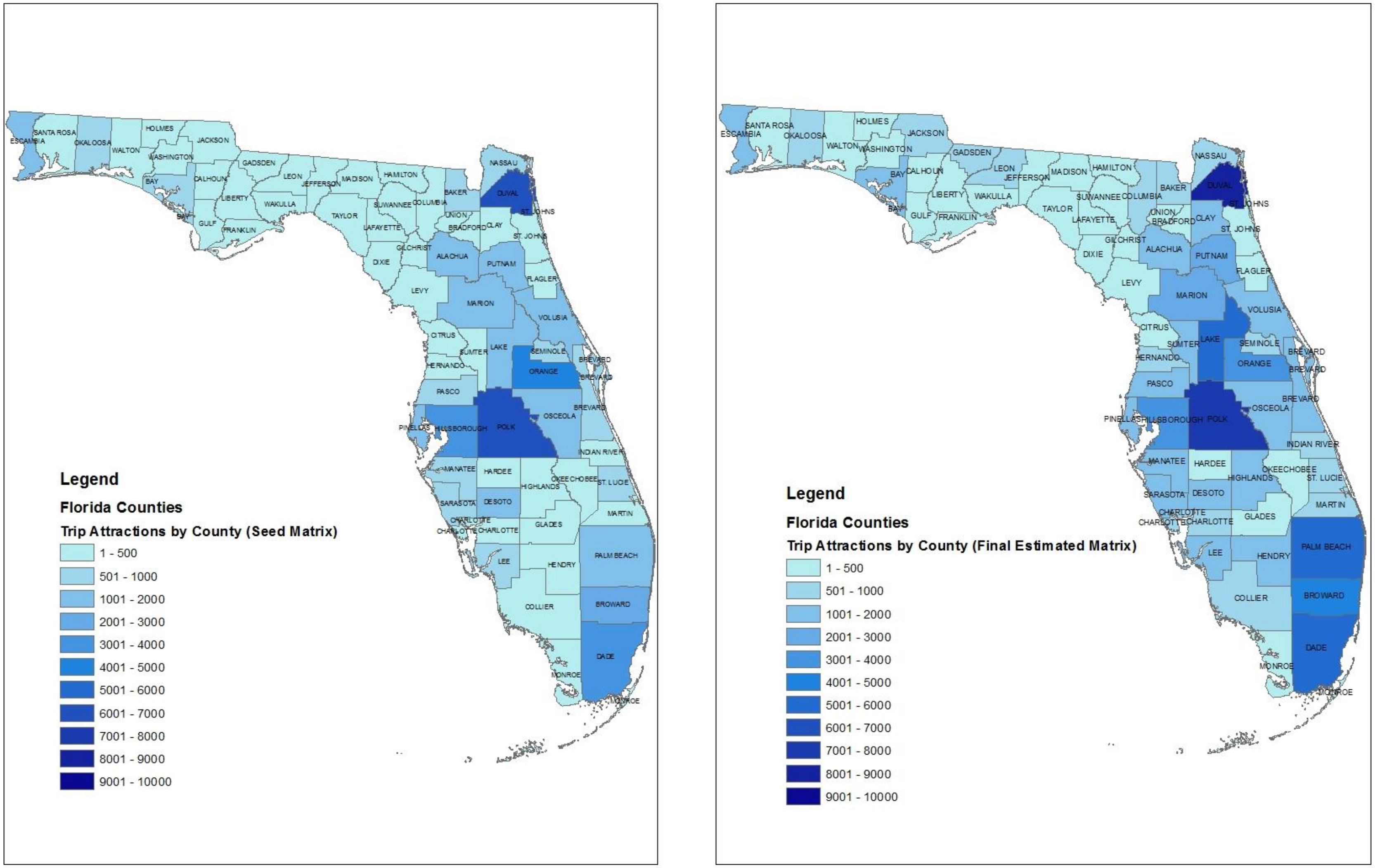


Figure 7.4 Comparison of Trip Attractions by County between Seed OD Matrix and Estimated OD Matrix

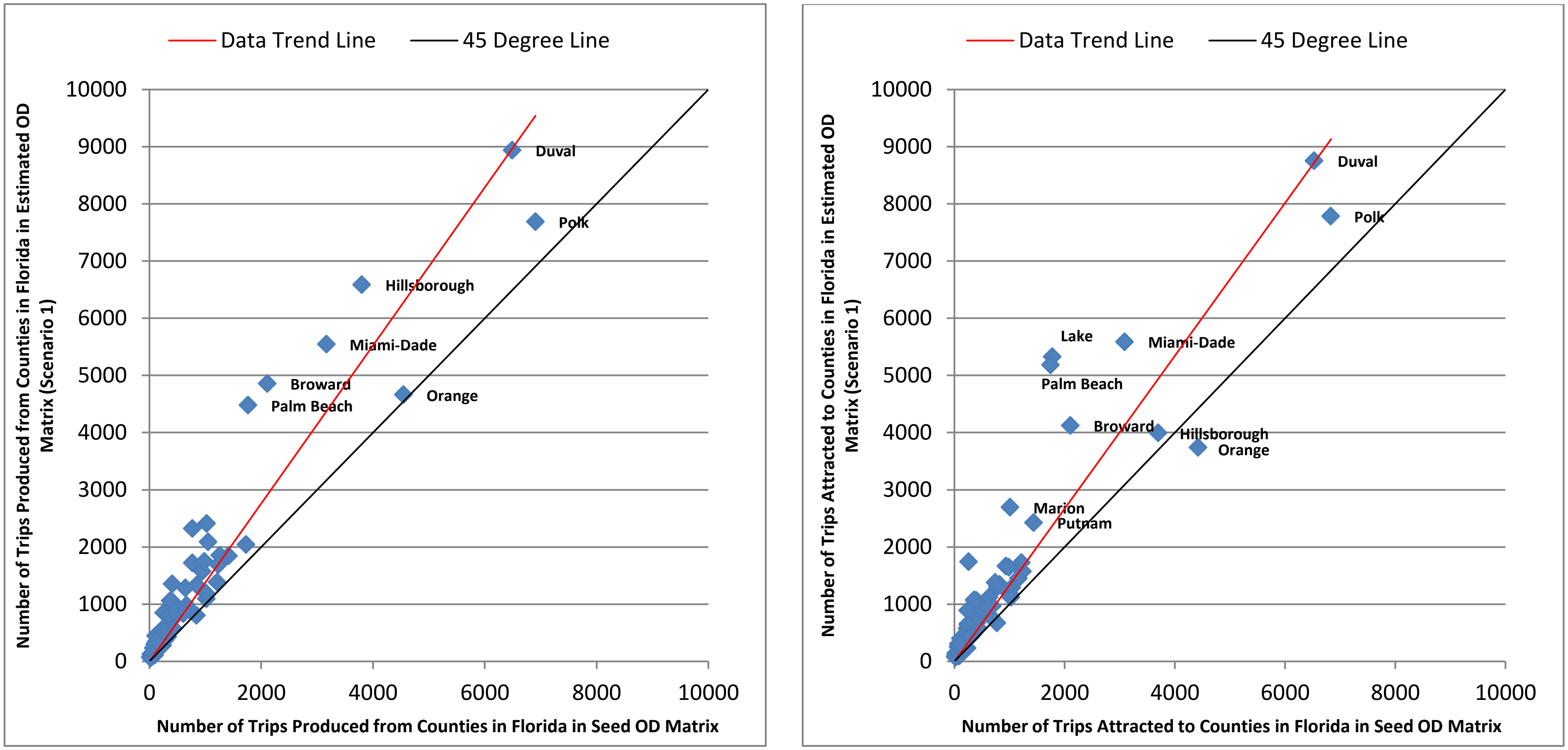


Figure 7.5 Comparison of Truck Trip Productions and Attractions between Seed and Estimated OD Matrix

Table 7.5 Destinations of Heavy Truck Trip Flows from Selected Counties in Florida

Origin \ Destination	Broward	Duval	Hillsborough	Miami-Dade	Orange	Palm Beach	Polk	...	...	
Broward (Seed matrix)	13.6%	5.4%	1.9%	18.5%	3.8%	10.4%	8.5%	...	...	100%
(Estimated matrix)	16.9%	2.0%	0.3%	28.0%	0.3%	29.0%	2.2%	...	...	100%
Duval (Seed matrix)	2.2%	16.0%	2.5%	2.7%	4.5%	1.5%	3.7%	...	...	100%
(Estimated matrix)	2.6%	16.5%	1.4%	2.9%	3.6%	1.4%	2.5%	...	...	100%
Hillsborough (Seed	0.8%	5.6%	14.4%	2.2%	6.4%	0.7%	15.5%	...	...	100%
(Estimated matrix)	0.4%	1.6%	11.1%	0.8%	3.6%	0.6%	15.9%	...	...	100%
Miami-Dade (Seed	10.9%	6.2%	3.0%	15.4%	4.7%	7.5%	7.5%	...	...	100%
(Estimated matrix)	16.4%	0.8%	0.9%	13.4%	0.5%	30.4%	1.2%	...	...	100%
Orange (Seed matrix)	1.8%	9.1%	5.2%	3.1%	10.4%	2.1%	15.1%	...	...	100%
(Estimated matrix)	1.0%	8.3%	4.1%	2.6%	10.7%	1.4%	15.7%	...	...	100%
Palm Beach (Seed	10.6%	7.7%	1.9%	10.2%	6.2%	10.9%	7.8%	...	...	100%
(Estimated matrix)	10.7%	13.0%	0.4%	12.8%	0.4%	15.8%	2.4%	...	...	100%
Polk (Seed matrix)	2.7%	4.3%	8.5%	3.1%	8.2%	2.0%	16.7%	...	...	100%
(Estimated matrix)	1.8%	2.3%	9.1%	1.8%	6.8%	2.0%	18.1%	...	...	100%

Table 7.6 Origins of Heavy Truck Trip Flows to Selected Counties in Florida

Origin \ Destination	Broward	Duval	Hillsborough	Lake	Miami-Dade	Orange	Palm Beach	Polk
Broward (Seed matrix)	13.6%	1.7%	1.1%	1.6%	12.6%	1.8%	12.5%	2.6%
(Estimated matrix)	19.9%	1.1%	0.3%	0.3%	24.3%	0.3%	27.2%	1.4%
Duval (Seed matrix)	6.7%	15.9%	4.4%	1.8%	5.7%	6.5%	5.5%	3.5%
(Estimated matrix)	5.7%	16.8%	3.2%	1.3%	4.7%	8.6%	2.4%	2.8%
Hillsborough (Seed matrix)	1.5%	3.2%	14.8%	4.7%	2.7%	5.5%	1.6%	8.6%
(Estimated matrix)	0.6%	1.2%	18.4%	14.8%	1.0%	6.3%	0.8%	13.5%
Miami-Dade (Seed matrix)	16.4%	3.0%	2.6%	4.2%	15.8%	3.4%	13.5%	3.5%
(Estimated matrix)	22.0%	0.5%	1.3%	0.5%	13.3%	0.7%	32.5%	0.9%
Orange (Seed matrix)	3.8%	6.3%	6.4%	13.7%	4.5%	10.7%	5.6%	10.1%
(Estimated matrix)	1.1%	4.4%	4.8%	3.6%	2.2%	13.3%	1.2%	9.4%
Palm Beach (Seed matrix)	8.9%	2.1%	0.9%	1.1%	5.8%	2.5%	11.0%	2.0%
(Estimated matrix)	11.6%	6.6%	0.4%	0.1%	10.2%	0.5%	13.7%	1.4%
Polk (Seed matrix)	8.8%	4.5%	16.0%	19.7%	6.9%	12.8%	7.7%	16.9%
(Estimated matrix)	3.4%	2.0%	17.4%	5.9%	2.4%	13.9%	3.0%	17.9%
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Total (Seed matrix)	100%	100%	100%	100%	100%	100%	100%	100%
Total (Estimated matrix)	100%	100%	100%	100%	100%	100%	100%	100%

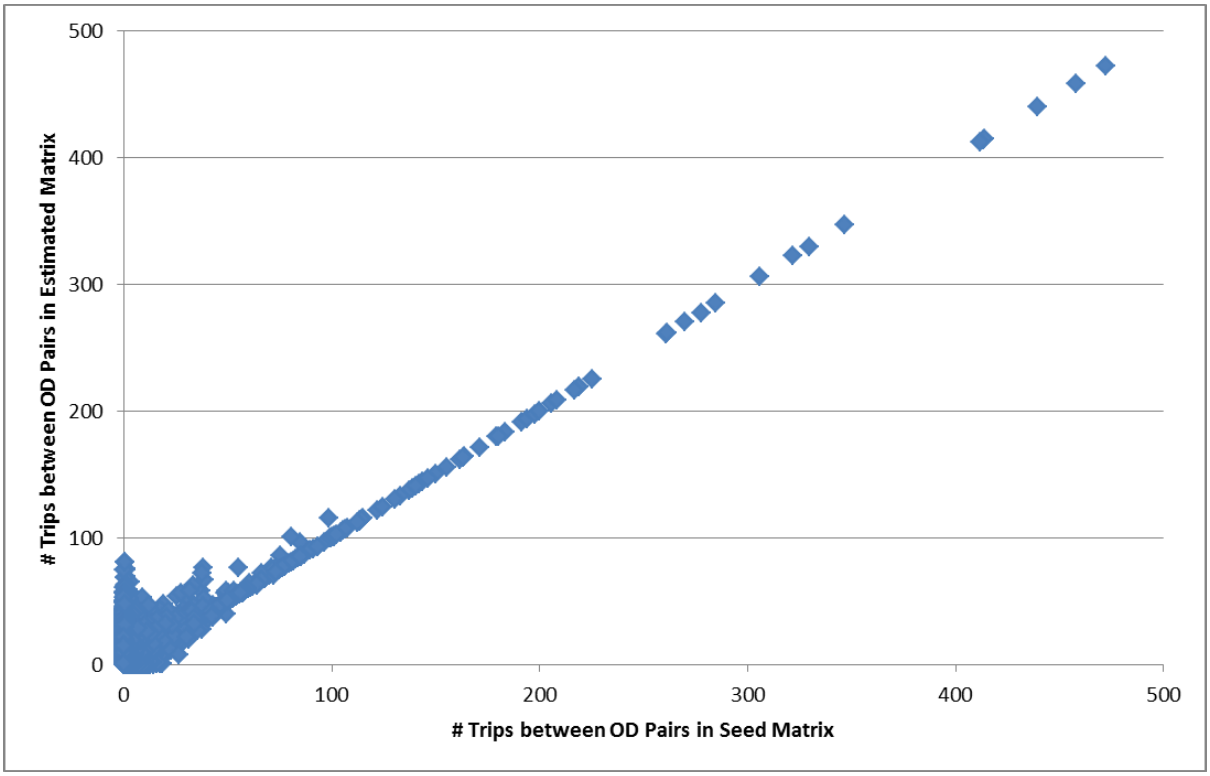


Figure 7.6 Comparison of Truck Trip Flows between Each OD Pair in the Seed and Estimated OD Matrix

CHAPTER 8: CONCLUSIONS AND FUTURE RESEARCH

8.1 Conclusions

This thesis explored the use of a large sample of GPS records of trucks movements with the focus on the state of Florida. More than 145 million GPS records were acquired from American Transportation Research Institute (ATRI) and the truck movements were analyzed and the following outcomes were obtained:

8.1.1 Algorithms to Convert ATRI's Raw GPS Data Streams into a Database of Truck Trips

The raw GPS data streams from ATRI need to be converted into a truck trip format to realize the full potential of the data for freight planning applications. This research resulted in algorithms to convert the raw GPS data into a database of truck trips. Two different algorithms were developed - one for ATRI's truck-GPS data with instantaneous speed information in the GPS records and the other for data without instantaneous speed information. The results from both the algorithms were subject to different validations to confirm that the algorithms can be used to extract accurate trip information from raw GPS data provided by ATRI. The resulting database comprises over 1.2 Million truck trips traveling within, into, and out of the state. This database of truck trips can be used for a variety of purposes, including the development of truck travel characteristics and origin-destination truck flow patterns for different geographical regions in Florida. The database can be used to calibrate and validate the next generation statewide freight travel demand model being developed by FDOT. In future work, this database can potentially be used to develop data on truck trip-chaining and logistics patterns in the state.

8.1.2 Analysis of Truck Travel Characteristics in Florida

The truck trip database developed from four months of ATRI's truck-GPS data was used to analyze a variety of truck travel characteristics in the state of Florida. The truck travel characteristics analyzed include trip duration, trip length, trip speed, time-of-day profiles, and origin-destination flows. Each of these characteristics were derived at a statewide level as well as for different regions in the state – Jacksonville, Tampa Bay, Orlando, Miami, and rest of Florida – defined based on the freight analysis framework (FAF) zoning system. It was found that the time-of-day profiles for all the four FAF zones in Florida show a single peak during the late morning period as opposed to a bi-modal peak typically observed for passenger travels for morning and evening peak periods.

8.1.3 Assessment of ATRI's Truck GPS Data and Its Coverage of Truck Traffic in Florida

This thesis resulted in a better understanding of ATRI's truck-GPS data in terms of its coverage of truck traffic in the state of Florida. This includes deriving insights on: (a) the types of trucks (e.g., heavy trucks and light trucks) present in the data, and (b) the geographical coverage of the data in Florida, and (c) the proportion of the truck traffic flows in the state covered by the data.

Based on discussions with ATRI and anecdotal evidence, it is known that the major sources of ATRI data are freight shipping companies that own large trucking fleets, which typically comprise tractor-trailer combinations (or FHWA vehicle type classification 8 to 13). However, a close observation of the data, through following the trucks on Google Earth and examining some travel characteristics of individual trucks, suggests that the data has a small but not-negligible proportion of trucks that are likely to be smaller trucks that do not necessarily haul freight over long distances. The project used simple rules to divide the data into two categories:

long-haul trucks or heavy trucks (considered to be FHWA vehicle classification 8 to 13), and short-haul trucks or light trucks. Specifically, trucks that did not make at least one trip of 100 mile length in a two week period and those that made more than 5 trips per day were considered "short-haul" trucks. Out of a total of 169,714 unique truck IDs in the data, about 4.6% were labeled as short-haul trucks (or light trucks) and separated from the remaining long-haul trucks (or heavy trucks). In future work, it will be useful to derive better definitions of heavy trucks and light trucks. While heavy trucks are of primary interest to FLSWM (assuming these are the long-haul freight carrying trucks), light trucks are also of potential use for updating the non-freight truck models. Further, extracting sufficient data on light trucks can potentially help understand truck movement within urban regions as well (because a considerable proportion of truck traffic in urban areas tends to comprise light trucks).

ATRI's truck GPS data represents a large sample of truck flows within, coming into, and going out of the State. However, the sample is not a census of all trucks traveling in the state. And it is unknown what proportion of heavy truck flows in the state is represented by this data. To address this question, truck traffic flows implied by one-week of ATRI's truck GPS data was compared with truck counts data from Florida Telemetric Traffic Monitoring Stations Sites (TTMS) truck traffic counts at over 200 locations in the state. The results from this analysis suggest that, at an aggregate level, the ATRI data provides 10.1% coverage of heavy truck flows observed in Florida. When the coverage was examined separately for different highway facilities (based on functional classification), the results suggest that ATRI data provides a representative coverage of truck flows through different types of highway facilities in the state.

The coverage of ATRI data was examined for different geographical regions in the state by examining the spatial distribution of the number of truck trips generated at a TAZ-level and at a county-level geography. In addition, the percentage of heavy truck traffic covered by ATRI data at different locations was examined. All these examinations suggest potential geographical differences in the extent to which ATRI data represents heavy truck traffic volumes at different locations in the state. For instance, truck trips generated from the Polk County were much higher than those generated from the Hillsborough and Miami Dade Counties. Further, the percentage of heavy truck traffic covered by ATRI data in the southern part of Florida (within Miami) and the southern stretch of I-75 is slightly lower compared to the coverage in the northern and central Florida regions. Such geographical differences (or spatial biases) can potentially be adjusted by combining ATRI's truck-GPS data with observed data on truck traffic flows at different locations in the state (from FDOT's TTMS traffic counting program).

8.1.4 Origin-Destination (OD) Tables of Statewide Truck Flows

An important outcome of this thesis was to use ATRI's truck-GPS data in combination with other available data to derive origin-destination (OD) tables of freight truck flows within, into, and out of the state of Florida. The OD flow tables were derived at the following levels of geographic resolution:

- TAZs of the Florida Statewide Model (FLSWM), with Florida and the rest of the country divided into about 6000 TAZs
- County-level resolution, where Florida is represented at a county-level resolution and the rest of the country is represented at a state-level resolution
- State-level resolution, where Florida and the rest of the country are represented at a state-level resolution.

As part of this task, first, the truck trip database developed from four months of ATRI's GPS data was converted into OD tables at the traffic analysis zone (TAZ)-level spatial resolution used in the Florida Statewide Model (FLSWM). Such an OD table derived only from the ATRI data, however, is not necessarily representative of the freight truck flows in the state. This is because the ATRI data does not comprise the census of trucks in the state. Although it is a large sample, it is not necessarily a random sample and is likely to have spatial biases in its representation of truck flows in the state. To address these issues, the OD tables derived from the ATRI data were combined with observed truck traffic volumes at different locations in the state (and outside the state) to derive a more robust origin-destination table that is representative of the freight truck flows within, into, and out of the state. To achieve this, a mathematical procedure called origin-destination matrix estimation (ODME) method is employed to combine the OD flow table generated from the ATRI data with observed truck traffic volume information at different locations within and outside Florida. The OD flow table estimated from the ODME procedure is likely to better represent the heavy truck traffic volumes in the state as it utilizes the observed truck traffic volumes to weigh the ATRI data-derived truck OD flow tables.

The truck flow OD tables derived in the project can be used for a variety of different purposes, as below:

- To understand the spatial distribution of truck travel demand in the region,
- To validate, calibrate, and update the heavy truck modeling components of FLSWM,
- Analysis of truck flows into and out of selected locations in the state.

8.2 Future Research

The work conducted in this project can be extended in several directions of interest to Florida, as discussed in this section.

8.2.1 Explore the Use of ATRI's Truck-GPS Data for Understanding Urban Freight Movements and Statewide Non-freight Truck Flows

A significant part of this thesis was aimed at generating data useful for the Florida Statewide Model (FLSWM); for example, statewide truck OD flows. In future work, it will be useful to explore if ATRI's truck-GPS data can be used to develop and understand truck flows within urban areas as well. As mentioned earlier, while the data predominantly comprises heavy trucks that tend to haul freight over long distances, the data contains light trucks that tend to serve local distribution and delivery. Extracting such trucks and analyzing their travel patterns to understand the extent to which the data covers urban truck flows is a fruitful avenue for future research. In addition, it will be useful to understand the gaps in this data in terms of what types of trucks and what industries are not represented in this data. This can potentially help in augmenting the data with other data sources for use in regional freight travel demand models.

It will be worth exploring the use of this data for generating non-freight truck travel patterns for FLSWM. Currently, the FLSWM uses quick-response freight manual (QRFM) techniques for modeling non-freight truck flows. While QRFM techniques are useful in the absence of data on non-freight truck flows, it is preferable to develop Florida-specific data to better model non-freight truck flows in the state.

8.2.2 Improvements to Origin-Destination Matrix Estimation (ODME)

The origin-destination matrix estimation performed in this study can be improved in different ways. First, the observed truck traffic volumes used in this study come from FDOT's telemetric traffic monitoring program (for over 200 locations in Florida), Georgia Department of transportation (for several locations in Georgia) and FHWA's VTRIS database (for locations outside Florida and Georgia). Within the timeframe of this study, the research team could not gather robust data on observed truck traffic volumes in several southeastern states. For example, there were little to no traffic count information for states such as Tennessee and a few other southeastern states. Providing robust truck traffic count data for southeastern states into the ODME procedure can potentially help in better estimating the truck flows into and out of the state. Second, the truck GPS-data used to develop the seed OD flow table (an input into the ODME procedure) is Florida-centric. The data does not necessarily provide a reliable picture of the truck flows between origins and destinations outside Florida. Therefore using more of ATRI's truck-GPS data, at least for the southeastern states other than Florida, can potentially help improve the ODME results. Third, the ODME procedure itself can be improved in different ways: (a) by allowing different constraints that are specific to different OD pairs, (b) by exploring the different weighting schemes used to expand the seed matrix, and (c) by improving the traffic assignment procedure based on observed route choice patterns of trucks. In this context, analyzing the route choice behavior of trucks is an important avenue for future research both for improving existing procedures used for traffic assignment and for improving the ODME procedure for estimating truck OD flows.

8.2.3 Development of Truck Trip Chaining and Logistics Data

This thesis resulted in procedures for identifying truck stops and truck trips from raw GPS data. This work can be extended further to derive truck trip chaining and logistics patterns from the data. In doing so, adding detailed land-use information can help in characterizing the truck travel patterns.

REFERENCES

Bernardin, V. L., Avner J., Sort J. (2011). Using Large Sample GPS Data to Develop an Improved Truck Trip Table for the Indiana Statewide Model, *TRB Innovation Papers.*

Gédéon, C., Florian, M., & Crainic, T. G. (1993). Determining origin-destination matrices and optimal multiproduct flows for freight transportation over multimodal networks. *Transportation Research Part B: Methodological*, *27*(5), 351-368.

Giuliano, G., Gordon, P., Pan, Q., Park, J., & Wang, L. (2010). Estimating freight flows for metropolitan area highway networks using secondary data sources. *Networks and Spatial Economics*, *10*(1), 73-91.

Gonzalez-Calderon, C., Holguin-Veras, J., Ban, X. (2012). Tour-based freight origin-destination synthesis. *Association for European Transport and Contributors 2012*.

Holguin-Veras, J., & Patil, G. R. (2007). Integrated origin-destination synthesis model for freight with commodity-based and empty trip models. *Transportation Research Record: Journal of the Transportation Research Board*, *2008*(1), 60-66.

Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Vallabhaneni, L., & Nippani, S. (2014). Development of a Tour-Based Truck Travel Demand Model using Truck GPS Data. In *Transportation Research Board 93rd Annual Meeting* (No. 14-4293).

Ma, X., McCormack, E. D., & Wang, Y. (2011). Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program. *Transportation Research Record: Journal of the Transportation Research Board*, *2246*(1), 92-100.

Ogden, K. W. (1978). The distribution of truck trips and commodity flow in urban areas: A gravity model analysis. *Transportation Research*, *12*(2), 131-137.

Stopher, P. R., Bullock, P., & Jiang, Q. (2003). Visualising trips and travel characteristics from GPS data. *Road and transport research*, *12*(2).

Tamin, O. Z., & Willumsen, L. G. (1989). Transport demand model estimation from traffic counts. *Transportation*, *16*(1), 3-26.

Wardrop, J. G. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers, Part II, Vol. 1,* pages 325–378, 1952.

Zargari, S. A., & Hamedani, S. Y. (2006). Estimation of freight OD matrix using waybill data and traffic counts in Iran roads. *Iranian Journal of Science & Technology, Transaction B, Engineering*, *30*(B1).