GPS Data for Truck-Route Choice Analysis of Port Everglades Petroleum Commodity Flow

BDV25-977-17

Final Report

Prepared for:
Florida Department of Transportation
Dr. Min-Tang Li, Project Manager
Mr. Frank Tabatabaee, Project Manager

Prepared by:
UNIVERSITY OF SOUTH FLORIDA COLLEGE OF ENGINEERING
University of South Florida
Dr. Abdul R. Pinjari
Mr. Mohammadreza Kamali
Ms. Trang Luong, EIT
Dr. Seckin Ozkul

January 2016

DISCLAIMER

METRIC CONVERSION

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km

TECHNICAL REPORT DOCUMENTATION

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle *GPS Data for Truck-Route Choice Analysis of Port Everglades Petroleum Commodity Flow*		5. Report Date *November 2015* 6. Performing Organization Code *USF*
7. Author(s) *Abdul R. Pinjari, Mohammadreza Kamali, Trang Luong, Seckin Ozkul*		8. Performing Organization Report No.
9. Performing Organization Name and Address *College of Engineering* *Department of Civil and Environmental Engineering* *University of South Florida* *4202 E. Fowler Avenue, ENB118, Tampa FL 33620*		10. Work Unit No. (TRAIS) 11. Contract or Grant No. *BDV25-977-17*
12. Sponsoring Agency Name and Address *Florida Department of Transportation* *Research Center* *605 Suwannee Street, MS 30* *Tallahassee, FL 32399-0450*		13. Type of Report and Period Covered *Final Report* *2/25/2015– 10/31/2015* 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This project investigated the use of truck-GPS data to derive travel patterns of petroleum-tanker trucks delivering petroleum commodities from Port Everglades (PEV) to gas stations and other fuel recipients in South Florida. The travel patterns of focus are the trip-chaining patterns and the travel paths (i.e., routes) of tanker trucks from Port Everglades (PEV) to their final delivery points. To this end, two months of truck-GPS data available from the American Transportation Research Institute (ATRI) were utilized for the 12-county region served by Port Everglades, including Miami-Dade, Broward, Palm Beach, Monroe, Martin, St. Lucie, Indian River, Okeechobee, Glades, Hendry, Lee, and Collier Counties. The truck route choice data and trip-chaining patterns derived in this project were utilized within a larger project aimed at exploring innovative ways to collect and analyze petroleum flow data in and out of Port Everglades. In addition, the procedures and the data developed in the project offer significant opportunities to understand truck trip chaining and route choice behavior in urban regions.

17. Key Word *truck-GPS data, petroleum flows, route choice, map-matching, truck trip chains*	18. Distribution Statement		
19. Security Classify. (of this report) *Unclassified*	20. Security Classify. (of this page) *Unclassified*	21. No. of Pages *66*	22. Price

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Department of Transportation (FDOT) managers on this project, Dr. Min-Tang Li from FDOT District 4 office and Mr. Frank Tabatabaee from FDOT central office, for their full support, constant guidance, and valuable feedback during this research. The authors also thank Mr. Peter Haliburton from Cambridge Systematics and Mr. Yongqiang Wu from Gannett Fleming for their continuous support and feedback as well as for securing fuel recipient datasets needed for the project. The truck-GPS data utilized in this project was provided by the American Transportation Research Institute (ATRI). Our thanks are due to Mr. Dan Murray and Mr. Jeff Short from ATRI, who played a key role in the provision of the truck-GPS data needed for this project. This project benefited from the data made available by Neil Kutchera from Port Everglades on the locations of the petroleum terminals in the port as well as his insight on the travel behavior of petroleum-tanker trucks. The project utilized a map-matching technique proposed in a recent paper coauthored by the first two authors of this report with Krishnan Viswanathan and Alireza Ermagun to derive truck travel routes using GPS data.

EXECUTIVE SUMMARY

Project Objectives

The goal of this task work order is to provide support to the Florida Department of Transportation (FDOT) District 4 office in an innovative, proof-of-concept data collection pilot project titled "*Port Everglades Petroleum Commodity Flow Pilot Study*". Funded by Federal Highway Administration (FHWA) through the SHRP2 C20 program, the purpose of this project is to explore innovative methodologies to collect and analyze petroleum flow data in and out of Port Everglades to better understand the supply-demand dynamics of the petroleum commodities in South Florida.

The information needed for the above project includes the petroleum origin and destination flow data describing the supply side and demand side of the petroleum products. Also needed is the truck travel path (or route choice) information of the petroleum-tanker trucks for their travel between Port Everglades (PEV) and the delivery locations such as gas stations and other fuel recipients.

The goal of this task work order is to use truck-GPS data from American Transportation Research Institute (ATRI) to derive petroleum-tanker trucks' travel paths (or routes) that the trucks take to travel from Port Everglades to their final delivery points including the intermediate stops that tanker trucks make along the route. To this end, the following specific tasks were performed in the project:

1) Acquire ATRI's truck GPS data needed for the project,
2) Gather, refine, and augment the existing fuel recipient data needed for the project,
3) Derive a database of petroleum-tanker truck trips from ATRI's raw GPS data,
4) Derive the trip chains of trucks traveling to and/or from PEV, and
5) Derive travel paths (or routes) of all trips of all trip chains going through PEV.

Each of these tasks is briefly discussed next.

Acquire ATRI's truck-GPS data needed for the project

For this project, the research team utilized ATRI's truck-GPS data for two months – September 2014 and March 2015 – in the following 12 Counties of South Florida: Miami-Dade, Broward, Palm Beach, Monroe, Martin, St. Lucie, Indian River, Okeechobee, Glades, Hendry, Lee, and Collier. To extract data from petroleum-tanker trucks that deliver petroleum products from PEV to the 12-county region, the project team first identified the petroleum terminals in PEV from which the tanker trucks load petroleum products. The boundaries of each petroleum terminal were identified using a polygon drawn around the terminal. Subsequently, ATRI extracted GPS data of all trucks in their database that were found within these polygons during September 2014 and March 2015. For all trucks that were found inside these polygons during the two months, ATRI extracted the GPS data of their travel in the 12-county region (as well as a few miles beyond) for the two months mentioned above. This resulted in a total of 242,418 raw GPS data records for subsequent use in the project.

To protect confidentiality of the raw GPS data, ATRI required University of South Florida (USF) to sign a non-disclosure agreement (NDA). According to the NDA, the raw GPS data must not be shared with anyone outside the research team. However, the agreement allows for the aggregate results and data products from the research to be submitted to FDOT.

Gather, refine, and enrich existing fuel recipients data for the 12-county region

The research team was provided with two sources of data on fuel recipients for the 12-county region, namely data from the Department of Revenue (DOR) fuel tax database and gas station location data from HERE. The DOR data included 2,315 geocoded facilities that comprised gas stations as well as other fuel recipients in the 12-county region such as government agencies, agricultural establishments, and industrial establishments that receive fuel via petroleum-tanker trucks from PEV. HERE data comprised 1,841 geocoded gas stations in the 12-county region. Neither datasets had a complete record of all fuel recipients in the 12-county region, even after combining the two datasets. In addition, both datasets had inaccuracies in the geocoding of gas stations or other fuel recipients. The research team rectified the locations of incorrectly geocoded gas stations and other fuel recipients in both datasets and also augmented around 90 new gas stations to the dataset. The new gas stations were found when the locations frequently visited (many of them were gas stations) by trucks in ATRI data were examined vis-à-vis the gas station locations in the DOR and HERE datasets. The refined and enriched final fuel recipient dataset that included data from DOR and HERE datasets as well as the 90 new gas stations was used for this project.

Derive a database of petroleum-tanker truck trips from ATRI's raw GPS data

The raw GPS data from ATRI were converted into a database of truck trips. To do so, the algorithm developed (for converting GPS data into trips) in a previous FDOT project by Pinjari et al. (2014) was utilized. The original algorithm developed for the previous FDOT project was geared toward deriving long-haul freight trips. This project, however, needs trips (and trip chains) of petroleum-tanker trucks that make local deliveries within and around urban areas. Therefore, some appropriate modifications were made to the algorithm to meet the project's needs. Specifically, for this project, the algorithm was modified to be able to derive local delivery trips in urban areas and to be able to derive trip chains (i.e., continuous chain of trips). The modified algorithm was used to convert two months of raw data comprising 242,418 GPS records into a total of 14,598 truck trips.

To validate and refine the trips derived from raw data, the land-use descriptions of the trip ends (origins and destinations) were identified for all 14,598 trips. To do so, all trip ends were first grouped into a smaller number of spatial "clusters" based on spatial proximity of trip ends. Subsequently, the land-use descriptions of the trip-end clusters were visualized and identified using Google Earth as well as the GIS layer of fuel recipients discussed earlier. Subsequently, the cluster-level land-use descriptions were assigned to all the individual trip ends within the cluster.

The spatial coordinates of the above-identified clusters at gas stations (i.e., clusters with land-use description as gas station) were compared with the coordinates of nearby gas stations in the HERE and DOR datasets. This helped in rectifying the geocoded locations of some gas stations in the combined HERE and DOR data as well as in identifying about 90 new gas stations not present in the HERE or DOR datasets.

The land-use descriptions of the trip ends were used to refine the initial set of trips extracted from the ATRI data. For example, trucks that did not serve the primary purpose of delivering petroleum products from PEV to fuel recipients were removed. In addition, trips ending on roadways at the study region boundaries or at the beginning or end of the data collection periods were removed. After all these procedures, a refined and validated set of 12,649 trips remained in the final trip file for subsequent use.

Derive trip chains of trucks traveling to and/or from PEV

In this task, a total of 1,320 trip chains were derived from 11,918 trips made by 92 unique trucks. Eight hundred and seven of these trip chains visited PEV at least once. Such trip chains mostly started or ended in either PEV or gas stations. The remaining 513 trip chains did not visit PEV. Those trip chains were mostly incomplete because the trucks had exited the boundaries of the study region (and GPS data was not available to derive the trip chains beyond the study region boundary). The 807 trip chains that visited PEV at least once were found to have excellent connectivity and consistency within trip chains, which is paramount for deriving routes for the entire trip chain. A descriptive analysis of the trip chaining characteristics was carried out using this data.

Derive travel paths (or routes) of all trips of all trip chains going through PEV

The last step in this project was to derive the travel paths (or routes) made by petroleum-tanker trucks for all trips in all the trip chains derived from the raw data. The procedure for deriving routes using GPS data consisted of two broad steps, namely (a) map-matching and (b) route generation. A simple yet effective procedure developed by Kamali et al. (2016, in press) was used to derive the routes. The travel routes were generated for a total of 11,907 trips that belonged to the 1,320 trip chains mentioned earlier. The route derived for each trip includes all the network links traversed by the truck between the origin and destination of the trip. Validation of the routes for 50 randomly selected trips indicated a high level of accuracy of the derived routes. The final product of the project was provided in the form of a GIS shapefile of the generated routes for 11,907 trips along with trip-level information such as trip length, trip time, and origin/destination TAZs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 : INTRODUCTION

1.1 Background

The goal of this task work order is to provide support to the Florida Department of Transportation (FDOT) District 4 office in an innovative, proof-of-concept data collection pilot project titled the "*Port Everglades Petroleum Commodity Flow Pilot Study*". Funded by Federal Highway Administration (FHWA) through the SHRP2 C20 program, the purpose of this project is to explore innovative methodologies to collect and analyze petroleum flow data in and out of Port Everglades (PEV) to better understand the supply-demand dynamics of the petroleum commodities in South Florida.

The information needed for the above project includes the petroleum origin and destination flow data describing the supply side and demand side of the petroleum products. Also needed is the truck travel path (or route choice) information of the petroleum-tanker trucks for their travel between PEV and the delivery locations such as gas stations and other fuel recipients.

The goal of this task work order is to use truck-GPS data from the American Transportation Research Institute (ATRI) to derive petroleum-tanker trucks' travel paths (or routes) that the trucks took to travel from Port Everglades to their final delivery points, including the intermediate stops that tanker trucks make along the route. To this end, the following specific tasks were performed in the project:

1) Acquire ATRI's truck GPS data needed for the project,
2) Gather, refine, and augment the existing fuel recipient data needed for the project,
3) Derive a database of petroleum-tanker truck trips from ATRI's raw GPS data,
4) Derive the trip chains of trucks traveling to and/or from PEV, and
5) Derive travel paths (or routes) of all trips of all trip chains going through PEV.

This report describes the above tasks in detail and presents the results from each task. Specifically, the report is structured as follows. Chapter 2 provides a background on the data used for this project, including American Transportation Research Institute (ATRI)'s truck GPS data and data on fuel recipients in the southeast Florida region. Chapter 3 describes the procedure used for converting ATRI's raw truck-GPS data into truck trips, along with results from validation and refinement of the derived truck trips. In addition, the chapter presents descriptive characteristics of the final validated and refined petroleum-tanker truck trips. Chapter 4 describes the process of deriving trip chains from the trips extracted from ATRI data. In addition, the chapter presents a descriptive analysis of the characteristics of trip chains. Chapter 5 describes the method used to derive travel paths (or routes) of the petroleum-tanker trucks for all trips in all trip chains derived from the raw GPS data. The procedure for deriving routes using GPS data consisted of two broad steps, namely (a) map-matching and (b) route generation. These two steps are explained in detail in this chapter, along with the validation results of the generated routes. An appendix associated with this chapter provides the description of the GIS shapefile of all the generated routes submitted to the FDOT central office. Finally, Chapter 6 concludes the report and identifies avenues for future research.

CHAPTER 2 : OVERVIEW OF ATRI DATA AND FUEL RECIPIENT DATA

2.1 Introduction

This chapter provides a background on the data used for this project, including American Transportation Research Institute (ATRI)'s truck-GPS data and data on fuel recipients (e.g., gas stations) in the southeast Florida region.

2.2 ATRI's truck-GPS data

The American Transportation Research Institute (ATRI) obtains truck position data from partnerships with trucking companies and data vendors throughout the U.S. and North America from a large sample of trucks that use onboard, wireless communications systems such as GPS. For this project, the research team utilized ATRI's truck-GPS data for two months, September 2014 and March 2015. At a minimum each record in ATRI's truck-GPS database included the following information:

- Unit information: a unique identifier for the truck (called Unique truck ID),
- Temporal information: time stamp of the time when the position of the truck was recorded in the following format: MM/DD/YYYY HH:MM:SS, and
- Geographical information: the latitude and longitude that identifies the spatial location of the truck at that time instance.

For each of the two months – September 2014 and March 2015 – ATRI provided data in two different formats: D1 format and D2 format. The characteristics of D1 and D2 data formats are listed below:

- D1 format:
    - Includes the truck's instantaneous speed (or spot speed), in addition to the above-mentioned information, for each GPS record.
    - Unique truck ID rotates (i.e., changes) every 24 hours. That means, even if it is the same truck, the unique truck ID changes every 24 hours around midnight. This makes it difficult to continuously trace the travel patterns of trucks in D1 data beyond 24 hours.
    - A sample record of D1 data looks as below.

Unique truck ID	Time/Date stamp	Speed	Latitude	Longitude
absdefghi12232123	05/03/2011 01:55:55	25	33.915932	-84.494760

- D2 format:
    - The GPS records do not include spot speed of the truck
    - Unique truck ID is static (i.e., the truck ID does not change from day to day). This makes it easy to continuously trace the travel of trucks in D2 data.
    - A sample record of D2 data looks as below.

Unique truck ID	Time/Date stamp	Latitude	Longitude
12232123	05/03/2011 01:55:55	33.915932	-84.494760

It should be noted that the "Unique truck ID" is a random-digit identifier assigned to each vehicle and cannot be used to identify the actual vehicle or to trace the carriers that provided the data. Therefore, one cannot utilize the "Unique truck ID" to identify whether a truck is a petroleum-tanker truck that carries petroleum products. As discussed in Section 2.3, other techniques are used in this project to extract GPS data of only petroleum-tanker trucks from ATRI's database.

To protect confidentiality of the data, ATRI required USF to sign a non-disclosure agreement (NDA). According to the NDA, the raw GPS data shared by ATRI with USF was to be used only for the purpose of analysis by USF researchers. The raw data must not be shared with anyone outside the research team. However, the agreement allows for the aggregate results and data products from the research to be submitted to FDOT as long as the locations of the trip origins, destinations, and intermediate locations are not revealed in a high spatial resolution. Once the NDA was in place, ATRI started sharing the raw truck-GPS data with USF. The data were shared through a secure FTP site that was used throughout the project for transferring data.

2.3 Geographical coverage of ATRI's truck-GPS data

In this project, the research team utilized two months (i.e., September 2014 and March 2015) of ATRI's truck-GPS data for the following 12-county region served by the Port Everglades (PEV): Miami-Dade, Broward, Palm Beach, Monroe, Martin, St. Lucie, Indian River, Okeechobee, Glades, Hendry, Lee, and Collier Counties. To extract such data of petroleum-tanker trucks that serve the purpose of delivering petroleum products from PEV to the 12-county region, the project team first identified the petroleum terminals in PEV from which the petroleum-tanker trucks loaded petroleum products. Initially a GIS shapefile containing 13 terminal points was provided to the research team, as shown in Figure 2.1. After the project management team removed one of these 13 terminals and added two additional terminals, a total of 14 petroleum terminals in PEV were used for subsequent tasks.

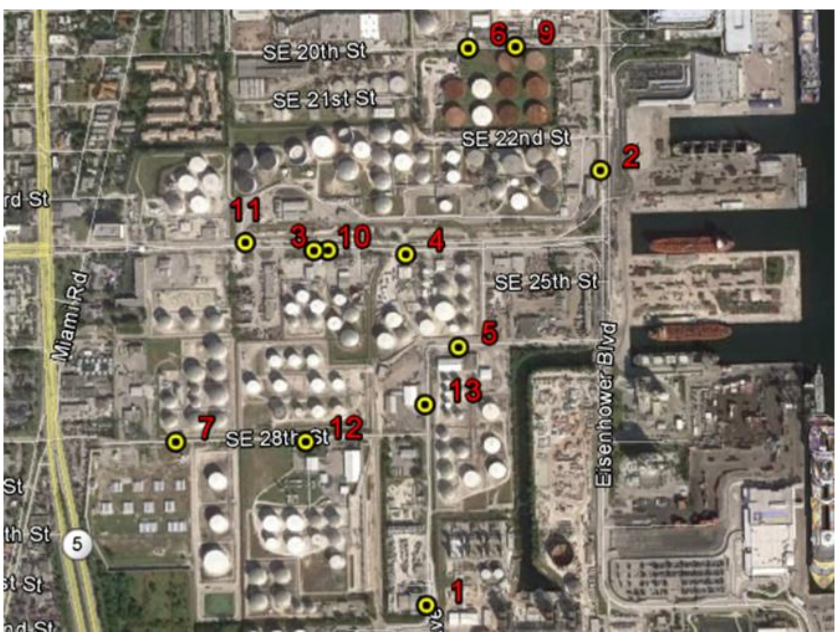


Figure 2.1. Location of petroleum terminals (yellow points) in PEV

To separate petroleum-tanker trucks from other trucks, a polygon was drawn around each terminal so that if a tanker truck (in ATRI's database) had stopped at that terminal to load petroleum commodities, it would be captured in the polygon. As an illustration, Figure 2.2 shows the polygon around terminal 1. The polygon is drawn in red and the terminal area within the polygon is depicted in a yellow circle. It should be noted here that the polygons were drawn tightly around the terminals so that the trucks captured in these polygons are most likely petroleum-tanker trucks (and not other trucks carrying other commodities). As shown in Figure 2.2, the east edge of the polygon is drawn tightly around the terminal so it does not extend into the adjacent roadway to avoid picking up other trucks from this roadway. The rest of the polygons around other terminals can be found in Appendix A. The polygons around all 14 terminals were saved in a GIS shapefile and sent to ATRI. Subsequently, ATRI extracted GPS data of all trucks (i.e., truck IDs) that were found within these polygons during the two months – September 2014 and March 2015.

Figure 2.2. Terminal 1 (in yellow circle) and the red polygon used for GPS data extraction

For all the truck IDs that were found inside the above-described polygons during the two months, ATRI extracted the GPS data for the 12-county region as well as a few miles beyond the 12-county region. The spatial boundary for which the GPS data of petroleum-tanker trucks were extracted is shown in Figure 2.3. The 12-county region is shaded in light blue color whereas the boundary for data collection is shown as a thick blue line. As can be observed, the spatial boundary extends a few miles beyond the 12-county region. The red dots in the figure are the truck-GPS data points extracted by ATRI for this project.

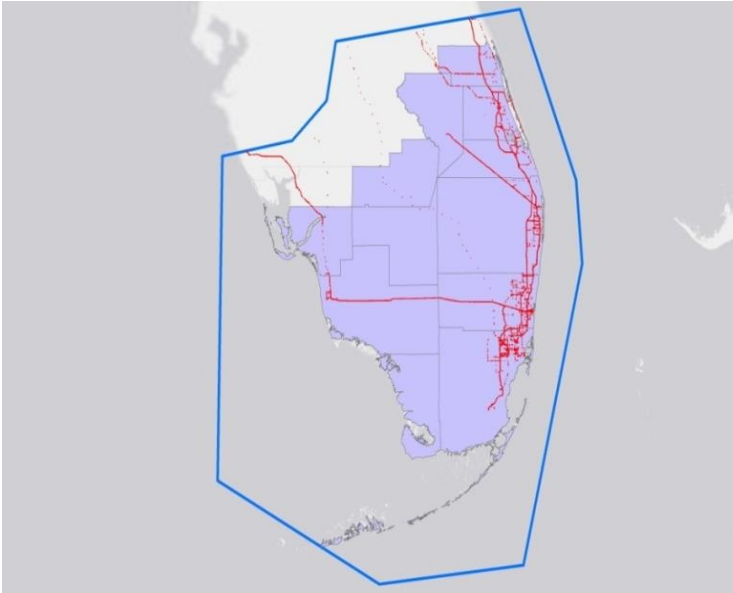


Figure 2.3. Spatial boundary of the data extracted and the GPS coordinates of the raw data

2.4 Descriptive information on ATRI's truck-GPS data

Table 2.1 presents basic descriptive characteristics of the truck-GPS data used for the project (for the two months of September 2014 and March 2015), including such information as the number of unique truck IDs and the total number of GPS points. Several observations can be made from this table. First, for both months, D2 data is significantly richer than D1 data in terms of the number of days the data is available for, the number of GPS records, and the number of trucks. Second, a direct addition of the number of truck IDs shown in 4 columns suggests a total of 141 truck IDs. However, it should be noted that this number is only the sum of all the truck IDs existing in both months of data, and it is not the total number of *unique* trucks in the two-month data. In fact, for D1 datasets, although the truck IDs change from day to day, the research team identified only 42 unique trucks out of the total 45 unique truck IDs. This was done by measuring the spatial and temporal gaps between the last GPS point of current truck ID and the first GPS point of the next truck ID. If the spatial and temporal gaps were less than one mile and less than 60 minutes, respectively, then the two truck IDs were considered to belong to the same truck. For the D2 datasets, 16 truck IDs were repeated between the September 2014 and March 2015 datasets, which brought the original number of 96 truck IDs in the D2 datasets down to 80 (44+52-16) unique trucks. In conclusion, data from a total of 122 unique trucks (42 trucks in D1 dataset + 80 trucks in D2 dataset) were used for this project.

Table 2.1. Attributes of the September 2014 and March 2015 GPS data

	September 2014		March 2015		Total
	D1	D2	D1	D2	
# Days of data	10	30	22	31	94
# Truck IDs	11	44	34	52	141
# GPS Points	8,621	86,606	35,182	112,009	242,418

An important attribute of GPS data is the temporal difference between consecutive GPS points (referred to as ping rate hereafter). The higher the ping rate, the more frequent the GPS data. Another key attribute of GPS data is the physical distance between consecutive GPS points (referred to as spatial gap hereafter). Tables 2.2, 2.3, 2.4, and 2.5 show cross-tabulations of the spatial gap and ping rates for both D1 and D2 datasets for September 2014 and March 2015. These tables show that D1 data is more frequent than D2 data for both months. Note, however, that the frequency of D2 data is also sufficient for deriving travel routes. Specifically, as can be observed from Table 2.3 and 2.5, close to 90 percent of the D2 data has a ping rate of less than 15 minutes.

Table 2.2. Cross-tabulation of ping rate versus spatial gap for September 2014 – D1 data

Ping rate (minutes) \ Spatial gap (miles)	< 1	1-5	5-15	15 <	Sum
< 1	45.1%	9.4%	0.0%	0.0%	54.5%
1-2	22.1%	21.1%	0.0%	0.0%	43.2%
2-5	0.4%	0.3%	0.0%	0.0%	0.7%
5-15	0.8%	0.1%	0.0%	0.0%	1.0%
15-20	0.2%	0.0%	0.0%	0.0%	0.2%
20-25	0.0%	0.0%	0.0%	0.0%	0.0%
25-30	0.1%	0.0%	0.0%	0.0%	0.1%
30-45	0.0%	0.0%	0.0%	0.0%	0.0%
45-60	0.2%	0.0%	0.0%	0.0%	0.2%
60-75	0.1%	0.0%	0.0%	0.0%	0.1%
> 75	0.0%	0.0%	0.0%	0.0%	0.0%
Sum	68.9%	31.0%	0.1%	0.0%	100.0%

Table 2.3. Cross-tabulation of ping rate versus spatial gap for September 2014 – D2 data

Ping rate (minutes) \ Spatial gap (miles)	< 1	1-5	5-15	> 15	Sum
< 1	17.1%	0.2%	0.0%	0.0%	17.2%
1-2	8.2%	0.7%	0.0%	0.0%	9.0%
2-5	11.3%	9.4%	10.0%	0.0%	30.8%
5-15	10.4%	7.7%	11.8%	1.6%	31.6%
15-20	2.8%	0.1%	1.0%	0.7%	4.6%
20-25	1.6%	0.0%	0.0%	0.0%	1.6%
25-30	1.5%	0.0%	0.0%	0.0%	1.5%
30-45	1.5%	0.0%	0.0%	0.0%	1.5%
45-60	0.5%	0.0%	0.0%	0.0%	0.5%
60-75	0.2%	0.0%	0.0%	0.0%	0.2%
> 75	1.0%	0.3%	0.2%	0.1%	1.6%
Sum	56.2%	18.4%	23.0%	2.4%	100.0%

Table 2.4. Cross-tabulation of ping rate versus spatial gap for March 2015 – D1 data

Ping rate (minutes) \ Spatial gap (miles)	< 1	1-5	5-15	15 <	Sum
< 1	42.7%	10.0%	0.0%	0.0%	52.7%
1-2	21.1%	22.4%	0.0%	0.0%	43.5%
2-5	0.9%	0.4%	0.0%	0.0%	1.2%
5-15	1.6%	0.1%	0.1%	0.0%	1.8%
15-20	0.3%	0.0%	0.0%	0.0%	0.3%
20-25	0.1%	0.0%	0.0%	0.0%	0.1%
25-30	0.1%	0.0%	0.0%	0.0%	0.1%
30-45	0.1%	0.0%	0.0%	0.0%	0.1%
45-60	0.1%	0.0%	0.0%	0.0%	0.1%
60-75	0.0%	0.0%	0.0%	0.0%	0.0%
> 75	0.1%	0.0%	0.0%	0.0%	0.1%
Sum	67.0%	32.9%	0.1%	0.0%	100.0%

Table 2.5. Cross-tabulation of ping rate versus spatial gap for March 2015 – D2 data

Ping rate (minutes) \ Spatial gap (miles)	< 1	1-5	5-15	15 <	Sum
< 1	19.0%	0.2%	0.0%	0.0%	19.2%
1-2	7.8%	1.2%	0.0%	0.0%	9.0%
2-5	9.3%	9.9%	10.8%	0.0%	30.0%
5-15	9.9%	7.5%	12.8%	1.4%	31.6%
15-20	2.6%	0.0%	0.6%	0.6%	3.9%
20-25	1.5%	0.0%	0.0%	0.0%	1.5%
25-30	1.3%	0.0%	0.0%	0.0%	1.3%
30-45	1.3%	0.0%	0.0%	0.0%	1.3%
45-60	0.6%	0.0%	0.0%	0.0%	0.6%
60-75	0.2%	0.0%	0.0%	0.0%	0.2%
> 75	1.0%	0.4%	0.1%	0.1%	1.5%
Sum	54.3%	19.2%	24.3%	2.1%	100.0%

2.5 Fuel recipient data

Gas stations are among the main delivery destinations of petroleum-tanker trucks that load fuel commodities at PEV. In addition to gas stations, there are other fuel recipients in the 12-county region, such as government agencies, agricultural establishments, and industrial establishments that receive fuel via petroleum-tanker trucks from PEV. Data of such fuel recipients, including gas stations, was secured from two different datasets. One dataset came from the Department of Revenue (DOR) database of fuel tax receipts (hereafter referred to as DOR data). The second dataset was purchased by FDOT from HERE data (hereafter referred to as HERE data). The two datasets are discussed below. A third source of information for gas stations is the ATRI data (i.e., the gas stations frequently visited by trucks in the ATRI data), which was used to augment the aforementioned two data sources.

2.5.1 Department of Revenue database of fuel tax receipts (DOR data)

This data set included 2,315 geocoded facilities comprising gas stations and other fuel recipients such as agricultural, industrial, and government facilities. Gas stations account for 69% of the facilities in DOR data. The remaining facilities are "other fuel recipients." When the geocoded gas station locations in this database were compared with several gas station locations from Google Earth, it was evident that this data (i.e., the geocoded database of 2,315 fuel recipients) was incomplete in its coverage of fuel recipients in the 12-county region. In addition, the geocoded locations of several gas stations were found to be inaccurate when compared against those in Google Earth. However, a strength of the DOR data is that it is the only source of data available in this project on fuel recipients other than gas stations.

2.5.2 HERE data

Version 2.11 of HERE data that was released in January 2015 was used in this project. HERE data comprised 1,841 gas stations in the 12-county region. There were no additional facility types other than gas stations in this dataset. HERE data was also incomplete in terms of encompassing all gas station in the 12-county region. In addition, inaccuracies were found in the geocoded locations of some gas stations in HERE data as well (when verified in Google Earth).

2.5.3 Combination of DOR and HERE data

As mentioned above, neither DOR data nor HERE data were complete in their coverage of all fuel recipients associated with PEV in the 12-county region. Moreover, while there was some degree of spatial overlap (either full overlap or close proximity) in the geocoded gas stations between the two datasets, not all gas stations in one dataset were found in the other dataset. Given this, the research team decided to combine both DOR and HERE datasets to generate a single dataset of fuel recipients in the 12-county region. In such combined dataset, the individual points complement each other in locations where the other dataset does not provide sufficient coverage. The geocoded locations of the fuel recipients in the combined data were verified against those in Google Earth. If needed, the locations in the combined database were rectified.

2.5.4 Additional gas station locations found using ATRI data

In the later tasks of this project, the research team found several additional locations that were not in the above described combined database of fuel recipients but were frequently visited by the trucks in ATRI data (see Chapter 3 for the procedure to derive truck-trip stops from ATRI data). Plotting such locations in Google Earth revealed that they were also gas stations, albeit they were not present in either the DOR dataset or the HERE dataset. The research team found about 90 such new gas station locations, which were augmented to the above data of fuel recipients obtained from DOR and HERE.

CHAPTER 3 : DERIVING TRUCK TRIPS

3.1 Introduction

This chapter describes the procedure used for converting ATRI's raw truck-GPS data into truck trips. The chapter also presents results from validation and refinement of the derived truck trips. The validation process includes verification of the accuracy of truck trip ends. The refinement process includes determination and elimination of trucks that are less likely to have carried petroleum products from PEV (such trucks might have carried goods other than petroleum products). In addition, the chapter presents descriptive characteristics of the final validated and refined petroleum-tanker truck trips.

3.2 Algorithm description

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

1) Clean, read, and sort raw GPS data in a chronological order for each truck ID. At the end of this step, all the GPS data belonging to each truck ID is grouped together in the chronological order.
2) Identify an initial set of truck trip stops (i.e., trip ends) based on spatial movement, time gap, and speed between consecutive GPS points. In this step, a truck is considered to have stopped at a destination if it stops (i.e., if the average travel speed between two consecutive GPS points is less than 5 mph) for at least 5 minutes. A truck stop of less than 5-minute duration is considered to be a traffic stop (i.e., not a valid destination) and, therefore, considered a part of the travel between origin and destination.
3) Conduct quality checks and refine or eliminate trips that do not satisfy quality criteria

The above identifying truck trip-ends procedure is based on an algorithm developed in a previous FDOT research project conducted by Pinjari et al. (2014). The original algorithm developed for the previous FDOT project was geared toward deriving long-haul freight trips. This project, however, needs trips (and trip chains) from petroleum-tanker trucks that make local deliveries around urban areas. Therefore, some appropriate modifications were made to the algorithm to meet the project's needs. Specifically, for this project, the algorithm was modified to derive local delivery trips in urban areas. These changes are detailed in Section 3.2.2.3.

3.2.1 Data cleaning and sorting

The raw GPS data was 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. Both D1 and D2 data showed that each truck ID was present in the dataset for at least one day. Therefore, all the truck IDs were kept in the dataset. The data were then sorted in a chronological order for each truck ID, beginning from the GPS record with the earliest date and time stamp.

3.2.2 Identification of 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 details of the algorithmic procedure in this step are presented in Figure 3.1, which is modified from the algorithm developed by Pinjari et al. (2014). Following is a list of the terms used in the algorithm along with their definitions:

1) *Travel distance* (td): Spatial (geodetic) distance between two consecutive GPS records.
2) *Travel time* (trt): Time gap between the two consecutive GPS records.
3) *Average travel speed* (trs): Average travel speed between consecutive GPS records (td/trt).
4) *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.
5) *Trip time* (tpt): Total time taken to travel from origin of the trip to the current GPS point.
6) *Trip speed* (tps): Average speed of the trip between the origin and the current GPS point.
7) *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)
8) *Destination dwell-time* (ddwt): Total duration the truck stops at the destination of a trip.
9) *Stop dwell-time* (sdwt): Duration of an intermediate stop (e.g., traffic stop).
10) *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.

3.2.2.1 Determination of truck stops and moving instances

An important component of the algorithm involved determining whether a truck was at a stop (i.e., rest) or in motion. As can be observed from the flowchart (Figure 3.1), the primary condition used to determine whether a truck was at a stop (which could be an origin, a destination, or simply an intermediate stop) or it was moving was based on the average travel speed between consecutive GPS data points. A cut-off speed of 5 mph was used, which means if the average travel speed between consecutive GPS records was less than 5 mph (i.e., if trs < 5 mph), the truck was assumed to have stopped.

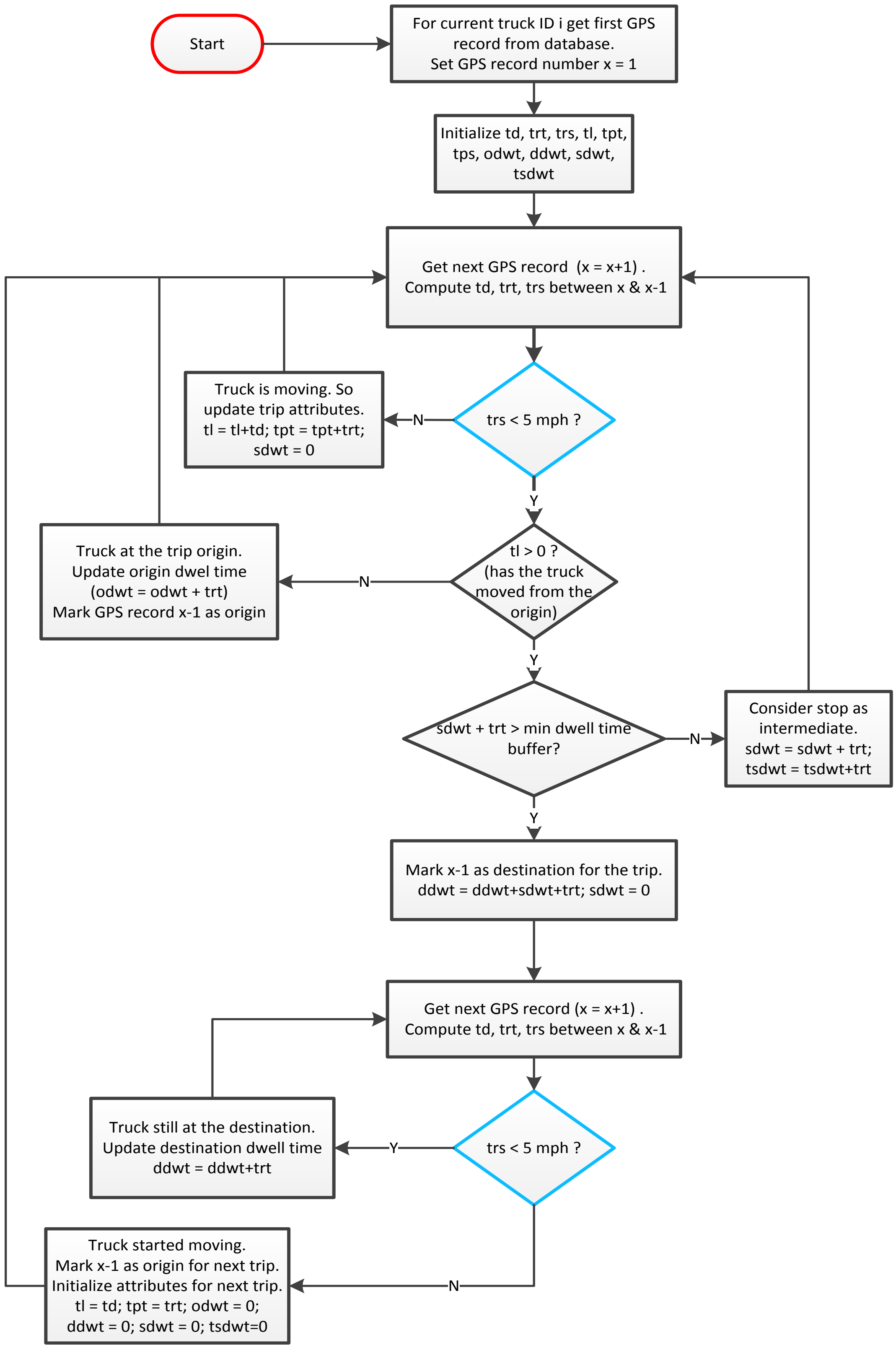


Figure 3.1. Algorithm for identifying truck trip ends from raw GPS data (modified from Pinjari et al. 2014)

3.2.2.2 Separation of intermediate traffic stops from trip destinations

The next step in the algorithm is to differentiate between traffic stops and real trip destinations. Pinjari et al. (2014) and Thakur et al. (2015) explain in full detail how this process is done. In a nutshell, once the algorithm detected a stop (i.e., trs < 5 mph) for a trip in progress (i.e., tl > 0), it started accruing the stop dwell-time (sdwt) based on the time elapsed between successive GPS data points (sdwt = sdwt + trt). If the truck started moving again (i.e., if trs > 5 mph) before the stop dwell-time reached the minimum dwell-time buffer value, then the stop was considered an intermediate stop, and the algorithm proceeded to find another stop. On the other hand, if the stop dwell-time exceeded the minimum dwell-time buffer value, then the stop was considered a valid destination. In this project, the minimum dwell-time buffer value was taken as 5 minutes. Stops of less than 5 minute duration were considered as traffic stops and, therefore, not as valid destinations.

3.2.2.3 Specific characteristics of the algorithm designed for this project

The original algorithm developed by Pinjari et al. (2014) was geared toward deriving long-haul freight trips. This project, however, needs trips (and trip chains) from petroleum-tanker trucks that make local deliveries around urban areas. Therefore, the following modifications were made to the algorithm to derive local delivery trips in urban areas.

First, a minimum dwell-time buffer of 5 minutes was used in order to capture all the possible stops that tanker trucks make. This was done to avoid missing any non-traffic stops made by the petroleum-tanker trucks. The Pinjari et al. (2014) algorithm used longer durations for minimum dwell-time buffer because of their focus on long-haul trips.

Second, in the algorithm explained by Pinjari et al. (2014), all the trips shorter than one mile were either removed or added to their previous trips. In this project, however, trips shorter than one mile were retained because short fuel delivery trips could happen, considering the notable number of gas stations located within a mile of PEV as well as within one mile of each other. Besides, it was important to retain such short trips to be able to derive continuous trip chains in subsequent tasks.

Third, Pinjari et al. (2014) eliminated truck stops in rest areas, regardless of the duration of those stops. In this project, however, the research team retained truck stops in rest areas as well. Retaining such trips allows the possibility that petroleum-tanker trucks could deliver fuel to gas stations in rest areas as well.

3.3 Results

The above-discussed procedure was applied to two months of raw data (September 2014 and March 2015) comprising 242,418 raw GPS records. This resulted in 14,598 truck trips. Table 3.1 shows a summary of the trips derived from the data. Summaries are provided for each month of the data, separately. The number of trips extracted, the number of unique truck IDs to which these trips belonged, and the average trip distance and trip speeds are presented for three different types of trips: (a) D1 trips (trips extracted from D1 data), (b) D2 trips (trips extracted from D2 data), and (c) trips from both datasets.

Note from Table 3.1 that the number of trips extracted from D2 data was larger than that from D1 data. For either D1 or D2 dataset, the average trip distances, average trip times, and average trip speeds were similar across the two months. A more detailed analysis of the characteristics of the trips extracted in this project is presented in Section 3.6.

Table 3.1. Descriptive statistics of the 14,598 trips extracted from raw GPS data

		D1 data	D2 data	Both D1 and D2 data
September 2014	Number of raw GPS records	8,621	86,606	95,227
	Number of trips extracted	92	6,435	6,527
	Number of unique truck IDs	11	44	55
	Average trip length (miles)	47	28	28
	Average trip time (minutes)	49	35	35
	Average trip speed (mph)	41	35	35
March 2015	Number of raw GPS records	35,182	112,009	147,191
	Number of trips extracted	344	7,727	8,071
	Number of unique truck IDs	34	52	86
	Average trip length (miles)	41	33	33
	Average trip time (minutes)	51	38	39
	Average trip speed (mph)	43	38	38
All two months	Number of raw GPS records	43,803	198,615	242,418
	Number of trips extracted	436	14,162	14,598
	Number of unique truck IDs	45	96	122
	Average trip length (miles)	42	31	31
	Average trip time (minutes)	51	37	37
	Average trip speed (mph)	43	37	37

3.4 Identification of land-use description at trip-end locations

Several reasons warrant the need to identify the land-use description of the trip origin and destination locations for each (and every) trip derived from the raw GPS data. First, the trip purposes (such as fuel delivery or other purposes) need to be identified based on the land-use descriptions of the trip ends. Second, some gas stations that received fuel from PEV but were not present in either the HERE data or the DOR data need to be identified. Identifying the land-use descriptions of all trip ends resulted in identifying 90 such gas stations, which were augmented to the combined HERE and DOR data discussed in the previous chapter. Third, the longitude and latitude coordinates of some gas stations or other fuel recipients were miscoded in the HERE and DOR data. This task helped in rectifying the coordinates of such gas stations.

In this project, the land-use description for each (and every) trip end was identified from observing the aerial image of the trip-end locations using Google Earth. However, manually examining the location of each and every 14,598 trip origins and 14,598 trip destinations would involve excessive amount of person hours and might lead to human errors. Therefore, all the trip ends were first grouped in to a smaller number of trip-end spatial clusters based on the spatial proximity of the trip ends. The idea is to manually identify the land-use descriptions of a smaller number of trip-end clusters than those of a large number of individual trip ends.

To describe the procedure to identify spatial clusters of the trip ends, the following two terms are used: (1) place ID and (2) unique ID. Place ID is an ID number for a unique location visited by several trips (where several trip ends are clustered together). A gas station is an example of a place ID. Unique ID is a combination of rounded trip end longitude (hereafter X) and latitude (hereafter Y) to three decimal places. A unique ID may be shared by several trip ends whose coordinates match to three decimal places. An example of a unique ID for several trip ends is: 27.122_-82.454. The procedure used to identify trip-end clusters is described below.

1) Round the latitude (Y) and longitude (X) coordinates of each GPS point to three decimal places. Doing so results in grouping all latitudes (and longitudes) within 320 ft of each other into one single rounded X coordinate (and Y coordinate). After rounding, there are unique values of rounded Y coordinates (e.g., 27.122) and X coordinates (e.g., -82.454).
2) For each trip end, combine its rounded Y and X coordinates to create its unique ID. Note that several trip ends may share the same unique ID.
3) For all the trip ends that share the same unique ID, extract the original GPS coordinates (i.e., non-rounded Y and X) of all trip ends and call them "unique ID representatives". In essence, unique ID representatives are the original coordinates of each of the trip ends that share the same unique ID.
4) First run (Do the following for each unique ID):
    a) Sort the dataset of unique ID representatives first by original Y coordinates and then by original X coordinates.
    b) Calculate the distances between all consecutive pairs of unique ID representatives.
    c) Calculate the numerical differences in GPS coordinates between all consecutive pairs of unique ID representatives.
    d) To identify place IDs, combine consecutive unique ID representatives that are less than 1000 ft apart (i.e., keep combining unique ID representatives until a consecutive pair of unique ID representatives is more than 1000 feet apart).
    e) Calculate the total distance between the first and the last unique ID representatives of a place ID. If the total distance is more than 1000 ft, the place ID is then subdivided into two or more place IDs to make the total distance less than 1000 ft.
    f) Select all new place IDs obtained after the first run.
5) Second run:
    a) Sort the first set of new place IDs first by original X and then by original Y coordinates[1] to further group place IDs that may be close enough to each other but were separated in the first run just because the data sorting was done on Y coordinates first. Perform the same steps from 4) b) to 4) e).
    b) Select all of the new place IDs obtained after the second run.
6) Repeat the same procedure for the third and fourth runs and identify the final set of place IDs, because the number of unique place IDs has most likely reached its minimum (i.e.,

[1] Note that the sorting in the first run was done first by original Y and then by original X coordinates.

the GPS data have been combined into the minimum possible number of place IDs or spatial clusters).

7) Label each place ID with "New cluster #," with # ranging from 1 to the total number of clusters.

After the clusters were created, their land-use descriptions were identified using Google Earth. Subsequently, the cluster-level land-use descriptions were assigned to all the individual trip ends within the cluster. In addition, the spatial coordinates of clusters at gas stations (i.e., clusters with land-use description as gas stations) were compared with the coordinates of nearby gas stations in the combined HERE and DOR data. This helped in rectifying the geocoded locations of some gas stations in the combined HERE and DOR data. Specifically, for each cluster, if any of the HERE or DOR gas station locations did not overlap exactly with the location of a cluster but at least one HERE or DOR location was found within 500 ft of the cluster, the corresponding coordinates of HERE or DOR location were updated to the exact coordinates of the cluster. Otherwise, the coordinates of that cluster were recorded as a new gas station or fuel recipient depending on the land-use description of the cluster (this is, if the cluster was not overlapping with or near any of the gas stations or fuel recipients in the HERE or DOR data). In this manner, as mentioned earlier, around 90 new gas stations were augmented to the combined HERE and DOR data of fuel recipients.

Tables 3.2 and 3.3 illustrate the land-use description distribution of origins and destinations, respectively, for all 14,598 trips after assigning the land-use descriptions of the clusters to the individual trip ends falling within the cluster. Table 3.2 shows the land-use descriptions of the trip origins and Table 3.3 shows the land-use descriptions of the trip destinations. As one can be observed from both the tables, around 40% of trip ends are observed in gas stations and at least 35% of trip ends are observed in PEV. Also note a small proportion of trip ends falling in other fuel recipients. This result is reasonable because these trips belong to tanker trucks that mainly deliver fuel from PEV to gas stations and other fuel recipients. Another dominant land-use description of the trip ends in both the tables is distribution centers (about 13% for both trip origins and trip destinations). When the research team closely examined the land-use descriptions of the locations visited by individual trucks, a few trucks in the data were found to have mainly traveled between distribution centers and, therefore, were not delivering fuel from PEV to fuel recipients. Such trucks were eliminated in further steps (see next section). Lastly, some trip ends were observed to be on the roadway. A closer examination suggested that a large proportion of trip ends falling on roadways were near the study area boundary.

Table 3.2. Land-use description distribution for the total 14,598 trip origins

Case	Location	Number of trip origins
1	Gas station	5,957 (40.8 %) trip origins
2	PEV	5,245 (35.9 %) trip origins
3	Other fuel recipients	747 (5.1 %) trip origins
4	Distribution center	1,906 (13.1 %) trip origins
5	On the road	463 (3.2 %) trip origins
6	Rest stop	280 (1.9 %) trip origins
	Total	14,598 (100%) trip origins

Table 3.3. Land-use description distribution for the total 14,598 trip destinations

Case	Location	Number of trip destinations
1	Gas station	5,552 (38 %) trip destinations
2	PEV	5,005 (34.3 %) trip destinations
3	Other fuel recipients	368 (2.5 %) trip destinations
4	Distribution center	1,910 (13.1 %) trip destinations
5	On the road	1,510 (10.3 %) trip destinations
6	Rest stop	253 (1.7 %) trip destinations
	Total	14,598 (100%) trip destinations

3.5 Refinement of the derived trip dataset

The initial set of trips extracted from the ATRI data was refined in different ways before proceeding further. First, trip-purpose distribution of the trips extracted from each unique truck was examined. Based on this, trucks that did not make any visit to/from PEV were removed from the trip file. Similarly, trucks that made only a small proportion of their total trips to PEV and/or fuel recipients but made a large proportion of trips to distribution centers were removed. These trucks most probably were not petroleum-tanker trucks (that serve the purpose of petroleum-product delivery from PEV) and therefore were not of interest in this project. Second, if the first or last trip of a truck has its origin or destination on the roadway, that trip was removed. This was done to have the first/last trip of a truck starting/ending at a valid location other than a roadway. In addition, trips with either origin or the destination or both origin and destination on the roadway were addressed in different ways, as appropriate. These trips were either joined to their next/previous trip or had the trip end replaced by the previous/next trip's destination/origin. Lastly, truck IDs that belonged to the same unique truck in D1 data were identified. Since truck IDs in this data rotate every 24 hours, the same truck may have two different truck IDs in two consecutive days. Such data were combined (into data for a single unique truck ID) based on the spatial and temporal proximity of the last few GPS data points from one day to the first few GPS data points from the next day. Finding those unique trucks and changing the truck ID from the second day back to the truck ID in the first day can help to follow the unique trucks in two consecutive days and build better trip chains. The outcomes from all the data cleaning/refinement tasks described above are listed below:

1) Removed 27 truck IDs that did not mainly deliver fuel commodities from PEV
2) Removed 36 trips at the start or end of data streams of individual trucks whose trip ends fell on the roadway

3) Removed trips with origin or destination on the roadway based on certain criteria. Specifically, trips that had their origin or destination on the roadway but not within one mile from any other trips made by that same truck were removed.
4) Joined D1 data from two consecutive days of data for 3 pairs of truck IDs into data for 3 unique truck IDs.

Tables 3.4 and 3.5 show the land-use description distribution of origins and destinations, respectively, after the above-mentioned trip cleaning/refinement tasks. As can be observed, when compared with Tables 3.2 and 3.3, the proportions of trip ends at "distribution center" and "on the road" have dropped after cleaning and/refining the trip dataset. As a result, a total number of 12,649 cleaned/refined and validated trips belonging to 92 unique truck IDs remained for further analysis.

Table 3.4. Land-use description distribution of 12,649 trip origins after refining the trip dataset

Case	Location	Number of trip origins
1	Gas station	5,683 (44.9 %) trip origins
2	PEV	5,163 (40.8 %) trip origins
3	Other fuel recipients	539 (4.3 %) trip origins
4	Distribution center	694 (5.5 %) trip origins
5	On the road	310 (2.5 %) trip origins
6	Rest stop	260 (2.1 %) trip origins
	Total	12,649 (100 %) trip origins

Table 3.5. Land-use description distribution of 12,649 trip destinations after refining the trip dataset

Case	Location	Number of trip destinations
1	Gas station	5,370 (42.5 %) trip destinations
2	PEV	4,968 (39.3 %) trip destinations
3	Other fuel recipients	278 (2.2 %) trip destinations
4	Distribution center	665 (5.3 %) trip destinations
5	On the road	1,129 (8.9 %) trip destinations
6	Rest stop	239 (1.9 %) trip destinations
	Total	12,649 (100 %) trip destinations

3.6 Characteristics of truck trips derived from ATRI data

This section provides descriptive characteristics of the validated and/or refined 12,649 truck trips derived as described earlier, such as the trip length, trip time, and average trip speed distributions. As illustrated earlier in Figure 2.3, most of the truck-GPS data is concentrated on the east coast of the study region while the rest of the data is stretched to the southwest region. This geographical distribution of GPS points impacts the distribution of trip characteristics, namely, trip length and trip time. Figure 3.2 shows the trip length distribution of all 12,649 trips.

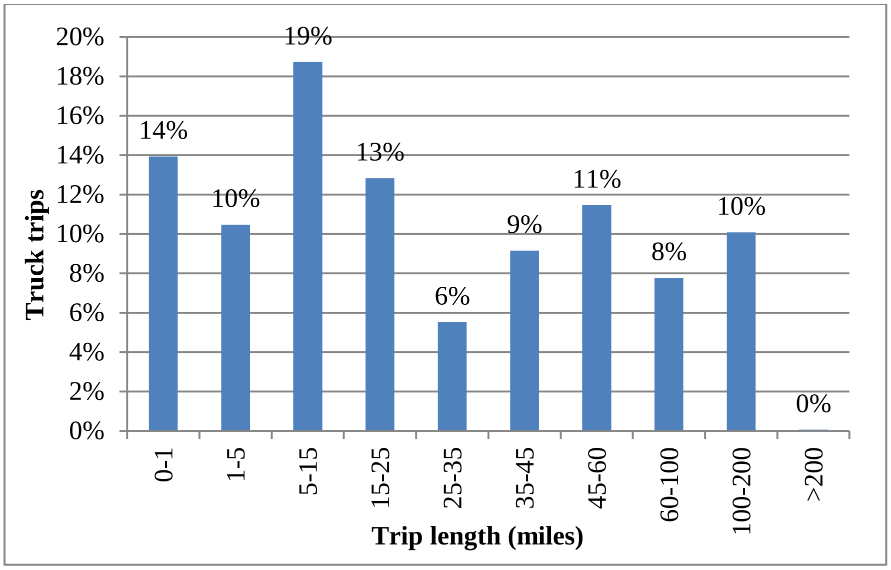


Figure 3.2. Trip length distribution (N = 12,649 trips)

As can be observed, the distribution is divided into two portions. The first portion belongs to trips less than 25 miles which mostly cover the eastern area of 12-county region. The second portion belongs to trips more than 25 miles that stretch into the western area of the 12-county region. This phenomenon is visually illustrated in Figure 3.3. In this figure, blue dots show destination points for trips shorter than 25 miles while red dots represent destination points for trips longer than 25 miles. As can be observed, red dots are stretched toward the west coast of Florida. Blue dots on the other hand, are mostly concentrated near PEV and east coast of Florida.

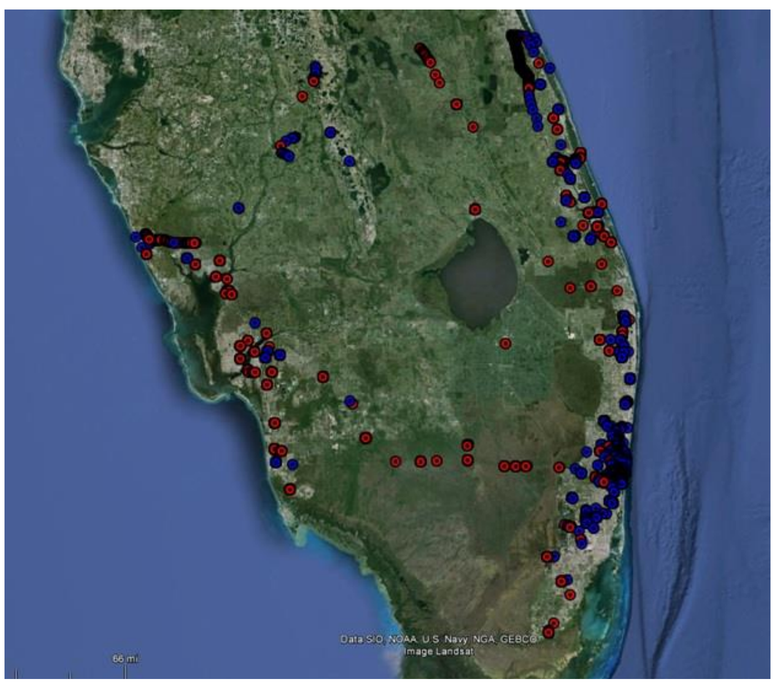


Figure 3.3. Destination points for trips shorter than 25 miles (blue dots) and longer than 25 miles (red dots) for 12,649 trips

Figure 3.4 illustrates the trip time distribution for all the refined/validated 12,649 trips derived. Similar to the trip length distribution, the trip time distribution is divided into two portions; one portion belongs to trips covering the eastern area of 12-county region whose trip time is less than 40 minutes, and the other portion belongs to trips covering the western area of 12-county region whose trip time is more than 40 minutes. It is worth noting here that the trip time represents the total travel time between trip origin and destination, including any traffic congestion stops of less than 5 minutes.

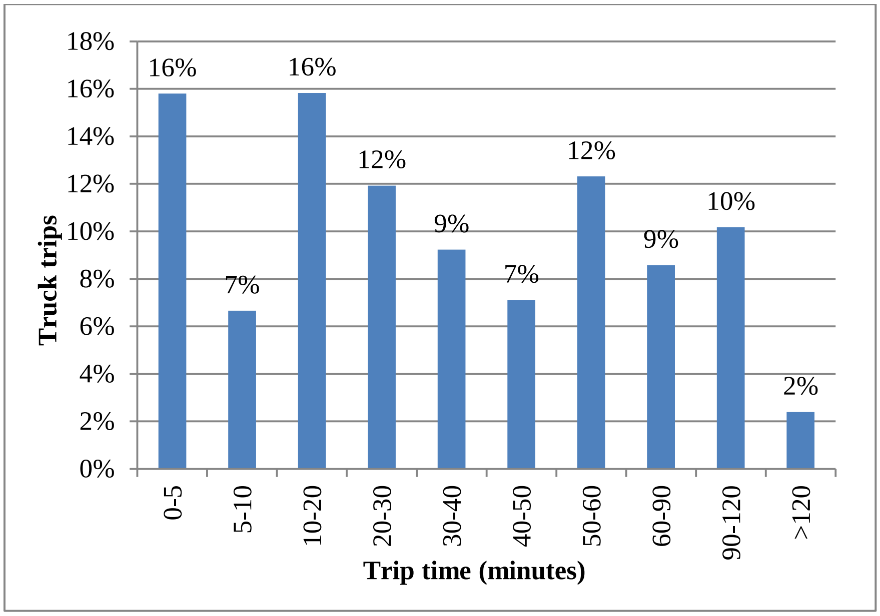


Figure 3.4. Trip time distribution (N = 12,649 trips)

Finally, Figure 3.5 shows the average trip speed distribution for 12,649 trips. As expected, the average speed of a vast majority of trips is less than 70 mph. Appendix B presents these distributions separately for each data type (i.e., D1 or D2).

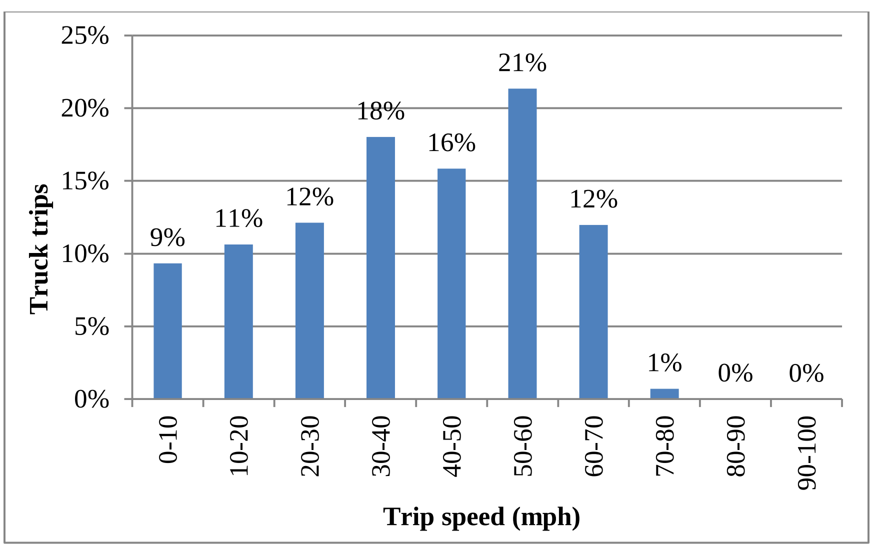


Figure 3.5. Average trip speed distribution (N = 12,469 trips)

CHAPTER 4 : DERIVING TRIP CHAINS

4.1 Introduction
This chapter describes the process of deriving trip chains from the 12,649 validated and/or refined trips extracted from the ATRI data. In addition, the chapter presents a descriptive analysis of the characteristics of trip chains.

4.2 Procedure to derive trip chains
A general and widely used definition of a trip chain is a series of trips made by a vehicle in chronological order. To define trip chains in this project, the 12,649 trips derived from the ATRI data were first used to construct 12,557 trip pairs (12,649 trips – 92 last trips for each truck ID = 12,557 trip pairs). In this context, a trip pair refers to two consecutive trips made by the same truck. Trip 1 and Trip 2 are an example of a trip pair illustrated in Figure 4.1.

4.2.1 Analysis of the spatial and temporal gaps between trip pairs
The extent of spatial gap and temporal gap between a trip pair can be used to determine whether the trip pair belongs to (i.e., both the trips belong to) the same trip chain or not. Figure 4.1 depicts the gap between a trip pair. Spatial gap is the spatial distance between Trip 1's destination and Trip 2's origin. Similarly, temporal gap is the temporal difference between Trip 1's destination and Trip 2's origin. The closer (spatially and temporally) the trips are in a trip pair, the more likely they belong to the same trip chain.

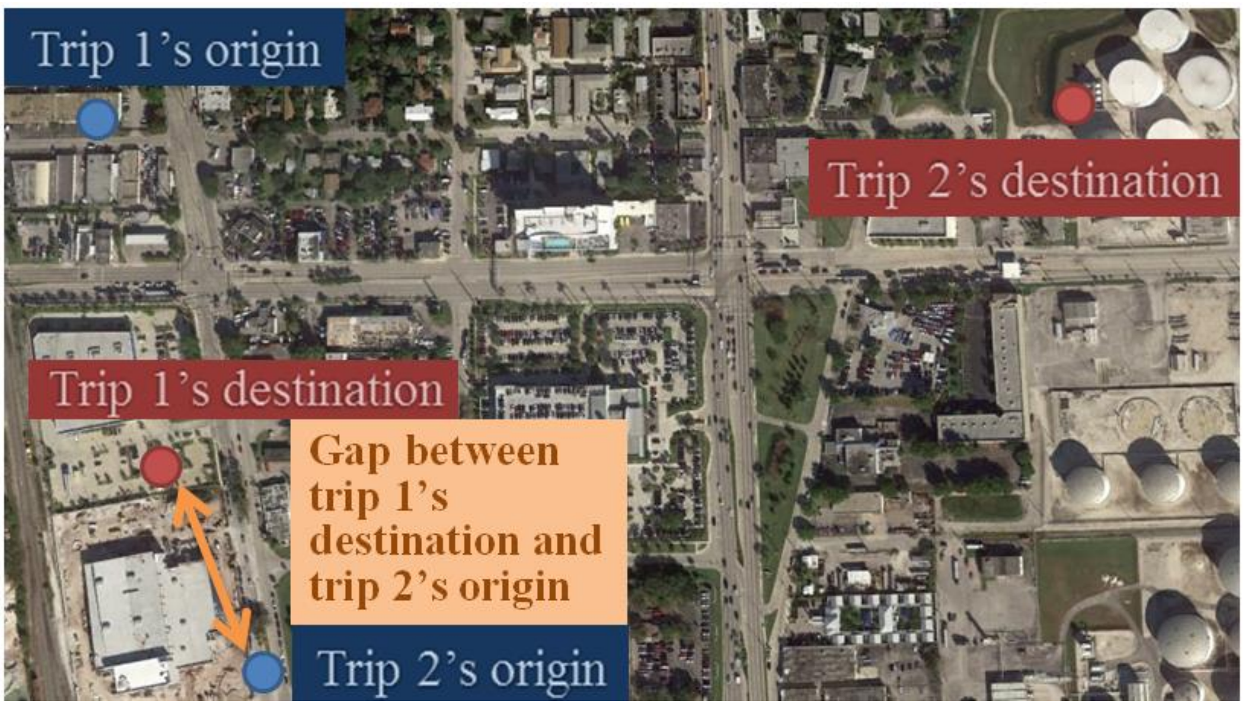


Figure 4.1. Illustration of the gap between two consecutive trips made by the same truck ID

The spatial gap and temporal gap distributions among 12,557 trip pairs are shown in Figures 4.2 and 4.3, respectively. One can observe from Figure 4.2 that the spatial gap is less than 1 mile for about 90% of the trip pairs. Figure 4.3 shows that only 13% of the trip pairs have a temporal gap of more than 240 minutes (four hours), Figure 4.4 depicts a two-way histogram of the relationship between spatial gap and temporal gap for 12,557 trip pairs. One can observe that only a small proportion of the trip pairs are separated by large spatial and temporal gaps.

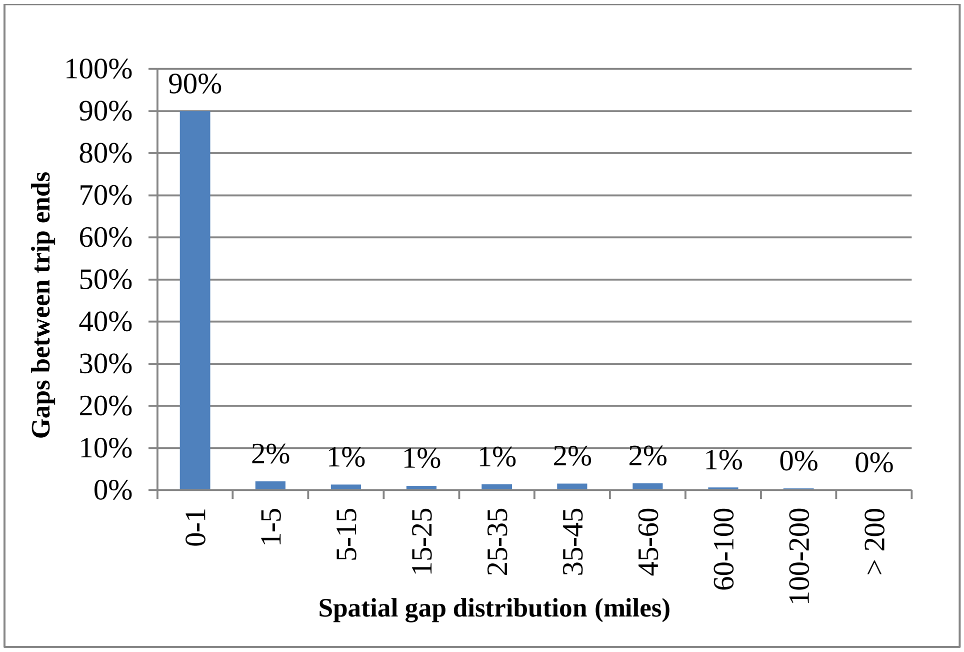


Figure 4.2. Spatial gap distribution among 12,557 trip pairs

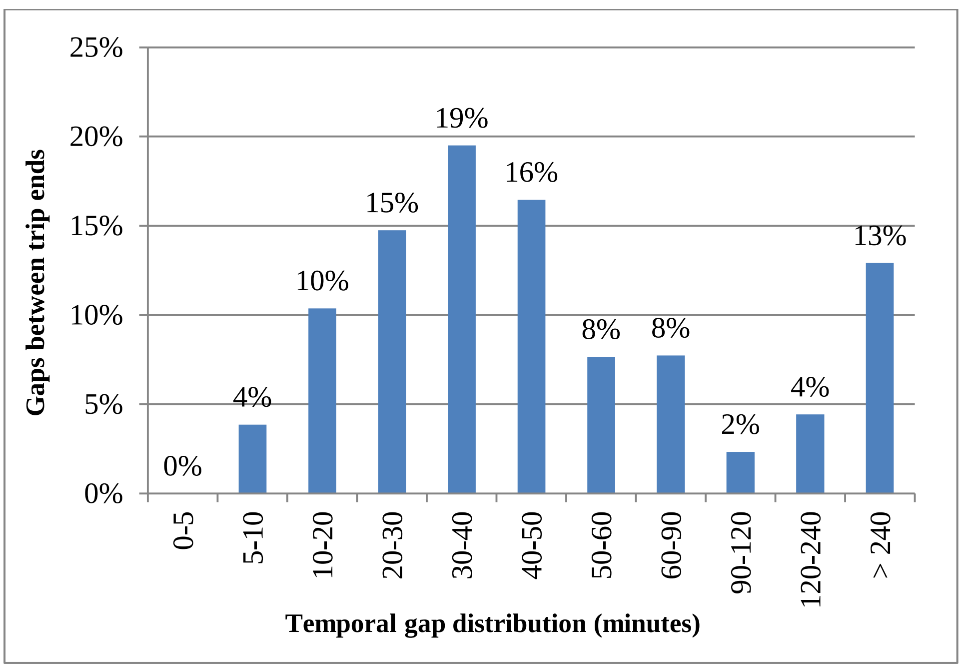


Figure 4.3. Temporal gap distribution among 12,557 trip pairs

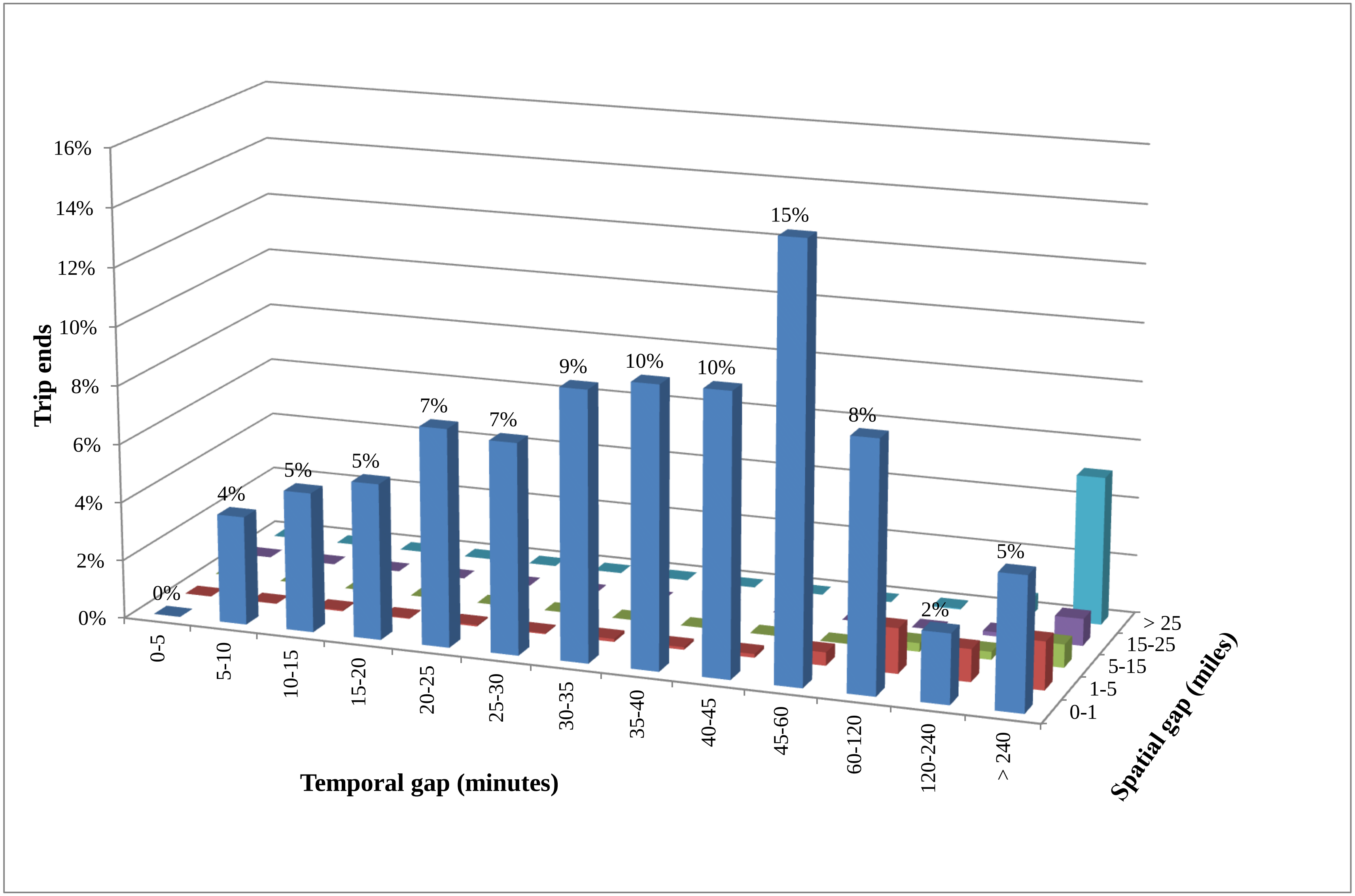


Figure 4.4. Two-way histogram of spatial gap vs. temporal among 12,557 trip pairs

Based on the above distributions of spatial and temporal proximity between trip pairs and discussions with the FDOT project management team, in this project, a trip chain was defined as a series of trips made by a truck based on the following criteria:

(1) The spatial gap between consecutive trips (trip pairs) in the chain should be < 1 mile, and
(2) The temporal gap between consecutive trips (trip pars) in the chain should be < 4 hours.

The above criteria imply that if either the temporal gap between one trip's destination and the next trip's origin is more than four hours or the spatial gap is more than 1 mile, the chain of trips would be broken. Using these criteria, a total of 1,320 trip chains were derived that corresponded to 11,918 trips made by 92 unique truck IDs. The remaining 731 trips were single trips that did not belong to any trip chains based on the criteria explained above (i.e., 12,649 final trips – 11,918 trips satisfying trip chain criteria = 731 single trips that did not satisfy the trip chain criteria).

To further verify the connectivity of trips in the trip chains built as described above, the research team compared the land-use description of the first trip's destination with the following trip's origin for each trip pair. Since the spatial gap of these trip pairs is required to be less than 1 mile and the temporal gap to be less than 4 hours, one would expect the destination land use of the first trip to match with (i.e., be the same as) the origin land use of the second trip. 86% trips (i.e., 10,624 trips) out of the total 12,557 trip pairs had a matching land-use description between the first trip destination and the second trip origin. Over 98% of these trip pairs with matching land uses had a spatial gap of less than 1 mile. This indicates a strong connectivity between each trip pair in the trip chains built as described above. It is also plausible to have the two trip ends of a trip pair with different land-use descriptions (as was the case for a small proportion of trip pairs), because it is possible (albeit for a small number of trip pairs) that a truck might move from one land use to another land use within a distance of 1 mile. In summary, the criteria used to analyze trips and derive trip chains may be viewed as reasonably robust for the purpose of the project.

4.3 Characteristics of the truck trip chains derived from ATRI data

As mentioned earlier, a total of 1,320 trip chains were built from 11,918 trips derived from the ATRI's truck-GPS data belonging to 92 unique truck IDs. Figure 4.5 shows the distribution of number of trips per trip chain. 62% of the trip chains have two to five individual trips per chain. It is noteworthy that there is a considerable proportion of trip chains with high number of trips. For example, 26 trip chains (2% of the total trip chains) shown in Figure 4.5 contain more than 50 individual trips. This may be because trucks could be tracked continuously for several consecutive days in a row.

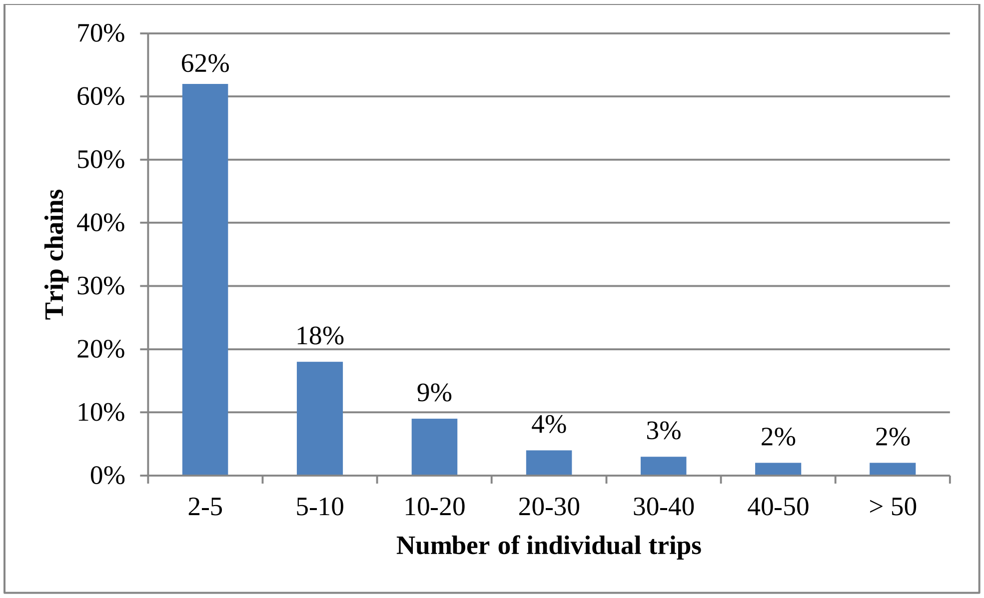


Figure 4.5. Distribution of number of trips per trip chain (N = 1,320 trip chains)

Figures 4.6 and 4.7 show the distributions for trip chain length (total mileage of the trip chain) and trip chain time (total duration of the trip chain), respectively. As can be observed, over 80% of trip chains are between 5 miles and 500 miles in length, and over 70% of trip chains are of duration between 30 minutes and 480 minutes (eight hours). A small percentage of trip chains are more than 1,000 miles and consequently more than 1,440 minutes (one day) long.

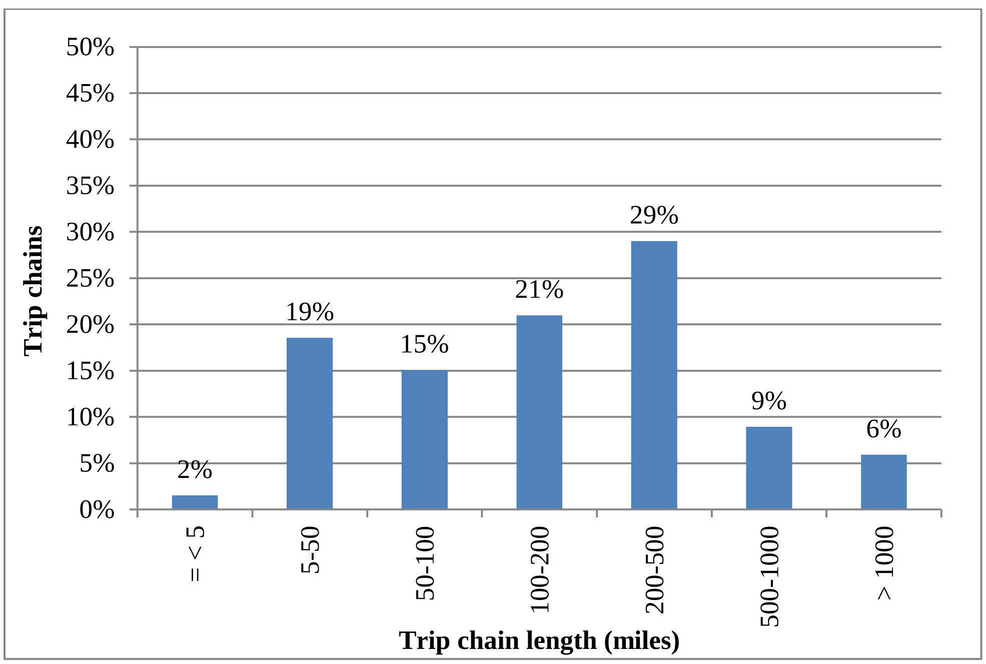


Figure 4.6. Trip chain length distribution (N=1,320 trip chains)

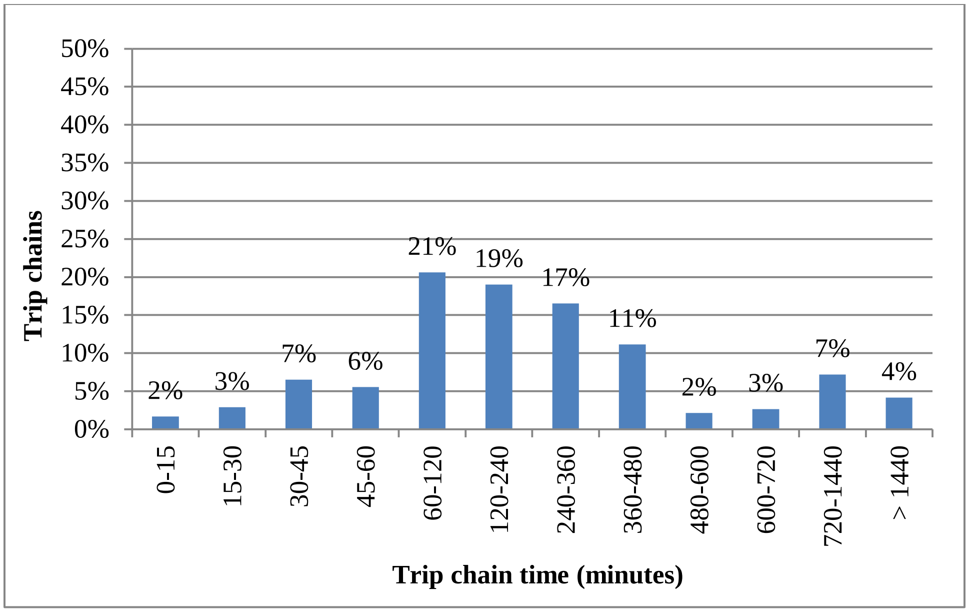


Figure 4.7. Trip chain time distribution (N = 1,320 trip chains)

4.4 Characteristics of the truck trip chains that visit PEV

Eight hundred and seven out of the total of 1,320 trip chains derived in this project included at least one trip end in PEV. The remaining trip chains did not involve any visit to PEV. Since the project focuses on petroleum product deliveries from PEV, this section presents basic descriptive characteristics of the 807 trip chains that included at least one visit to PEV. Table 4.1 illustrates the land-use description distribution of beginning locations (i.e., trip origin of the first trip) and ending locations (i.e., trip destination of the last trip) of the 807 trip chains that visited PEV at least once. One can observe from this table, as expected, that the proportion of trip ends (of both origins and destinations) at PEV and gas stations and other fuel recipients in this table is higher than those of other land uses. The only anomaly is the relatively high share (i.e., about 20%) of trip chain destinations on the road. This is because some tanker trucks deliver fuel commodities outside the study region's boundary and therefore, their last GPS points in the data were captured on the road.

Table 4.1. Land-use description distribution for origins and destinations of 807 trip chains that visited PEV at least once

Case	Location	Number of trip chain origins	Number of trip chain destinations
1	Gas station	249 (30.9 %) trip chain origins	153 (19.0 %) trip chain destinations
2	PEV	388 (48.1 %) trip chain origins	369 (44.7 %) trip chain destinations
3	Other fuel recipients	66 (8.2 %) trip chain origins	66 (8.2 %) trip chain destinations
4	Distribution center	55 (6.8 %) trip chain origins	52 (6.4 %) trip chain destinations
5	On the road	32 (4.0 %) trip chain origins	159 (19.7 %) trip chain destinations
6	Rest stop	17 (2.1 %) trip chain origins	8 (1.0 %) trip chain destinations
	Total	807 (100 %) trip chain origins	807 (100 %) trip chain destinations

Figure 4.8 shows the distribution of number of trips per trip chain for the 807 trip chains that involved at least one visit to PEV. 41% of the trip chains have two to five individual trips. Figures 4.9 and 4.10 show the distributions for trip chain length (total mileage of trip chain) and trip chain time (total duration of trip chain), respectively, for 807 trip chains that visited PEV at least once. As can be observed, around 90% of the trip chains are 5 miles to 500 miles long, and over 70% of the trip chains last between 30 minutes and 480 minutes (eight hours). Only a small percentage of the trip chains are more than 1,000 miles and consequently more than 1,440 minutes (one day) long.

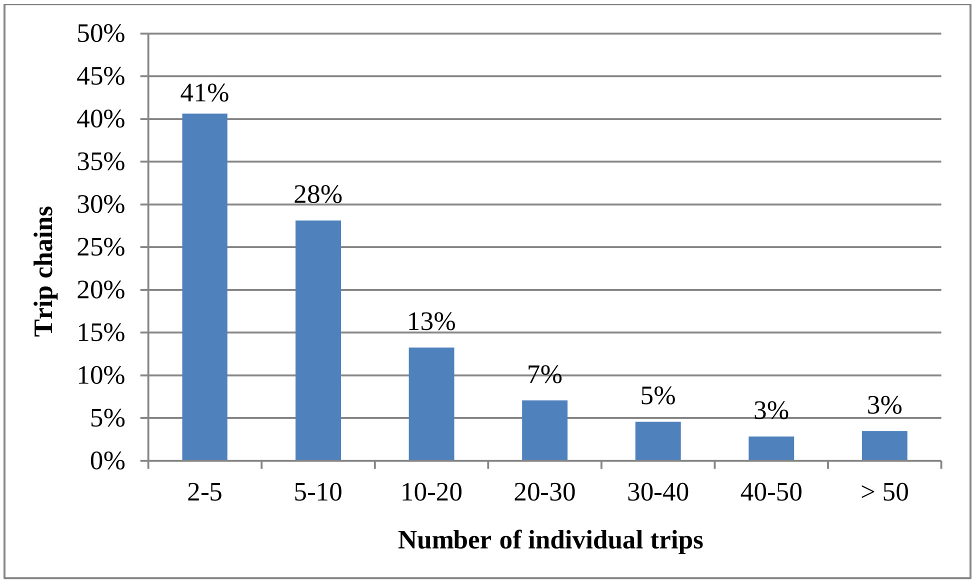


Figure 4.8. Distribution of the number of trips per trip chain for trip chains visiting PEV at least once (N = 807 trip chains)

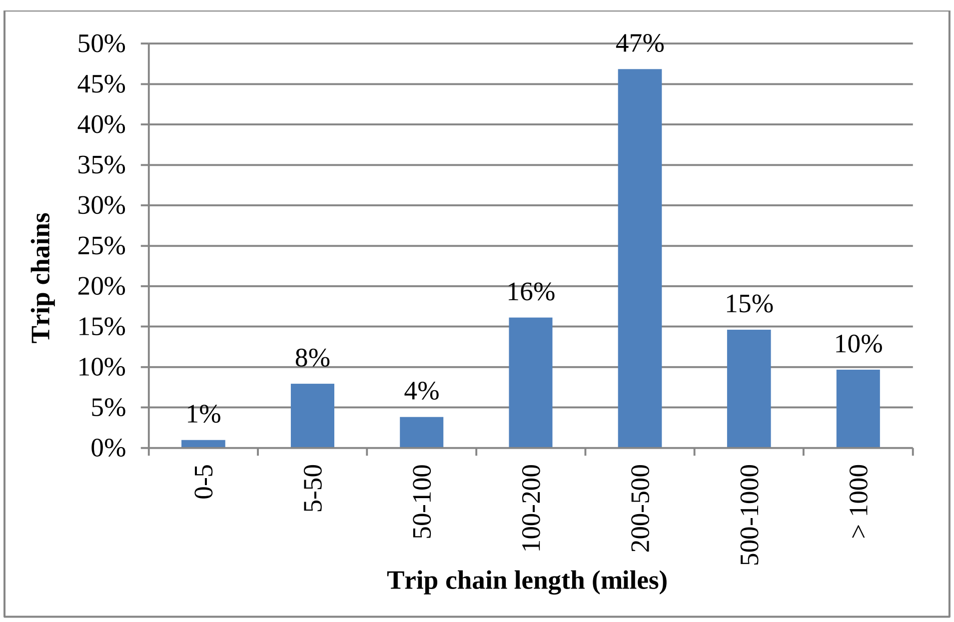


Figure 4.9. Distribution of trip-chain length for trip chains visiting PEV at least once (N = 807 trip chains)

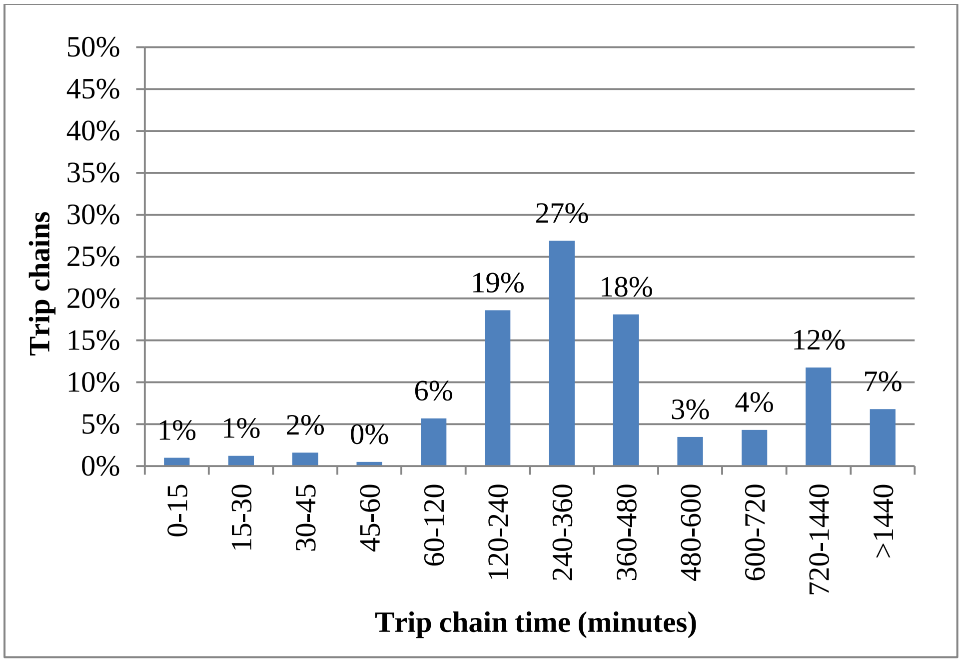


Figure 4.10. Distribution of trip-chain time for trip chains visiting PEV at least once (N = 807 trip chains)

Figures 4.11 and 4.12 respectively show the starting time and ending time profiles of 807 trip chains. Corresponding to the trip chain start time profile shown in Figure 4.11, there are multiple peaks – one for 6:00 A.M. to 9:00 A.M. corresponding to the A.M. peak, one around noon, and a third peak around the typical P.M. peak time. These multiple peaks are in agreement with the project management team's input that petroleum-tanker trucks tend to operate in different shifts at different times of the day. In both figures, there are spikes close to midnight. This is probably an artifact of the nature of the raw data, as the data streams data usually start or end around midnight (12:00 A.M.). Another interesting pattern is that the trip end profile shows a peak around 5:00 A.M., in addition to the typical P.M. peak period. Further analysis is needed to understand these temporal profiles.

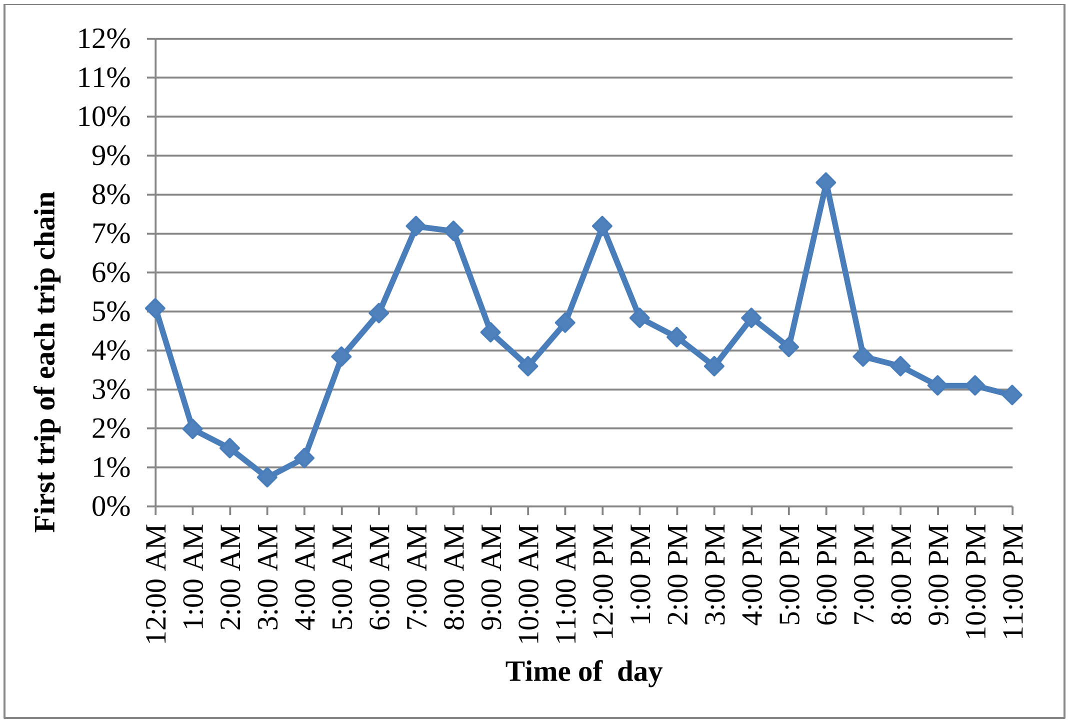


Figure 4.11. Profile of starting time of the trip chains that visit PEV at least once (N = 807 trip chains)

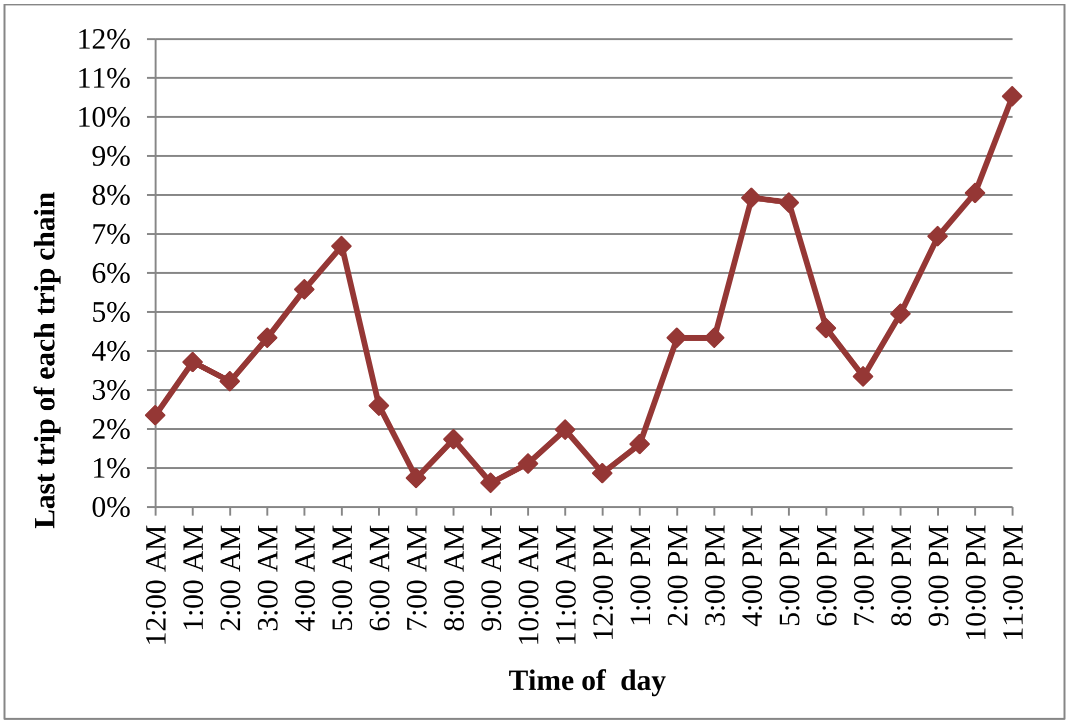


Figure 4.12. Profile of ending time of the trip chains that visit PEV at least once (N = 807 trip chains)

Table 4.2 shows the location description distribution of the 513 trip chains that did not visit PEV at all. In this table, the percentage of on-the-road trip ends is quite high for both origins and destinations. On the other hand, the proportions of PEV and gas stations are relatively low compared to those in Table 4.1. This is because many of these trip chains were originally part of a bigger chain but were cut because their trucks had crossed the study area boundary.

Table 4.2. Land-use description distribution for origins and destinations of 513 trip chains that did not visit PEV

Case	Location	Number of trip chain origins	Number of trip chain destinations
1	Gas station	181 (34.3 %) trip chain origins	46 (9.0 %) trip chain destinations
2	Other fuel recipients	28 (4.5 %) trip chain origins	14 (2.7 %) trip chain destinations
3	Distribution center	47 (9.2 %) trip chain origins	27 (4.3 %) trip chain destinations
4	On the road	174 (33.9 %) trip chain origins	388 (74.6 %) trip chain destinations
5	Rest stop	83 (16.2 %) trip chain origins	38 (7.4 %) trip chain destinations
	Total	513 (100 %) trip chain origins	513 (100 %) trip chain destinations

CHAPTER 5 : DERIVING TRIP ROUTES

5.1 Introduction

The last step in this project was to derive the travel paths (or routes) made by petroleum-tanker trucks for all trips in the trip chains discussed in Chapter 4. The procedure for deriving routes using GPS data consisted of two broad steps: (a) map-matching and (b) route generation. These two steps are explained in detail in this chapter, along with the validation results of the generated routes. The travel routes were generated for a total of 11,907 truck trips derived from ATRI data. Appendix C provides the description of the GIS shapefile of all the generated routes (the shapefile was submitted to the FDOT central office).

5.2 Map-matching

Map-matching is a technique that uses a combination of GPS location data and roadway network data to identify the correct link that has been traversed by the vehicle on the network. Map-matching is the first step towards generating the route taken by a vehicle on the network. A number of map-matching methods exist in the literature. Most of these methods are applicable for matching high-frequency GPS data (i.e., GPS data streams with small time intervals, such as at most one minute) to a roadway network. The GPS data available for this project, however, had varying frequencies ranging from a few minutes to a few hours. Therefore, this project needs an appropriate map-matching methodology for low frequency GPS data.

The map-matching algorithm used in this project is a modified version of an algorithm introduced by Yang et al. (2005) and modified by Kamali et al. (2016, in press). This methodology results in a reasonably high level of accuracy while also being easy to implement for large streams of data. It can be applied to the GPS travel data between the origin and the destination of a trip to map-match the GPS data to the underlying highway network. Kamali et al. (2016, in press) explain the details of this map-matching and route generation procedure. The core of the methodology is briefly described below.

0) Step 0. Remove GPS data within 30 feet of origin and destination points of the trip. This was done because the GPS data very close to origin or destination points might be difficult to accurately map-match to the network because the spatial error of GPS data may disrupt the correct spatial sequence of GPS points near origin and destination locations. Besides, the non-disclosure agreement signed between USF and ATRI requires the protection of spatial location data close to the origin and destination locations.
1) Step 1: Overlay the GPS data points on the highway network. Use ArcGIS Network Analysist tool to find the closest and second closest highway network links to each GPS point. Let D1 and D2 denote the distance from a GPS point to closest link and second closest link, respectively.
2) Step 2: If D1 > 500 ft, remove the GPS point. That is, GPS points that have no roadway links within their 500 ft. buffer would be removed. This step eliminates such GPS points to avoid matching them to a wrong roadway link.
3) Step 3: If $\frac{D2}{D1} > 2$, go to *"Step 4"*, or else go to *"Step 5"*.
4) Step 4: If D1 + D2 > 35 ft, match the GPS point to the closest link; otherwise, remove the GPS point. This step helps in avoiding matching GPS points to a wrong link at interchanges or near entry/exit ramps. GPS points near/at intersections or entry/exit

ramps are very difficult to map-match because of two reasons: dense links in the highway network and the GPS data's geographical inaccuracy. Since the typical GPS data accuracy is about 16.4 ft. (5 meters) according to a Department of Defense report (2001), it is reasonable to require that D1+D2 be greater than twice of GPS maximum accuracy (i.e., 35 feet).

5) Step 5: Make a 65-ft. buffer around each GPS point that did not satisfy the ratio criterion in *"Step 3"*. If there is only one intersection node falling in that buffer, then match the point to the intersection. Otherwise (i.e., if more than one intersections exist in the buffer), remove the GPS point to avoid map-matching it to a wrong intersection. This step is applicable for map-matching a GPS point that is close to an intersection. If the GPS point is near an intersection and only one intersection node falls in the 65-ft. buffer, the GPS point is matched to the intersection node. However, some intersections are coded as more than one node in the network. Consequently, two or more nodes might fall inside the 65 ft. buffer around a GPS point. Since the software cannot intelligently decide which node the GPS point should be matched to, GPS points that have two or more intersection nodes falling inside their buffers were removed.
6) Step 6: Remove any trip that has less than 5 GPS points (including origin and destination) after all the above steps. This is because for such trips with a small amount of GPS data (less than 5 GPS points) may not be sufficient information to accurately derive routes.

Using the above-described procedure, a total of 11,907 trips were successfully map-matched. Only 11 trips out of the 11,918 original trips that had less than 5 remaining GPS points were removed. The next step in deriving routes was generating the routes using map-matched GPS points, as discussed below.

5.3 Route generation

ArcMap 10.3 Network Analyst extension was employed to generate the routes for map-matched GPS points. Specifically, for each trip, the entire travel route was generated by deriving the shortest time path between the origin and destination and through the intermediate, map-matched roadway links between the origin and destination. To do so, Network Analyst utilizes a modified version of Dijkstra's algorithm to find shortest time path between two points.

The final output of route generation is a GIS shapefile with routes for 11,907 trips belonging to 1,320 trip chains made by 92 trucks. The route generated for each trip is in the form of the series of network links traversed by a truck between the origin and the destination of the trip. Along with this information, the GIS shapefile includes other attributes of the trip, including the origin and destination TAZ, County locations, trip length, and trip duration. Figure 5.1 shows an overview of the generated routes for 11,907 trips that belong to 1,320 truck trip chains in the 12-county region. The GIS shapefile of 11,907 trip routes along with its documentation was submitted to the FDOT central office. Appendix C provides a description of this GIS shapefile. Using the directions provided in this appendix, one can utilize the GIS shapefile to view (or extract) the route of a single trip, the routes of all trips in a trip chain, or the routes of all trips of all trip chains made by a truck (given by the unique truck ID).

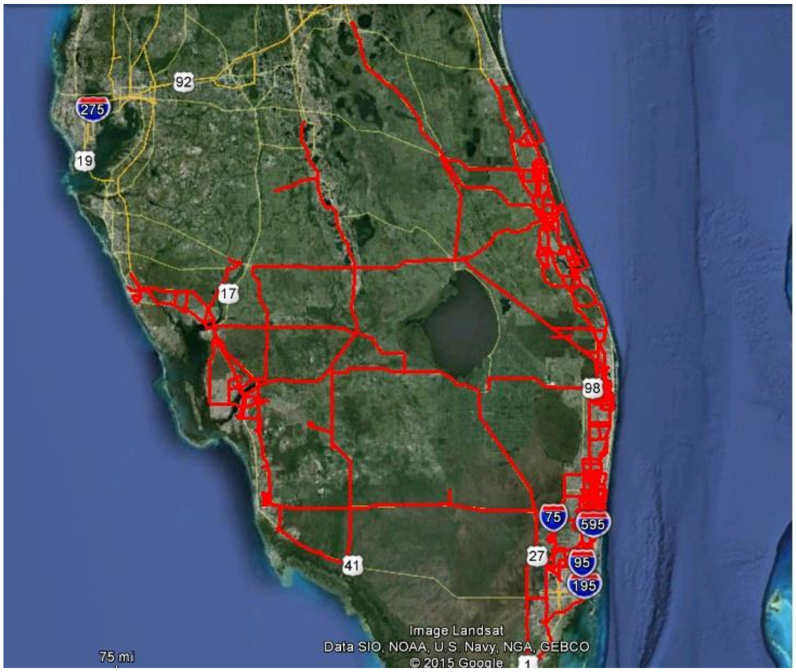


Figure 5.1. Depiction of routes generated for 11,907 petroleum-tanker-truck trips

5.4 Validation of the generated routes

The routes generated using the procedure described above were validated in terms of consistency and feasibility. A route is considered to be consistent if:

1) The direction of the travel is consistent throughout the entire route,
2) There are no loops throughout the entire route, and
3) There are no missing links in a route.

And the route is considered feasible only if:

4) There are no infeasible maneuvers throughout the entire route (e.g., making a turn on a straight bridge or an immediate change from a freeway to an arterial road).

The routes generated for 50 trips were selected and then validated using Google Earth. Figure 5.2 illustrates the trip length distribution of the 50 trips selected for validation checks. This trip length distribution shows that this 50-trip sample represents the trip lengths in the whole range of 11,907 trips available in the data.

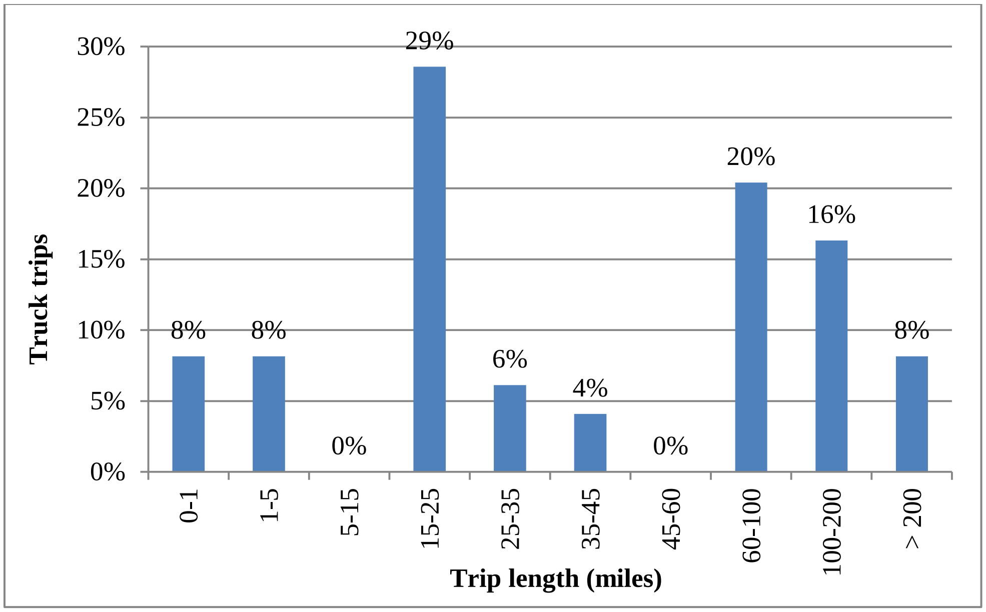


Figure 5.2. Trip length distribution of 50 trips selected for validation of generated routes

Table 5.1 presents the details of 50 validated routes. The column labeled "status" indicates if a route passed the feasibility and consistency checks. Only two out of the 50 routes were flagged "Not Ok" because of an unrealistic loop in the derived routes.

Table 5.1 Trip information for 50 trips selected for validation of generated routes

Case	Route number	Trip time (minutes)	Trip length (miles)	Status
1	2	97.3	94	Ok
2	3	63.9	62	Ok
3	4	1.2	4	Ok
4	6	0.2	1	Ok
5	7	3.0	15	Ok
6	9	0.1	1	Ok
7	10	0.8	5	Ok
8	11	79.4	86	Ok
9	12	21.5	27	Ok
10	13	102.8	120	Ok
11	14	22.2	43	Ok
12	15	24.8	56	Ok
13	16	3.0	10	Ok
14	17	79.4	89	Ok
15	18	60.8	70	Ok
16	19	138.7	154	Ok
17	20	21.7	39	Ok
18	21	21.1	42	Ok
19	23	78.9	89	Ok
20	24	21.4	26	Ok
21	26	96.1	115	Ok

Table 5.1 (continued)

Case	Route number	Trip time (minutes)	Trip length (miles)	Status
22	27	99.9	110	Ok
23	28	40.4	44	Ok
24	29	136.6	138	Ok
25	30	0.4	2	Ok
26	31	135.6	126	Ok
27	32	28.8	32	Ok
28	33	100.0	93	Ok
29	34	65.7	63	Ok
30	35	23.3	37	Ok
31	36	22.4	35	Ok
32	37	33.7	59	Ok
33	38	2.3	7	Ok
34	39	24.3	41	Not Ok
35	46	36.9	70	Ok
36	47	33.3	46	Ok
37	48	23.0	44	Ok
38	49	20.6	29	Ok
39	50	15.9	28	Ok
40	1993	212.3	256	Not Ok
41	2248	224.7	200	Ok
42	5836	176.7	155	Ok
43	6082	202.2	180	Ok
44	6091	181.9	184	Ok
45	6280	205.6	191	Ok
46	6372	177.1	156	Ok
47	10009	165	133	Ok
48	11678	183.5	283	Ok
49	12010	34	27	Ok
50	12088	25	18	Ok

Figure 5.3 illustrates an example of a route that was followed on Google Earth for feasibility and consistency checks. The image in the second row of the figure depicts the whole path of Trip #3, with its origin and destination identified. The third row illustrates the route's origin in a rest stop. Next image shows an intermediate point on the route which is located at an overpass. As can be observed, the estimated route does not show any anomaly such as a loop or infeasible maneuver at this location. This image demonstrates the accuracy of the procedure used to generate routes; particularly because the route goes through a dense network of roadway links, which can result in high chance of generating an infeasible or inconsistent route. The last image shows the destination of the route which is close to a gas station. In summary, this figure illustrates an example of a generated route that successfully passes the feasibility and consistency checks.

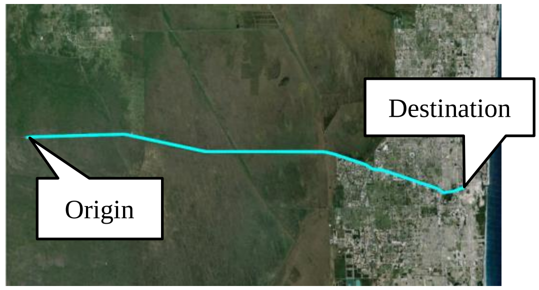
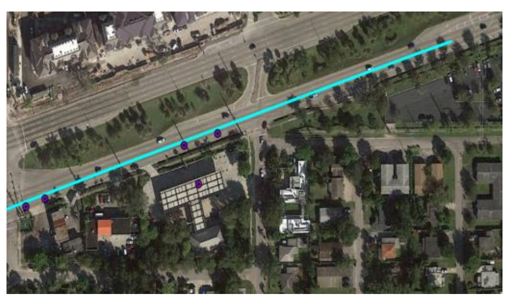
Trip number	3
The whole route	
Origin	
Overpass	
Destination	

Figure 5.3. An example of following a route for feasibility and consistency check

CHAPTER 6 : CONCLUSION

6.1 Summary

This project investigated the use of truck-GPS data to derive travel patterns of petroleum-tanker trucks delivering petroleum commodities from Port Everglades (PEV) to various fuel recipients such as gas stations in South Florida. The travel patterns of focus are the trip chaining patterns and the travel paths (i.e., routes) of tanker trucks from Port Everglades (PEV) to their final delivery points. To this end, the following tasks were pursued in the project:

1) Acquire ATRI's truck GPS data needed for the project,
2) Gather, refine, and augment the fuel recipient data needed for the project,
3) Derive a database of truck trips from ATRI's raw GPS data,
4) Derive the trip chains of trucks originating at PEV, and
5) Derive truck travel paths for the trucks traveling between PEV and gas stations.

The procedures used for and the outcomes from each of these tasks are briefly discussed next. It is worth noting here that the truck route choice data and trip chaining patterns derived in this project are utilized within a larger project funded by FHWA's SHRP 2 C20 program and conducted by FDOT's District 4 office titled "*Port Everglades Petroleum Commodity Flow Pilot Study*" aimed at exploring innovative ways to collect and analyze petroleum flow data in and out of Port Everglades.

6.2 Acquire ATRI's truck-GPS data needed for the project

For this project, the research team utilized ATRI's truck-GPS data for two months – September 2014 and March 2015 – in the following 12 Counties of South Florida: Miami-Dade, Broward, Palm Beach, Monroe, Martin, St. Lucie, Indian River, Okeechobee, Glades, Hendry, Lee, and Collier. To extract data from trucks that deliver petroleum products from PEV to the 12-county region, 14 petroleum terminals in PEV, from which the tanker trucks load petroleum products, were identified. Subsequently, ATRI extracted GPS data of all trucks in their database that were found near these terminals during September 2014 and March 2015 to capture their travel in the 12-county region (as well as a few miles beyond). This resulted in a total of 242,418 raw GPS data records for subsequent use in the project.

6.3 Gather, refine, and enrich existing fuel recipients data for the 12-county region

Two sources of data on fuel recipients for the 12-county region were used in this project – the Department of Revenue (DOR) fuel tax database and gas station location data from HERE. Neither datasets had a complete record of all fuel recipients in the 12-county region, even after combining the two datasets. In addition, both datasets had inaccuracies in the geocoding of gas stations or other fuel recipients. Incorrectly geocoded gas stations and other fuel recipients in both datasets were rectified first. Next, around 90 new gas stations were added to the combined DOR and HERE dataset. The new gas stations were found when the locations frequently visited (many of them were gas stations) by trucks in the ATRI data were examined vis-à-vis the gas station locations in the DOR and HERE datasets.

6.4 Derive a database of petroleum-tanker truck trips from ATRI's raw GPS data

The raw GPS data from ATRI was converted into a database of truck trips. To do so, the algorithm developed (for converting GPS data into trips) in a previous FDOT project by Pinjari et al. (2014) was modified for the purpose of the project. Specifically, for this project, the algorithm was modified to be able to derive local delivery trips in urban areas and to be able to derive trip chains (i.e., continuous chain of trips). The modified algorithm was used to convert two months of raw data comprising 242,418 GPS records into a total of 14,598 truck trips.

To validate and refine the trips derived from raw data, the land-use descriptions of the trip ends (origins and destinations) were identified for all 14,598 trips. To do so, all trip ends were first grouped into a smaller number of spatial "clusters" based on spatial proximity of trip ends. Subsequently, the land-use descriptions of the trip-end clusters were visualized and identified on Google Earth as well as the GIS layer of fuel recipients discussed earlier. Next, the cluster-level land-use descriptions were assigned to all the individual trip ends within the cluster. Land-use descriptions of the trip ends were used to refine the initial set of trips extracted from the ATRI data. In addition, trips ending on roadways at the study region boundaries or at the beginning or end of the data collection periods were removed. After all these procedures, a refined and validated set of 12,649 trips remained in the final trip file for subsequent use.

6.5 Derive trip chains of trucks traveling to and/or from PEV

In this task, a total of 1,320 trip chains were derived from 11,918 trips made by 92 unique trucks. 807 of these trip chains visited PEV at least once. Such trip chains mostly started or ended in either PEV or gas stations. The remaining 513 trip chains did not visit PEV at least once. Those trips were also mostly incomplete because the trucks had exited the boundaries of the study region (therefore, sufficient information was not present to derive complete trip chains in such cases). The 807 trip chains that visited PEV at least once were found to have very good connectivity and consistency within trip chains, which is paramount for deriving routes for the entire trip chain. A descriptive analysis of the trip chaining characteristics was carried out using this data.

6.6 Derive travel paths (or routes) of all trips of all trip chains going through PEV

The procedure for deriving routes using GPS data consisted of two broad steps: (a) map-matching and (b) route generation. A simple algorithm developed by Kamali et al. (2016, in press) was used to derive the routes. The travel routes were generated for a total of 11,907 trips that belonged to the 1,320 trip chains mentioned earlier. The route derived for each trip includes all the network links traversed by the truck between the origin and destination of the trip. Validation of the routes for 50 randomly selected trips indicated a high level of accuracy of the derived routes. The final product of the project was provided in form of a GIS shapefile of the generated routes for 11,907 trips along with trip-level information such as trip length, trip time, and origin/destination TAZs.

6.7 Opportunities for future research

The data developed in the project offer significant opportunities to understand petroleum-tanker truck trip chaining and route choice behavior in urban regions. While the current project focused on developing this data, future work on analyzing the determinants of truck trip chaining and route choice patterns can help in the development of petroleum-tanker truck routing policies in the south Florida region. The procedures developed in the project for building truck trip chains and travel route datasets from raw GPS data may be applied to other urban regions and for other purposes as well. For example, development of truck trip chains (or tours) for all major types of commodities transported in metropolitan regions is of interest for tour-based modeling of truck travel patterns in urban areas. The route choice data developed in the project is paramount for analysis and modeling of petroleum-tanker truck paths and flows from PEV. Further, development of route choice data for larger datasets with a large number of trips between OD pairs of interest may offer opportunities to measure and analyze the diversity of the routes chosen between OD pairs. Such analysis can inform the generation of choice sets for route choice modeling. In summary, this project resulted in procedures and data that can be used for a variety of fruitful endeavors to better understand petroleum-tanker truck travel behavior.

.

REFERENCES

Kamali, M., Ermagun, A., Viswanathan, K., Pinjari, A.R. (2016). **"**Development of Truck Route Choice Data Using Truck GPS." In *Proceedings of Transportation Research Board 95th Annual Meeting*, in press.

Pinjari, A. R., Zanjani, A. B., Thakur, A., Irmania, A. N., and Kamali, M. (2014). Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and Planning (Florida Department of Transportation Research Report BDK84-977-20). Tampa, FL: University of South Florida. http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PL/FDOT-BDK84-977-20-rpt.pdf.

Thakur, A., Zanjani, A. B., Pinjari, A. R., Short, J., Mysore, V., and S. F. Tabatabaee. (2015). Development of Algorithms to Convert Large Streams of Truck-GPS Data into Truck Trips. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2529, pp. 66-73.

U.S. Department of Defense (DOD) (2001). *Global Positioning System Standard Positioning Service Performance Standard*, Third Edition. Washington, D.C.: Government Printing Office. http://www.gps.gov/technical/ps/2001-SPS-performance-standard.pdf.

Yang, J. S., Kang, S. P., & Chon, K. S. (2005). The map matching algorithm of GPS data with relatively long polling time intervals. *Journal of the Eastern Asia Society for Transportation Studies, Vol. 6*, 2561-2573.

APPENDIX A : POLYGONS AROUND PORT EVERGLADES FUEL TERMINALS

This appendix provides the aerial images of all polygons drawn around fuel terminals at Port Everglades (PEV). The polygon boundaries are shown in red and the actual fuel terminals are shown in yellow circles. The red polygons were sent to ATRI to identify trucks that were found inside the polygons in a two-month period (September 2014 and March 2015) and to extract the GPS data of those trucks as they traveled in the South Florida region in those two months.

Figure A.1. Terminal 1 (in yellow circle) and the red polygon used for GPS data extraction

Figure A.2. Terminal 2 (in yellow circle) and the red polygon used for GPS data extraction

Figure A.3. Terminal 5 (in yellow circle) and the red polygon used for GPS data extraction

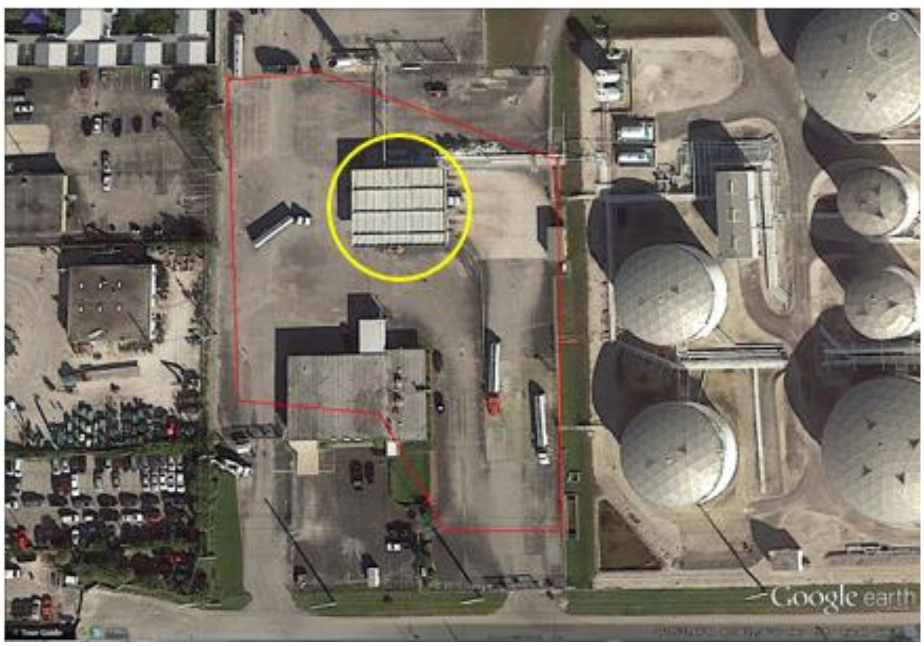


Figure A.4. Terminal 7 (in yellow circle) and the red polygon used for GPS data extraction

Figure A.5. Terminal 11 (in yellow circle) and the red polygon used for GPS data extraction

Figure A.6. Terminal 12 (in yellow circle) and the red polygon used for GPS data extraction

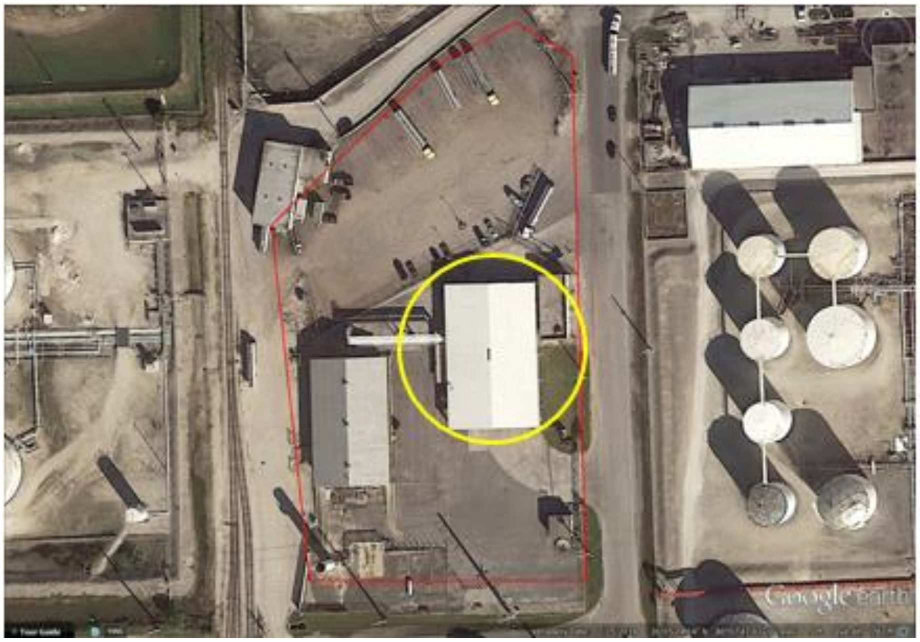


Figure A.7. Terminal 13 (in yellow circle) and the red polygon used for GPS data extraction

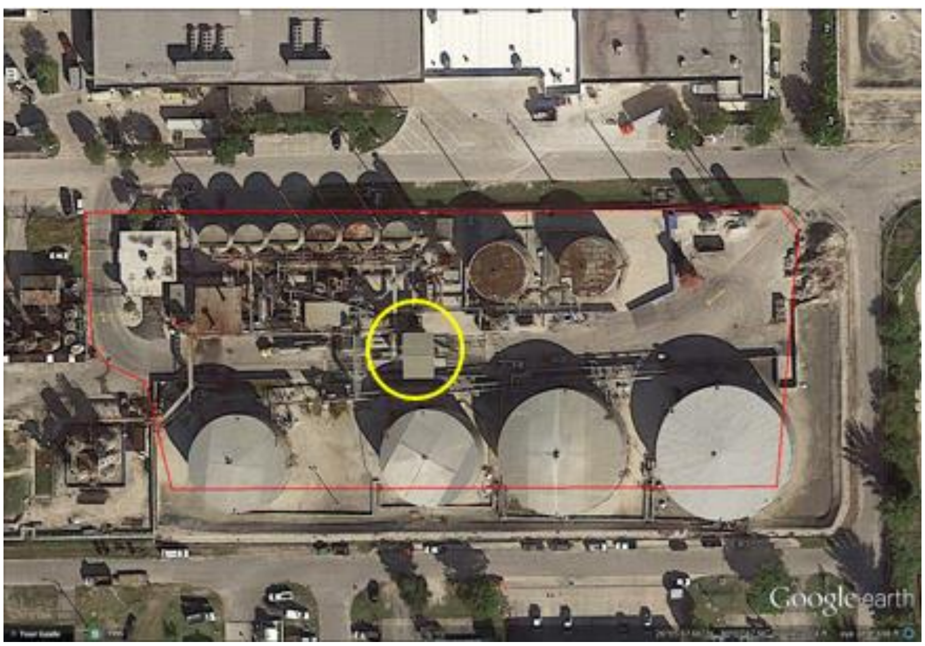


Figure A.8. Terminal 14 (in yellow circle) and the red polygon used for GPS data extraction

Figure A.9. Terminal 15 (in yellow circle) and the red polygon used for GPS data extraction

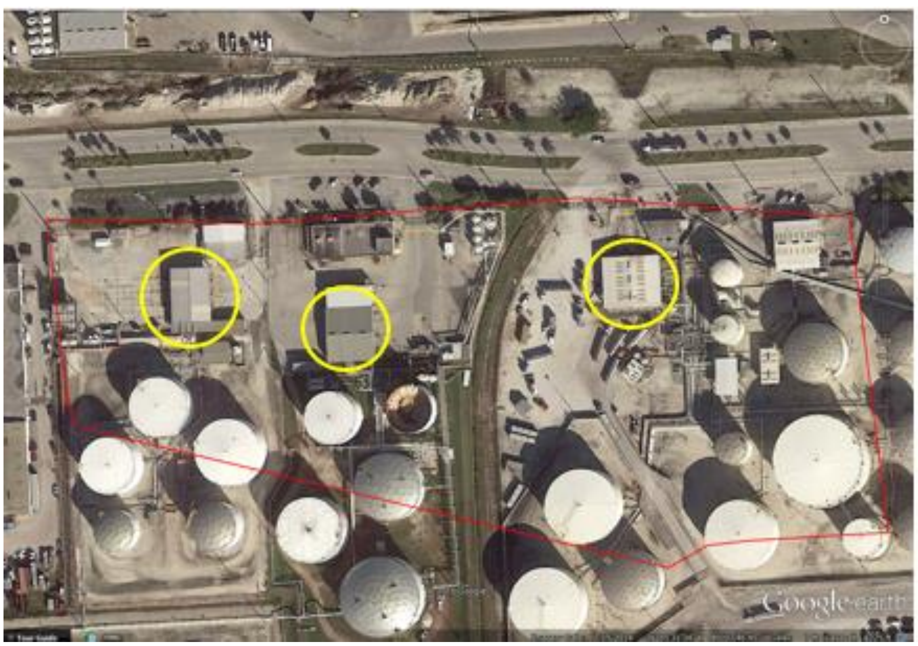


Figure A.10. Terminals 3, 4, and 10 (in yellow circle) and the red polygon used for GPS data extraction

Figure A.11. Terminals 6 and 9 (in yellow circle) and the red polygon used for GPS data extraction

APPENDIX B : DISTRIBUTIONS OF TRIP LENGTH, TRIP TIME, AND AVERAGE TRIP SPEED FOR TRIPS DERIVED FROM DIFFERENT DATASETS

This appendix provides histograms of all trip length, trip time, and trip average speed distributions for all trips extracted from two months of data. The above-mentioned distributions are provided for each data type (i.e., D1 data and D2 data) separately.

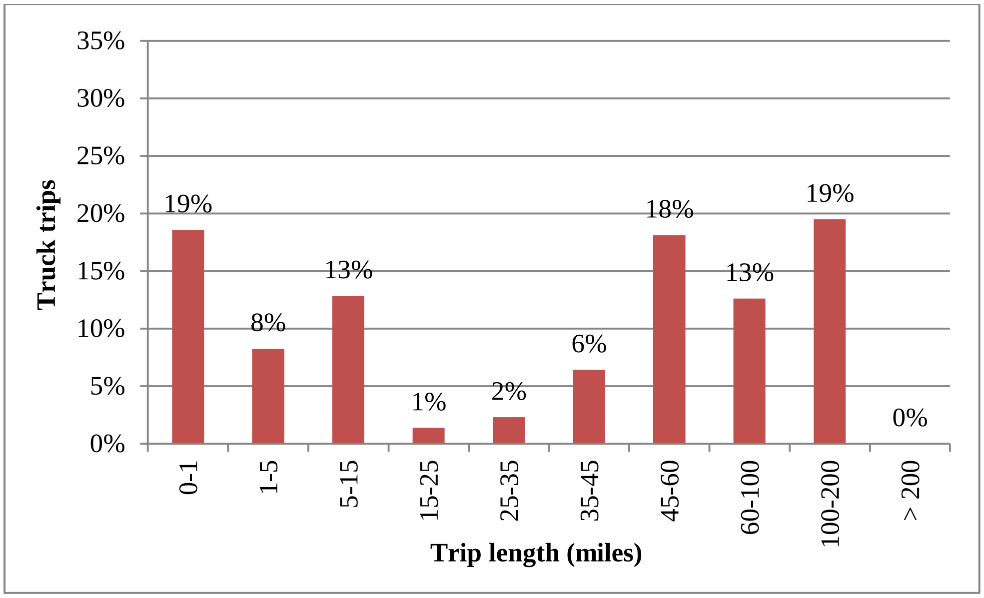


Figure B.1. Trip length distribution for all trips extracted from D1 data (N = 436 trips)

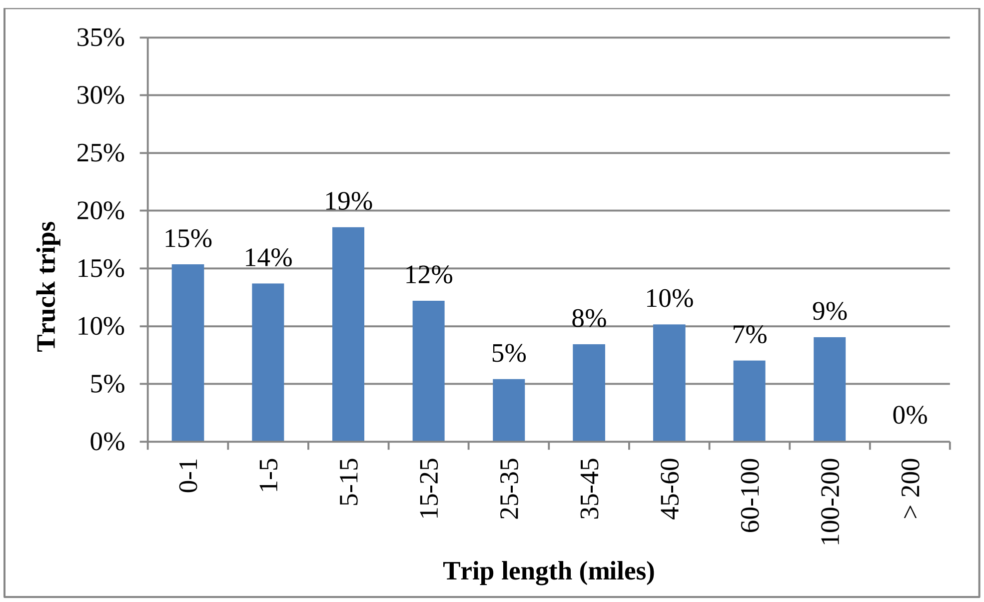


Figure B.2. Trip length distribution for all trips extracted from D2 data (N = 14,162 trips)

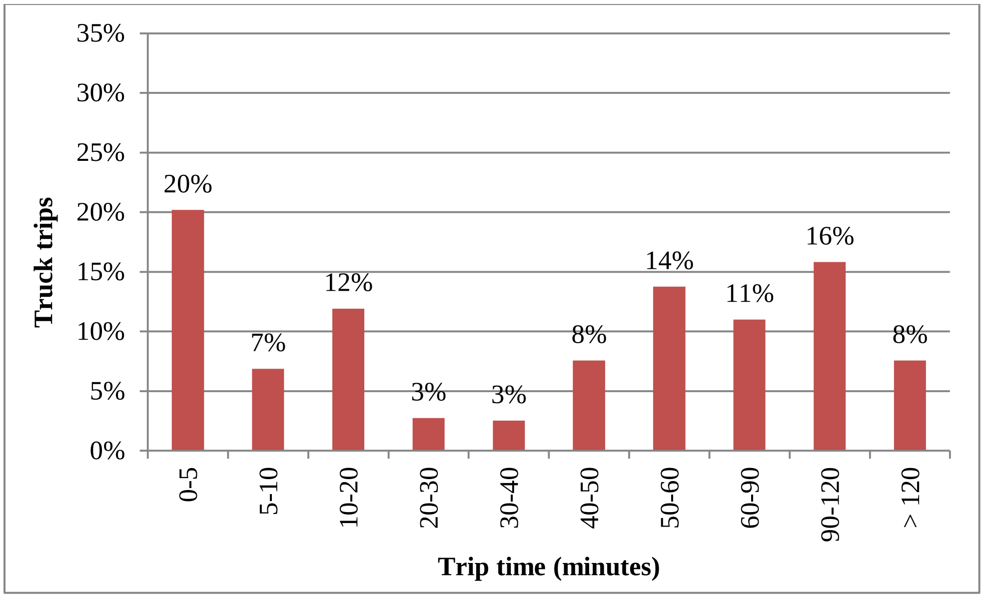


Figure B.3. Trip time distribution for all trips extracted from D1 data (N = 436 trips)

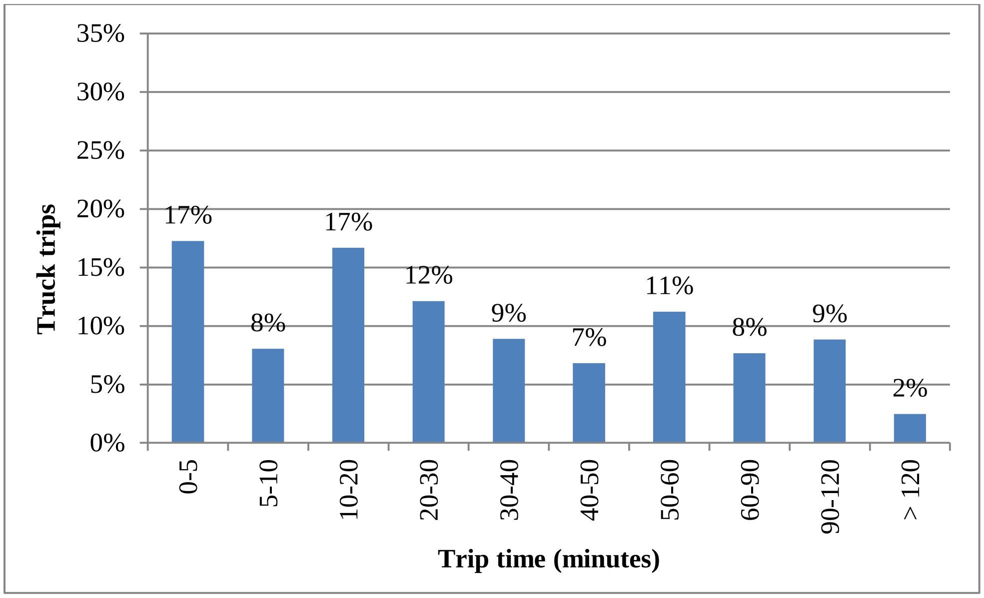


Figure B.4. Trip time distribution for all trips extracted from D2 data (N = 14,162 trips)

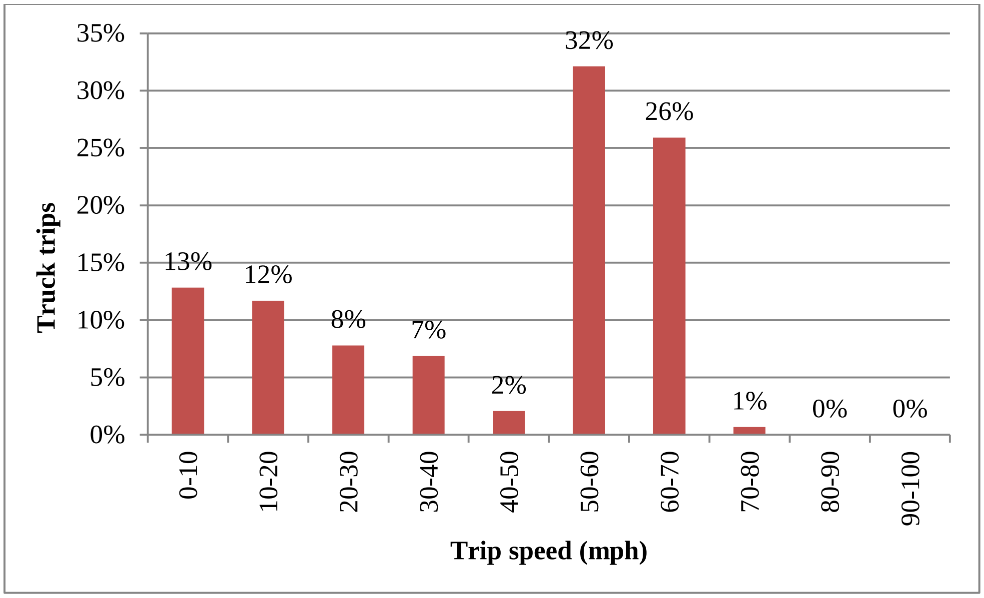


Figure B.5. Trip average speed distribution for all trips extracted from D1 data (N = 436 trips)

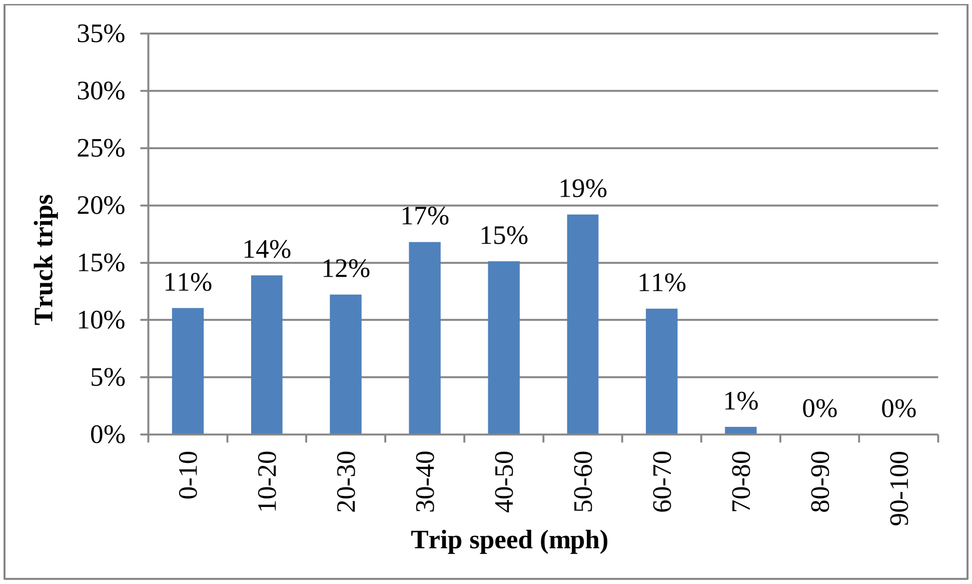


Figure B.6. Trip average speed distribution for all trips extracted from D2 data (N = 14,162 trips)

APPENDIX C : DOCUMENTATION OF THE GIS SHAPEFILE OF TRUCK ROUTES DERIVED FROM ATRI DATA

The GIS shapefile of truck routes consists of a series of network links traversed by 11,907 truck trips derived from ATRI's GPS data. The Navteq highway network was used for deriving these routes. Each record in the attribute table of this shapefile represents a network link traversed by one (or more) of the truck trips. For each of these link records, attributes containing trip-level information (e.g., trip length) have been appended. It is important to note that the routes for all trips in this shapefile belong to 1,320 trip chains. Table C.1 presents all attributes present in the shapefile's attribute table. Detailed definition of each attribute is discussed next.

Table C.1. Attribute table fields and definitions

Attribute name in the shapefile	Description
OBJECTID	(This field should be neglected)
Shape	(This field should be neglected)
Tr_ID	Truck ID number
T_ChN_Tr	Trip chain number starting from one for each truck ID
T_Num	Trip number starting from one for each trip chain
T_Chn_Num	Trip chain number starting from one and ending at 1320
Unique_trip_ID	Combination of truck ID, trip chain number, and trip number (in the mentioned order)
LINK_ID	Network link ID number
T_Start_D	Trip start date
T_Start_Tm	Trip start time
Org_TAZ	Origin TAZ
Org_Cnty	Origin county
T_End_D	Trip end date
T_End_Tm	Trip end time
Dest_TAZ	Destination TAZ
Dest_Cnty	Destination county
T_Lngth	Trip length (miles)
T_Time	Trip time (minutes)
Dest_Dw_Tm	Destination dwell time (minutes)
Org_Lctn	Origin location description
Dest_Lctn	Destination location description
end_in_PEV	Trip chain visiting PEV
cntrd_X	Longitude of link centroid
cntrd_Y	Latitude of link centroid
GPS_num	Number of map-matched GPS points for each trip
O_GPS_num	Number of original GPS points for each trip
Links_num	Number of links for each trip
Shape_Length	(This field should be neglected)

OBJECTID: This field counts the number of records in the attribute table. This field has been created by ArcGIS throughout building the final GIS file and does not provide any information on routes, links, trips, or trucks.

Shape: This field shows the value "Polyline" for all the records in the attribute table. This field has been created by ArcGIS throughout building the final GIS file and does not provide any information on routes, links, trips, or trucks.

Truck ID number (Tr_ID): Unique ID number assigned to each individual truck. To visualize the routes of all trips belonging to a specific truck ID, select all link records belonging to that truck ID and view them in a GIS layer.

Trip chain number (T_ChN_Tr): Serves as an ID number for trip chains by a truck ID, starting from one for each truck ID. This number starts from one for each truck ID and shows the chronological order of trip chains. To visualize the routes of all trips belonging to a specific trip chain of a specific truck ID, select all link records belonging to that specific truck ID and that specific trip chain number (and view them in a GIS layer).

A trip chain is a series of consecutive trips made by a truck (i.e., consecutive trips belonging to a single truck ID). Within a single trip chain, the spatial gap and temporal gap between current trip's destination and the next trip's origin are less than one mile and less than four hours, respectively.

Trip number (T_Num): Serves as an ID number for trips in a trip chain, starting from one for each trip chain. The order of the trip number in each trip chain is in the chronological order of trips in each trip chain. To visualize the route of a specific trip belonging to a specific trip chain of a specific truck ID, select all link records belonging to that specific truck ID, trip chain number, and trip number (and view them in a GIS layer).

Overall trip chain number (T_Chn_Num): An overall unique trip chain number, starting at 1 and ending at 1320. (1,320 total trip chains). This column does not identify the truck ID to which the trip chain belongs.

Unique_Trip_ID: Combination of truck ID, trip chain number (**T_ChN_Tr**), and trip number (in the mentioned order).

Link ID number (LINK_ID): Unique ID number assigned to each specific link on Navteq network. For each link in this shapefile, link information (street name, facility type, etc.) can be retrieved from the Navteq network using LINK_ID.

Trip start date (T_Start_D): Start date of the trip.

Trip start time (T_Start_Tm): Start time (time of day) of the trip.

Origin TAZ (Org_TAZ): Traffic Analysis Zone (TAZ) number in which the trip started.

Origin County (Org_Cnty): Name of the County in which the trip starts.

Trip end date (T_End_D): End date of the trip.

Trip end time (T_End_Tm): End time of the trip.

Destination TAZ (Dest_TAZ): TAZ number in which the trip ended.

Destination County (Dest_Cnty): Name of the county in which the trip ended.

Trip length (T_Lngth): Trip length in miles (obtained from the algorithm to convert raw GPS data into trips).

Trip time (T_Time): Trip duration in minutes (obtained from the algorithm to convert raw GPS data into trips), including traffic stops of duration less than 5 minutes.

Destination dwell time (Dest_Dw_Tm): Dwell time at the destination of the trip in minutes. This attribute along with "Destination location description" (Dest_Lctn) can be used to identify fuel delivery trip ends.

Origin location description (Org_Lctn): Land use description of the trip start location.

Destination location description (Dest_Lctn): Land use description of the trip end location.

end_in_PEV : Indicates whether a truck visited the PEV at least once during this trip chain. If a trip chain visits PEV at least once this field shows the value one, otherwise it shows zero.

Link centroid longitude (cntrd_X): This field shows the geographical longitude of each link centroid in decimal degrees. This number is obtained using "GCS_North_American_1983" (NAD 1983) coordinate system.

Link centroid latitude (cntrd_Y): This field shows the geographical latitude of each link centroid in decimal degrees. This number is obtained using "GCS_North_American_1983" (NAD 1983) coordinate system.

Number of map-matched GPS points in each trip (GPS_num): Each trip consists of several map-matched GPS points. This field shows the number of map-matched GPS points extracted for each trip.

Number of original GPS points in each trip (O_GPS_num): Each trip consists of several GPS points before map-matching. This field shows the number of original GPS points extracted for each trip.

Number of links in each trip (Links_num): A route on Navteq network has been generated for each truck trip. This field shows the number of network links existing in the route of each trip.

Shape_Length: This field is produced during generating the final shapefile and it shows the length of each record in meters. This field should be neglected.

How to visualize the routes of a specific truck trips or that of a trip chain?
To visualize the route of a specific truck trip, select all the link records belonging to that trip's truck ID, trip chain number, and trip number (and view them in a GIS layer). This can be achieved using ID_CHN_NUM.

To visualize the route of a specific trip chain (i.e., the route of all trips in the trip chain), select all the link records with that specific "Trip chain number" (and view them in a GIS layer). This can be achieved using T_Chn_Num.

To visualize all the trips belonging to a certain truck ID, select all the link records belonging to that specific truck ID (and view them in a GIS layer). This can be achieved using Tr_ID.