

November 2017

Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida

Trang D. Luong

University of South Florida, trangluong@mail.usf.edu

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Urban Studies and Planning Commons](#)

Scholar Commons Citation

Luong, Trang D., "Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida" (2017). *Graduate Theses and Dissertations*.

<http://scholarcommons.usf.edu/etd/7052>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida

by

Trang D. Luong

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

Major Professor: Abdul R. Pinjari, Ph.D.
Robert Bertini, Ph.D.
Pei-Sung Lin, Ph.D.
Seckin Ozkul, Ph.D.

Date of Approval:
November 1, 2017

Keywords: truck-GPS data, route diversity, route overlap, route dominance, network structure

Copyright © 2017, Trang D. Luong

DEDICATION

To the students who have acquired sincere interests in transportation modeling

ACKNOWLEDGEMENTS

My sincere gratefulness is due to major advisor, Dr. Abdul R. Pinjari, for his attentive guidance and thoughtful discussions about graduate research and coursework. His meticulous dedication to research has taught me the importance of in-depth studies. His wholehearted support signifies beyond fully funding my Master's studies: the respect for my diverse involvement at USF and efforts of guiding my professional growth.

My gracious thanks are to Divyakant Tahlyan for his rigorous commitment to research on transportation modeling, and Mohammadreza Kamali for his initial work on deriving routes and counting the number of unique routes.

My hearty appreciation is extended to Drs. Robert Bertini, Pei-Sung Lin, Seckin Ozkul, for participating in my thesis committee, reinforcing the value of research, and offering me full concentration on completing this work.

The data used in this research was obtained from projects funded by the Florida Department of Transportation (FDOT) and the United States Department of Transportation (USDOT). This study partially contributes to the report of the research project funded by the USDOT National University Transportation Center (UTC) Consortium led by the Center for Advanced Infrastructure and Transportation (CAIT), Rutgers University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National University Transportation Center, USDOT, or FDOT.

TABLE OF CONTENTS

LIST OF TABLES	iii
LIST OF FIGURES	iv
ABSTRACT.....	v
CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 Literature Review.....	2
1.3 Research Objectives.....	3
1.4 Thesis Organization	4
CHAPTER 2: DATASETS.....	5
2.1 Raw Truck GPS Data.....	5
2.2 Deriving Truck Routes.....	5
2.2.1 Converting Truck-GPS Data to Truck Trips.....	5
2.2.2 Deriving Truck Routes from Truck Trips	6
2.2.2.1 Improving Map-matching Procedure for Short-haul Routes	7
2.3 Database of Truck Routes for Diversity Analysis	9
CHAPTER 3: DEVELOPMENT OF ROUTE DIVERSITY METRICS.....	15
3.1 Introduction.....	15
3.2 Number of Unique Routes	15
3.3 Route Variability Metrics	18
3.3.1 Average Commonality Factor.....	18
3.3.2 Average Path Size	18
3.3.3 Non-overlapping Index	19
3.3.4 Illustrations of Route Variability Metrics	19
3.4 Route Dominance Metrics	21
3.4.1 Standardized Variance of Route Usage	21
3.4.2 Standardized Shannon Entropy of Route Usage.....	22
3.4.3 Illustrations of Route Dominance Metrics.....	23
3.5 Illustration.....	24
CHAPTER 4: MODELING METHODOLOGY.....	26
4.1 Introduction.....	26
4.2 Count Data Models for Number of Observed Unique Routes.....	26
4.3 Fractional Response Models for Average Path Size and Standardized Shannon Entropy of Route Usage.....	27

CHAPTER 5: DESCRIPTIVE ANALYSIS	28
5.1 Introduction.....	28
5.2 Diversity Metrics	28
5.3 Potential Determinants of Diversity.....	32
5.3.1 Trip Characteristics	33
5.3.2 OD Location Characteristics.....	33
5.3.3 Network Structure.....	33
CHAPTER 6: ESTIMATION RESULTS	35
6.1 Introduction.....	35
6.2 NB Regression Model for Number of Unique Routes.....	35
6.3 Fractional Response Models for Average Path Size.....	39
6.4 Fractional Response Models for Standardized Shannon Entropy of Usage	40
CHAPTER 7: CONCLUSIONS AND FUTURE RESEARCH.....	41
7.1 Summary of Research.....	41
7.2 Findings and Conclusions of Research.....	42
7.3 Opportunities for Future Research.....	44
REFERENCES	45

LIST OF TABLES

Table 2.1 Dataset Summary	11
Table 5.1 Descriptive Statistics of Diversity Metrics	28
Table 5.2 Descriptive Statistics of Explanatory Variables for Route Diversity Analysis	32
Table 6.1 Estimation Results of Truncated Negative Binomial Regression of Number of Unique Routes	37
Table 6.2 Estimation Results of Fractional Response Models for Average Path Size	39
Table 6.3 Estimation Results of Fractional Response Models for Standardized Shannon Entropy of Usage	40

LIST OF FIGURES

Figure 2.1 Geographical Distribution of all Long-haul Truck Routes.....	10
Figure 2.2 Geographical Distribution of all Short-haul Truck Routes	10
Figure 2.3 Trip Length Distributions of Long-haul and Short-haul Trips.....	12
Figure 2.4 Trip Time Distributions of Long-haul and Short-haul Trips.....	12
Figure 2.5 Florida Weigh Stations Map.....	13
Figure 2.6 Florida Seaport System.....	14
Figure 3.1 Number of Unique Routes Distribution for Long-haul and Short-haul Datasets	16
Figure 3.2 Scatter Plots of the Number of Unique Routes versus the a) Number of Trips, b) Number of Truck IDs, and c) Direct OD Distance.....	17
Figure 3.3 Correlations of a) Average Commonality Factor, b) Average Path Size, c) Non- overlapping Index with the Number of Unique Routes	20
Figure 3.4 Correlation of a) Standardized Variance of Route Usage and b) Standardized Shannon Entropy of Route Usage with the Number of Unique Routes	23
Figure 3.5 Correlation between Two Dominance Metrics.....	24
Figure 3.6 Examples of Unique Routes for a Long-haul OD Pair and a Short-haul OD Pair	25
Figure 5.1 Number of Unique Routes Distribution	29
Figure 5.2 Average Commonality Factor Distribution.....	30
Figure 5.3 Average Path Size Distribution	30
Figure 5.4 Non-overlapping Index Distribution	30
Figure 5.5 Standardized Variance of Usage Distribution.....	31
Figure 5.6 Standardized Shannon Entropy of Usage Distribution.....	31
Figure 5.7 Long Ellipse, Short Ellipse, and Circular Buffers.....	34

ABSTRACT

This thesis presents a comprehensive exploratory analysis of truck route choice diversity in the state of Florida, for both long-haul and short-haul truck travel segments. We employ six metrics to measure three different dimensions of diversity in truck route choice between any given origin-destination (OD) pair. These dimensions are: (1) number of distinct routes used to travel between the OD pair, (2) the extent of overlap (or lack thereof) among the routes, and (3) the evenness (or the dominance) of the usage of different unique routes. The diversity metrics were utilized to examine truck route choice diversity from over 73,000 truck trips that were derived from over 200 million GPS records of a large truck fleet. Descriptive analysis and statistical modeling of the diversity metrics offered insights on the determinants of various dimensions of truck route choice diversity between an OD pair. The results could be used to improve choice set generation algorithms for truck route choice modeling as well as in planning truck route policies and investments.

CHAPTER 1: INTRODUCTION

1.1 Background

Highway freight mobility is critical to a region's economic growth. An essential step toward enhancing highway freight mobility is to improve our understanding of freight movement. Analyzing truck routes can help design short-term truck routing policies aimed at mitigating congestion and improving travel time reliability. Due to limited data on truck movements, however, truck route choice has been an understudied dimension of freight movement. The recent availability of global positioning systems (GPS) data has started to fill this gap. A few studies have used GPS data to understand route choice behaviors of freight trucks or to derive freight performance measures (Brown and Racca, 2012; Liao, 2014; Wang et al., 2016; Woodard et al., 2017). However, not much attention has been paid to gain a better understanding of the truck travel patterns, particularly on the degree of diversity among truck routes.

Analyses of the degree of truck route diversity have important applications in both planning and modeling practices. Transportation planners can use route variability measures to evaluate the performance of transportation networks, particularly for the demand of designated truck routes or toll roads. More diversity in the chosen routes implies higher resiliency of the roadway system, especially for routine infrastructure maintenance and rebuilding efforts for emergency recovery. Identifying OD pairs with higher diversity can help trucking companies in optimizing routing strategies. For modeling applications, route diversity measures, such as the number of expected routes and their amount of route overlap, can be useful criteria to guide route choice set generation algorithms, and subsequently improve the generated routes.

1.2 Literature Review

The availability of GPS-data in recent years has offered three major venues for route choice research, such as building route choice models, examining the choice set algorithms, and analyzing behaviors of the chosen routes. Route choice research has utilized the recently available GPS-data to build route choice models, examine the choice set algorithms, and analyze behaviors of the chosen routes. Several advanced route choice models have utilized this rich GPS data collected to analyze routes taken by automobiles (Levinson and Zhu, 2013). There are still limited route choice models build for truck mode. Hess et al., 2015 is the only recent study that explored a large set of truck-GPS data collected in a dense urbanized network. Generating a sensible and exhaustive route choice set is the focus of not only the algorithm itself but also on the robustness of the route choice model estimates (e.g.: Bekhor et al., 2006; Bliemer and Bovy, 2008). Bovy, 2009 suggested four aspects of the choice set: (1) the sufficient number of route alternatives, (2) the inclusion of observed routes, (3) the plausibility of the route hierarchical sequence, and the (4) diversity of the generated routes. Those criteria need to be determined by a model that can estimate the reasonable number of diverse alternatives for any given OD pair.

In addition to developing route choice models, researchers have deployed GPS data to analyze the route deviation or the variability of observed routes. Jan et al., 2000 concluded that most chosen routes differed from the shortest time path by comparing matches of few similar OD pairs. Papinski et al. 2009 estimated that up to 20% of travelers deviated from their planned route. Papinski and Scott, 2009 found that the observed routes are longer than both shortest time and shortest distance routes because travelers also consider many other route characteristics. Spissu et al., 2011 quantified the time and distance that the observed routes deviated from minimum-cost routes, while examining the variability of daily routes chosen by same or different individuals.

Zhu and Levinson's empirical test of the user-optimal equilibrium principle in trip assignment using all chosen routes is the latest study on the extent of route variability. Among those few studies, the datasets were limited by the size of only tens or hundreds of observed routes. Only a few indices were applied to quantify the extent of route deviation or route variability, such as path/time/distance deviation index and overlapping index (Jan et al., 2000; Spissu et al., 2011).

1.3 Research Objectives

Based on the available datasets of truck routes derived for two FDOT projects, the objective of this thesis is to conduct a comprehensive exploratory analysis of truck route choice diversity in Florida for both long-haul and short-haul travel segments. The term diversity is used to characterize the differences in observed routes used by trucks along three different dimensions: (1) the number of different routes used by trucks for travel between an OD pair, (2) the amount of overlap/similarity among different routes used between an OD pair, and (3) the evenness (or, otherwise, the dominance) in usage of different routes. Specifically, the study addresses two broad questions: (1) How to measure the degree of diversity in the routes trucks use to travel between an OD pair? (2) What factors influence the diversity of truck route choice between an OD pair? To this end, six metrics were used to measure the following three different dimensions of diversity in route choice between a given OD pair: (1) number of different routes used between the OD pair, (2) extent of overlap (or lack thereof) among the routes, and (3) evenness (or the dominance) of the use of different unique routes between that OD pair. These metrics were applied to quantify truck route choice diversity using large streams of more than 200 million GPS records. Next, statistical models were estimated to explore the influence of various determinants on the three dimensions of route choice diversity between different OD pairs. The models provided insights

into the influence of truck travel characteristics, OD location characteristics, and network structure characteristics between an OD pair on the diversity of route choice between that OD pair.

1.4 Thesis Organization

Chapter 2 describes the truck-GPS data and the derived route datasets used in this thesis. Chapter 3 describes the metrics used to quantify diversity in truck route choice. The statistical models used in this thesis and empirical results are presented in Chapter 4. Chapter 5 documents descriptive analysis of diversity metrics and explanatory variables. Chapter 6 summarizes empirical findings from statistical models on the determinants of diversity. Chapter 7 concludes this thesis with findings and recommendations for future research.

CHAPTER 2: DATASETS

2.1 Raw Truck GPS Data

The truck-GPS data used in this thesis was obtained from the American Transportation Research Institute (ATRI) for two Florida Department of Transportation (FDOT) funded projects (Pinjari et al., 2014; Tahlyan et al., 2017). The data used to derive long-haul truck trips (trips longer than 50 miles) comprised more than 145 million GPS records corresponding to a fleet of nearly 50,000 freight trucks. The long-haul GPS data spanned spatially over the state of Florida and temporally over a four-month period (March–June 2010). The data used to derive short-haul trips (trips shorter than 50 miles) comprised more than 96 million GPS records corresponding to a fleet of nearly 110,000 freight trucks and spanned six counties of the Tampa Bay region in Florida. Temporally, the short-haul data corresponded to first 15 days in October 2015, December 2015, April 2016, and June 2016.

2.2 Deriving Truck Routes

2.2.1 Converting Truck-GPS Data to Truck Trips

The raw truck-GPS data were first converted into a database of truck-trips using algorithms developed by Thakur et al. (2015) and later refined by Pinjari et al. (2015) for the same data. The overall procedure to convert ATRI's truck-GPS data into a database of truck trips is summarized in the following three broad steps:

1. Clean, read, and sort raw GPS data in chronological order for each truck ID. At the end of this step, all GPS data belonging to each truck ID are grouped together in 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, is considered part of the travel between origin and destination.
3. Conduct quality checks and refine or eliminate trips that do not satisfy quality criteria.

2.2.2 Deriving Truck Routes from Truck Trips

To derive the chosen route for each long-haul trip, raw GPS records corresponding to each trip were map-matched using the procedure developed by Kamali et al. (2016) to high-resolution NAVTEQ roadway networks provided by FDOT. The long-haul truck routes were derived by Kamali et al. (2016). The 2010 NAVTEQ network used to derive long-haul routes comprised more than 1.5 million links and 5.8 million nodes. To match the time frame when the GPS data was collected, the 2015 NAVTEQ network used for short-haul routes comprised over 1.8 million links and more than 6.9 million nodes. The short-haul truck routes were derived by the author and her colleagues as one of the main tasks in the report “Development and Analysis of Truck Route Choice Data for the Tampa Bay Region using GPS Data” (Tahlyan et al., 2017). The procedure for deriving routes using GPS data consisted of three broad steps: 1) map-matching data set preparation, 2) map-matching process, and 3) route generation. Section 2.2.2.1 documents the improvements of the first two steps of the map-matching procedure developed by Kamali et al. (2016) to better suit the nature of short-haul truck trips.

2.2.2.1 Improving Map-matching Procedure for Short-haul Routes

1. Map-matching data set preparation: The raw GPS data of each trip were refined based on the following criteria: a) Removing GPS coordinates within a 1,000-ft radius of the origin or destination of each trip. For trips that started or ended in urban areas with a high density of highway network links, it was not easy to accurately map-match the raw GPS data because of the absence of many minor roadway network links in the Navteq network used in the study. Mismatching these GPS points would lead to loops (circuitous maneuvers) in the generated routes. As this step also removes the origin and destination GPS points, these points were later added back to the set of GPS points corresponding to a trip. b) Space-sampling the remaining GPS coordinates to be at least 1,000 ft apart. This was done to eliminate GPS points that were too close to each other and did not help enhance the accuracy of matching the points to the road network. The space-sampling approach is different from the time-sampling approach mentioned in Kamali et al. (2016), but space sampling helps to keep consistency across the spatial distribution of the consecutive GPS points. In the time-sampling approach, GPS points can still be very close to each other after a period of time (e.g., 1 or 2 minutes when a truck stops at a traffic light, etc.). c) Based on the remaining GPS points in each trip, removing trips with less than 5 GPS points, as the number of GPS points in those trips was considered below the number needed to accurately derive the travel route.

2. Map-matching process: 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. The author modified the map-matching procedure used in Kamali et al. (2016), which was originally proposed by Yang et al. (2005).

First, all the GPS points that were within 500-ft buffer around the highway interchanges were removed. Points close to highway interchanges are difficult to map-match and can lead to major detours from the actual route. This was done by drawing a 500-ft buffer around the highway interchanges; these interchanges were identified using the attributes present in the Navteq network that indicated if a network link belongs to a highway interchange. The GPS points falling within the 500-ft buffer were identified by intersecting the GPS data layer with the 500-ft buffer layer around the highway interchanges.

Second, the distance of each GPS point from the first nearest link in the network was calculated. This distance was denoted as D1, and all GPS points where $D1 > 500$ ft were removed. Subsequently, distance of each GPS point from the second nearest link in the network was calculated and denoted as D2. This was done using the “Generate Near Distance Table” tool in the ArcGIS environment.

Third, the angle between the geographic north and a perpendicular line drawn from each GPS point to the first- and second-nearest links in the network was calculated. If the location of the link was east of the location of the GPS point, the angle was measured in the clockwise direction; if the location of the link was west of the GPS point, the angle was measured in the anti-clockwise direction. These angles were denoted as A1 and A2. If $170^\circ < (A1 + A2) < 190^\circ$, the GPS point was supposedly between two parallel roads and was difficult to be map-matched accurately. All such GPS points were removed from the dataset. Of the points that were removed, those that fell within the 65-ft buffer of just one roadway intersection were retained and were map-matched to the intersections instead of the links.

Finally, all trips with fewer than five GPS points were removed, as these trips did not have enough GPS points for accurate route determination.

2.3 Database of Truck Routes for Diversity Analysis

To analyze route choice and the diversity therein, it is useful to aggregate trip end locations to larger spatial units to observe a sufficient number of trips to get an uncensored view of the various routes trucks choose between two locations. Even if a substantial number of trips was observed between disaggregate OD locations, it might not exhibit the complete alternatives in route choice due to lack of diversity in the truck drivers, operators and/or nature of the trucking businesses. Furthermore, practical implementations of route choice analysis and modeling consider spatially-aggregated units such as traffic analysis zones (TAZs). There are 5,403 TAZs in the Florida statewide travel demand model (FLSWM). The size of each TAZs varies based on each region's traffic, population, and employment densities. Therefore, in this thesis, all trip end locations were aggregated to the TAZs defined in Florida's statewide travel demand model.

Based on empirical observations, TAZ OD pairs that had at least 50 trips for long-haul data and at least 30 trips for short-haul data were selected, as OD pairs with fewer trips might not offer a complete picture of truck route diversity. Table 2.1 summarizes the attributes and selection criteria for both long-haul and short-haul truck trips. All subsequent analyses were built parallel for both datasets to compare and contrast the degree of route diversity for long-haul and short-haul trips. From a set of more than 78,000 routes, the final long-haul dataset used in this analysis comprised 277 TAZ OD pairs with a total of 30,263 routes that were longer than 50 miles as shown in Figure 2.1. Thakur et al., 2015 estimated that ATRI GPS data only cover about 10% of truck traffic volumes in Florida. Figure 2.2 shows the geographical distribution for 527 short-haul OD pairs comprising 42,884 routes refined from a set of more than 230,000 routes. As trips shorter than 5 miles would not have many route choice options for truck travel, all short-haul routes are between 5 and 50 miles long.

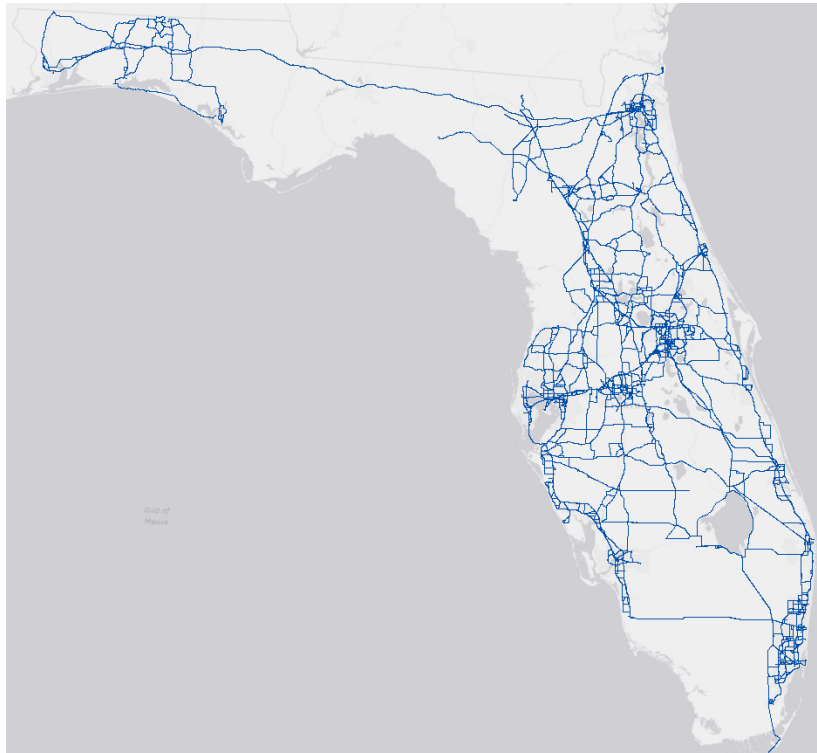


Figure 2.1 Geographical Distribution of all Long-haul Truck Routes

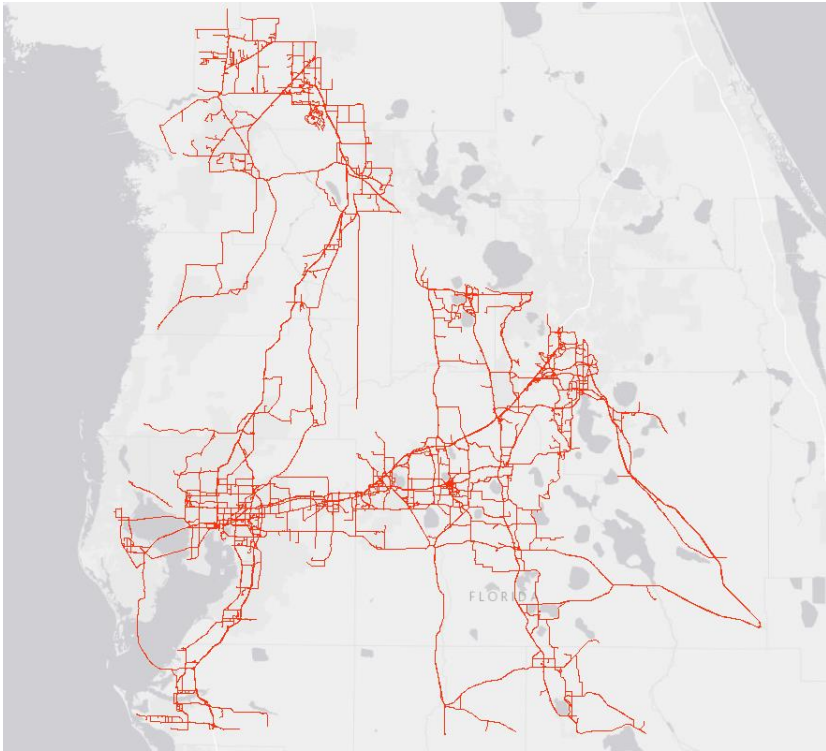


Figure 2.2 Geographical Distribution of all Short-haul Truck Routes

Table 2.1 Dataset Summary

Attributes	Long-haul	Short-haul
<i>Original datasets</i>		
Collected period	March, April, May and June 2010	October, December 2015 and April, June 2016
Coverage region	Entire State of Florida	Six counties in the Tampa Bay (mid-west of Florida)
Number of raw GPS records	96 million	145 million
Number of derived routes	78,381	233,329
<i>Selection criteria for route diversity analysis</i>		
Minimum trip length	50 miles	5 miles
Minimum number of trips per OD pair	50	30
<i>Refined datasets for route diversity analysis</i>		
Number of truck routes	30,263	42,884
Number of OD pairs	277	527

As all selected truck trips have their corresponding traversed routes along the road network, the terms “trip” and “route” are used interchangeably. Figure 2.3 presents the trip length distribution of the trips selected for analyzing truck route diversity. More than half of long-haul trips are between 50 and 100 miles long, with the maximum trip length of 470 miles. The trip time distribution shown in Figure 2.4 indicates that the average truck speed is less than 60 miles per hour, especially for long-haul trips.

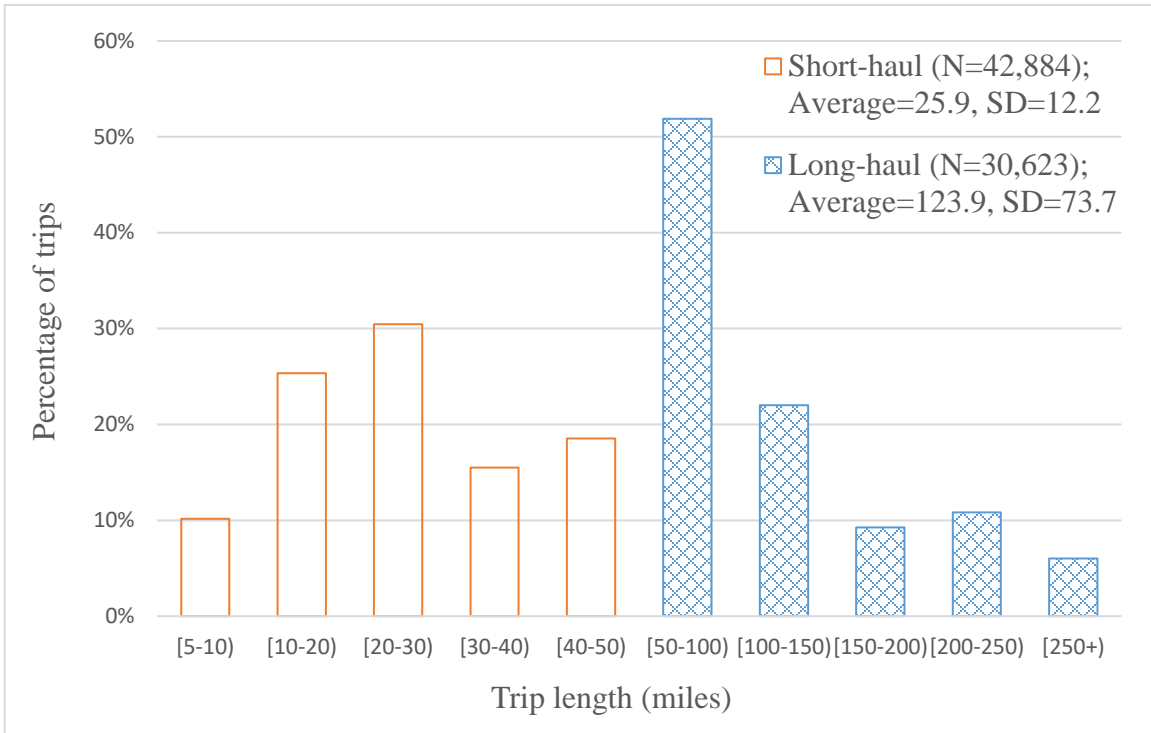


Figure 2.3 Trip Length Distributions of Long-haul and Short-haul Trips

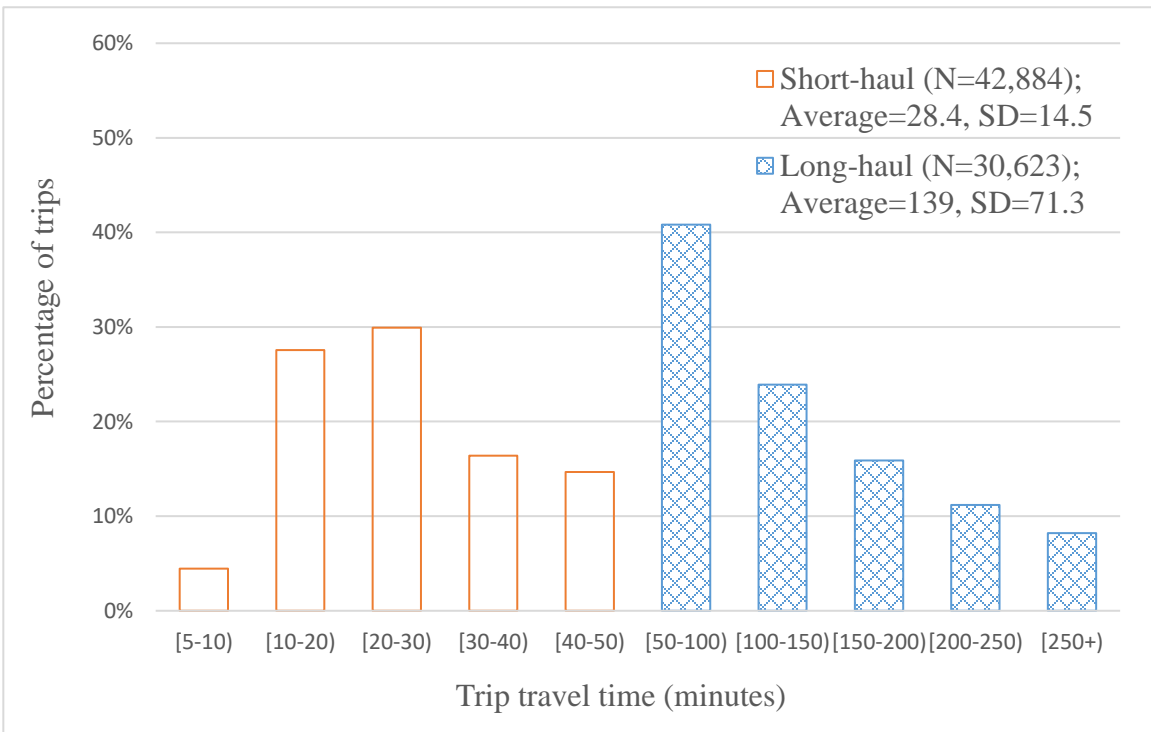


Figure 2.4 Trip Time Distributions of Long-haul and Short-haul Trips

Figure 2.5 presents the locations of all 20 active weigh stations that are distributed along or near the Interstate system. The long-haul routes shown in Figure 2.1 pass through 18 out of 20 weigh stations. The White Springs Weigh-In-Motion Station may have trips going out of / into Florida while this analysis only considers routes trips within Florida. Truck trips going through the Old Town Static Station may have not been a part of an OD pair with at least 50 trips. Three weigh stations—Wildwood, Seffner and Hopewell are within the study boundaries for the short-haul dataset.

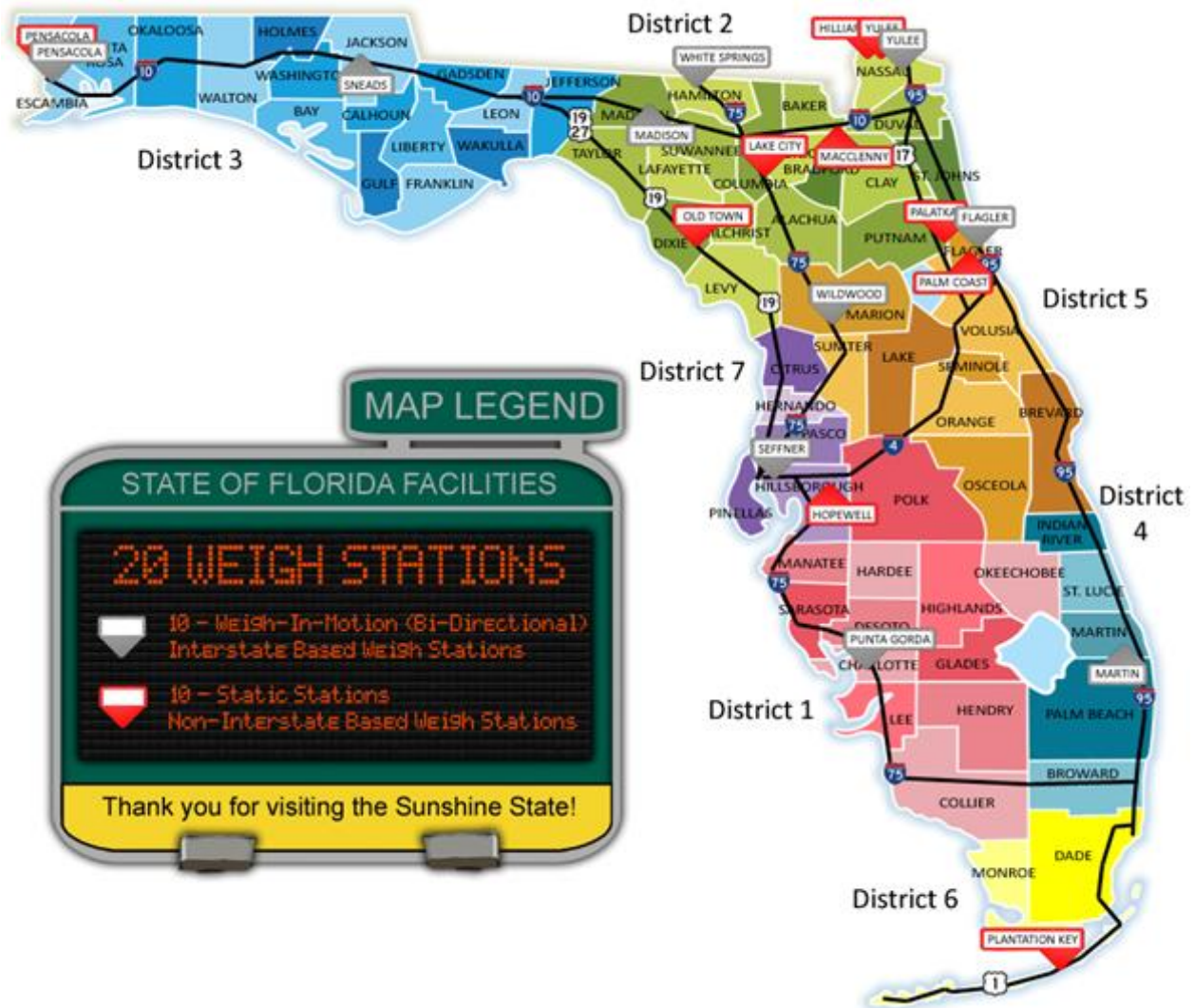


Figure 2.5 Florida Weigh Stations Map

Figure 2.6 presents the locations of ten cargo seaports that were visited by long-haul routes. Detailed analysis of the route distributions confirms that the Port Tampa Bay and Port Manatee are the two popular origins and destinations for short-haul truck trips. It can be observed that freight truck is the primary mode to transport goods into and out of seaports.



Figure 2.6 Florida Seaport System

CHAPTER 3: DEVELOPMENT OF ROUTE DIVERSITY METRICS

3.1 Introduction

To measure diversity in truck route choice between a given OD pair, the following six metrics were employed: (1) number of unique routes, (2) average commonality factor, (3) average path size, (4) non-overlapping index, (5) standardized variance of route usage, and (6) standardized Shannon entropy of route usage. The first metric measures the number of unique routes traveled by trucks between an OD pair. The next three metrics measure the extent of overlap (or lack thereof) among the observed unique routes, which are referred to as the route variability metrics. The last two metrics measure the evenness (or, otherwise, dominance) in the usage of the routes between the OD pair, which are referred to as the route dominance metrics. These three dimensions together provide a complete picture of the diversity in truck route choice between an OD pair.

3.2 Number of Unique Routes

Many routes traveled between an OD pair are different by only a few links. To determine a set of distinct or unique routes traveled between an OD pair, we used the commonality factor proposed by Cascetta et al. (1996). Commonality factor (C_{ij}) between routes i and j is defined as:

$$C_{ij} = l_{ij} / \sqrt{L_i L_j} \quad (3.1)$$

where L_i and L_j represent the length of routes i and j , respectively, and l_{ij} is the length of the shared portion between the two routes. The two routes are referred to as unique from each other if the commonality factor between the two routes is below 0.95. To determine the number of unique routes observed between an OD pair, all routes between that OD pair are arranged in an ascending order of route length. The shortest route is the first unique route. The commonality factor of each

subsequent route is computed with respect to all previous unique routes to determine if it is a unique route (if C_{ij} is less than 0.95). The result of this process is a set of unique routes between an OD pair, where the commonality factor between any two unique routes is less than 0.95. The size of this unique route set represents the number of unique routes used between that OD pair. As a result, the initial large set of routes derived for all trips is reduced to a smaller set of unique routes per OD pair as shown in Figure 3.1. It is important to note that even for a large number of unique routes between an OD pair, there still may not be many competitive alternative routes since each unique route can overlap up to 95% with any other routes.

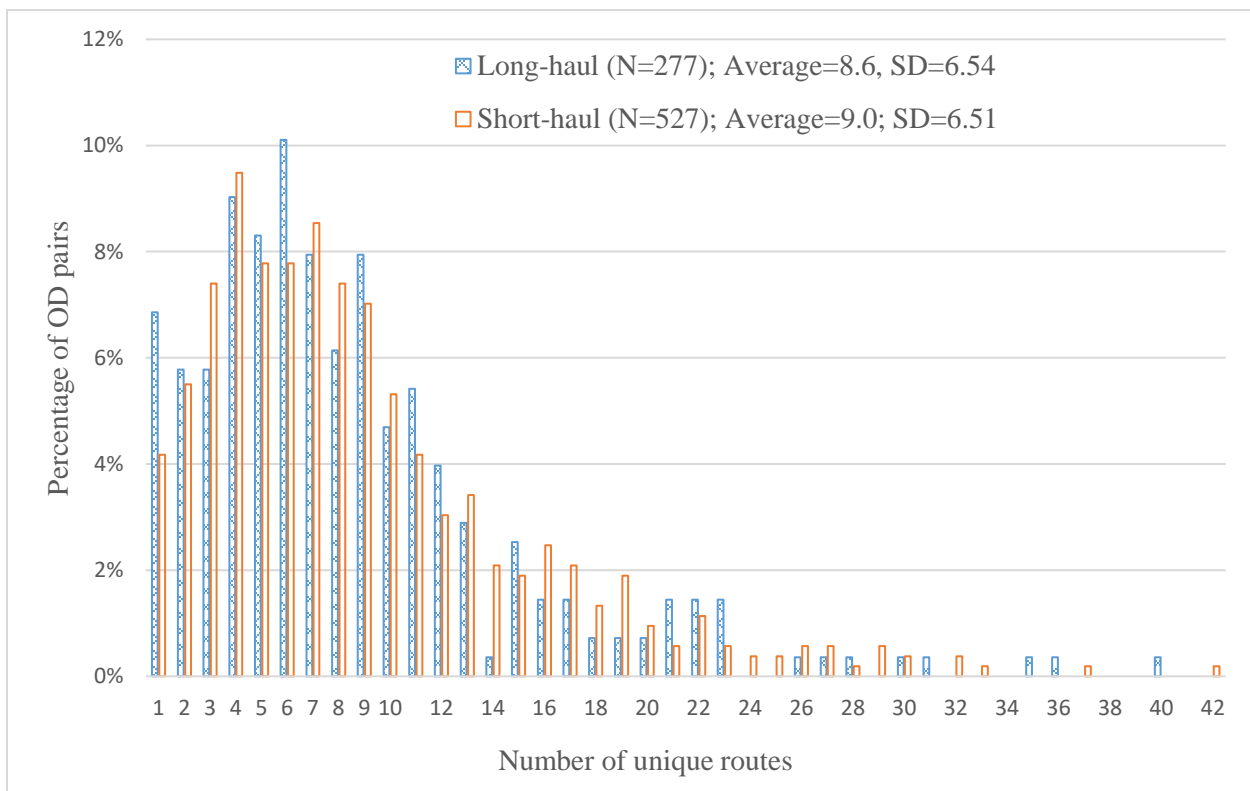


Figure 3.1 Number of Unique Routes Distribution for Long-haul and Short-haul Datasets

Figure 3.2 depicts the number of unique routes as a function of a) number of trips, b) number of truck IDs per OD pair, and c) direct OD distance. Each sub-figure shows the Pearson's correlation coefficient to quantify the correlation of the independent and dependent variables.

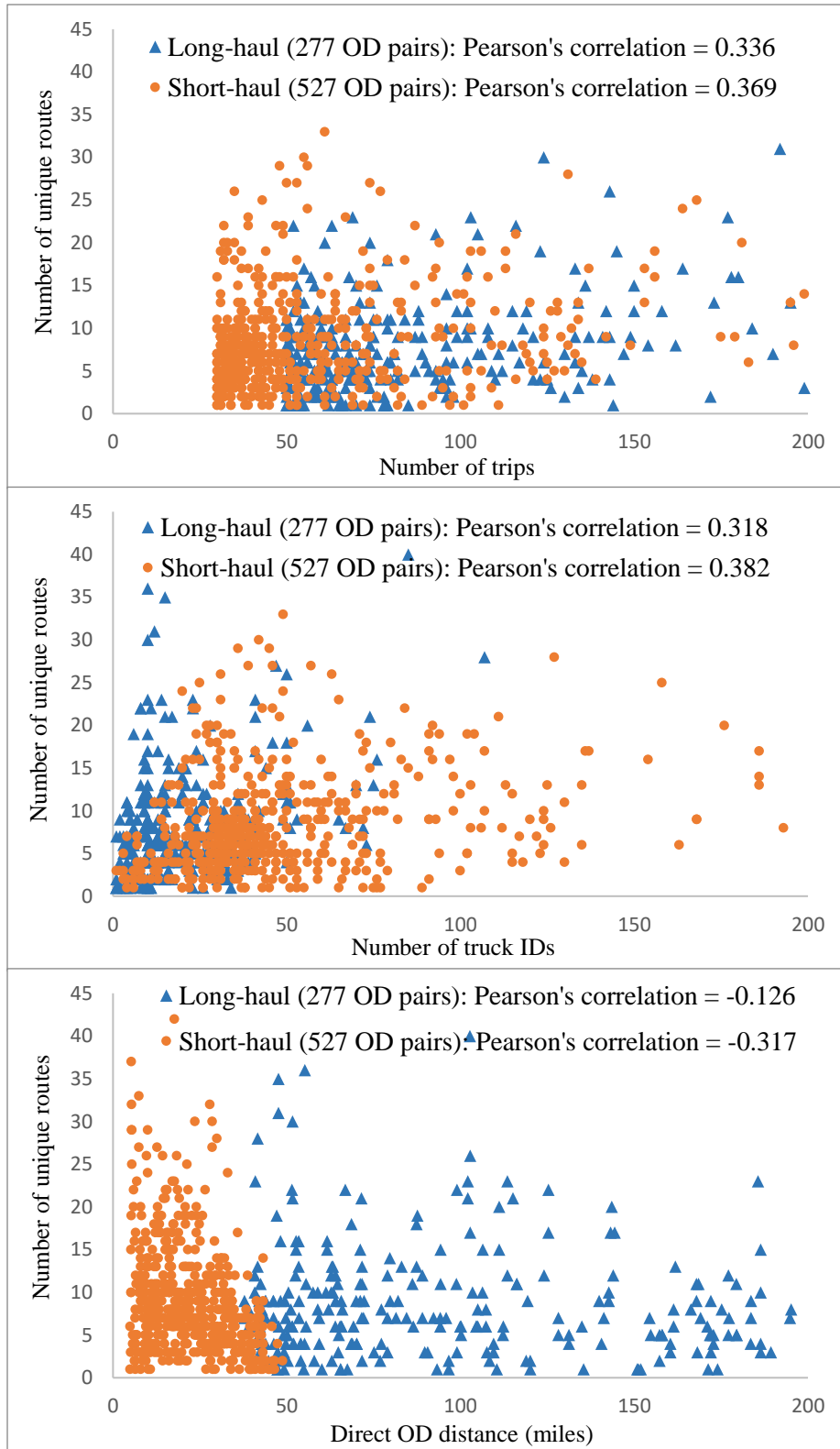


Figure 3.2 Scatter Plots of the Number of Unique Routes versus the a) Number of Trips, b) Number of Truck IDs, and c) Direct OD Distance

As shown in Figure 3.2 a), the short-haul dataset has a stronger positive correlation between number of trips and unique routes. Figure 3.2 b) shows a relative linearly increasing trend for the number of unique routes as the number of truck IDs increases. The negative Pearson correlation coefficients in Figure 3.2 c) imply some decrease in the number of unique routes as the distance between origin and destination increases.

3.3 Route Variability Metrics

3.3.1 Average Commonality Factor

Average commonality factor for a given OD pair is the mean value of the commonality factors computed across all pairs of unique routes between that OD pair. Since the earlier metric (number of unique routes) does not consider the extent of overlap (or lack thereof) between the unique routes, this metric measures the degree of overlap between all unique routes in an OD pair. Ranging between 0 and 1, an average commonality factor value closer to 0 (or 1) represents low (or high) overlap between the unique routes.

3.3.2 Average Path Size

Path size is a commonly-used metric in the route choice literature to measure the degree of overlap of two routes between an OD pair. Proposed by Ben-Akiva and Bierlaire (1999), the path size for a unique route i is defined as:

$$PS_i = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i} \right) \frac{1}{\sum_{j \in k} \delta_{aj}} \quad (3.2)$$

where Γ_i is the set of all links composing route i , l_a is the length of link a , L_i is the length of route i , and δ_{aj} is equal to 1 if a route j belonging to the unique route set k uses link a , and zero otherwise. The maximum possible value of PS is 1, and the minimum value tends to 0. A route with no overlap with any other routes has a PS value 1. Average PS in an OD pair is the mean value of all PS computed for all unique routes between that OD pair.

3.3.3 Non-overlapping Index

Complementary to the above two metrics, the degree of non-overlap among the unique routes between an OD pair is quantified using the non-overlapping index. This index is measured as the ratio between the total length of links (on unique routes) that were used only once to the total length of all links (on unique routes) that were used at least once. This index ranges between 0 and 1, where a value closer to 1 represents low overlap among unique routes.

3.3.4 Illustrations of Route Variability Metrics

Figure 3.3 demonstrates the distributions of the three aforementioned route variability metrics across OD pairs as well as their correlations with the number of unique routes. The degree of route variability generally increases for OD pairs with higher number of unique routes shown by the decrease of average commonality factor and increase of non-overlapping index. There is a strong non-linear relationship between the number of unique routes and average path size. In particular, a dense concentration of OD pairs on the top left corner of Figure 3.3 a) indicates that many OD pairs with fewer unique routes also have higher average commonality factors, which means the majority of OD pairs only have a few alternative routes, and these routes also overlap considerably with each other. Figure 3.3 b) shows a sharp decrease of average path size as the number of unique routes increases. While there is no apparent correlation between the number of unique routes and the non-overlapping index, some OD pairs have the non-overlapping index of zero, which means all road links are used by more than one unique route.

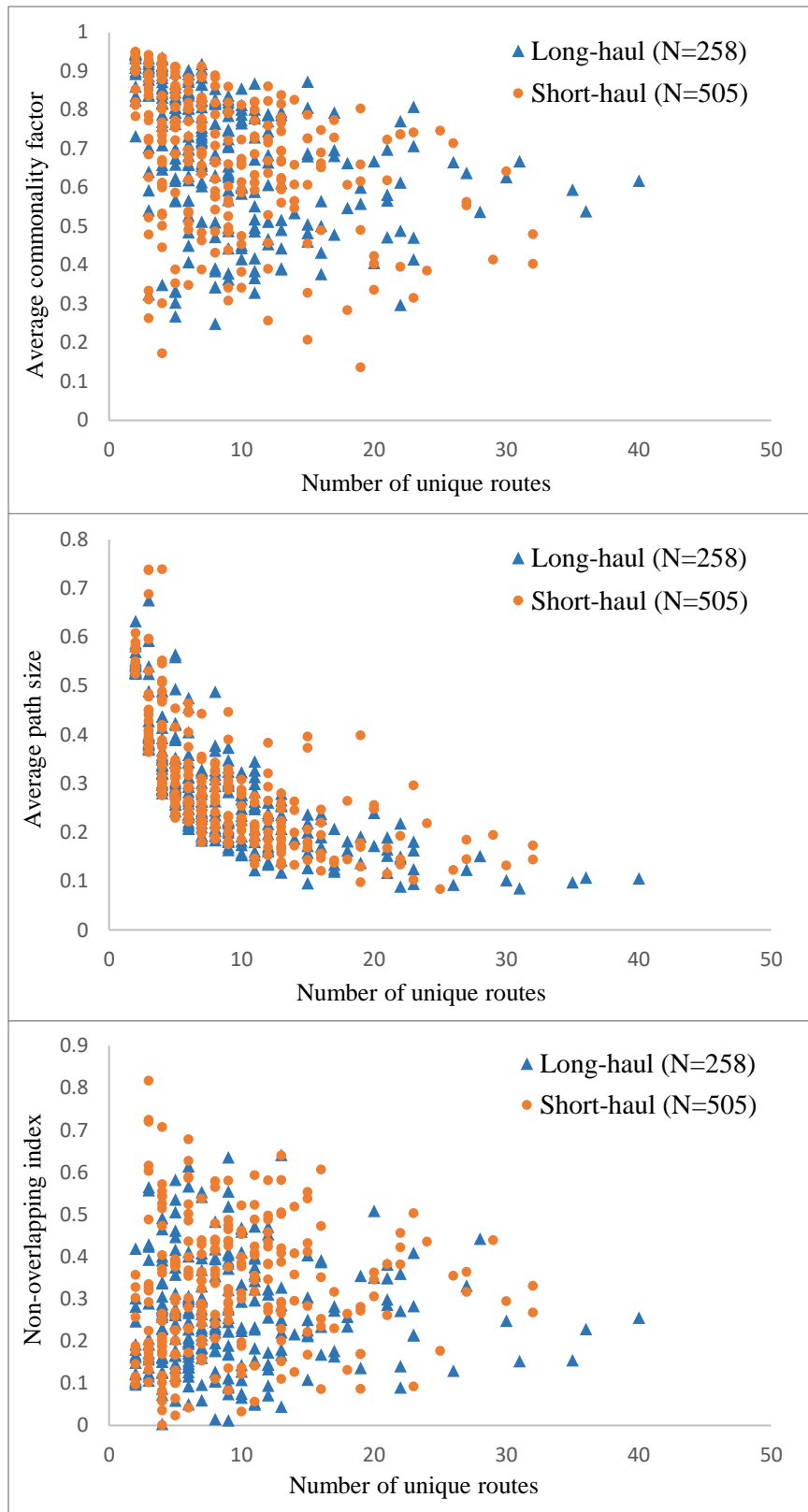


Figure 3.3 Correlations of a) Average Commonality Factor, b) Average Path Size, c) Non-overlapping Index with the Number of Unique Routes

3.4 Route Dominance Metrics

3.4.1 Standardized Variance of Route Usage

Another dimension of diversity is based on the evenness of the usage of different unique routes between an OD pair. The most even usage is when all observed trips between an OD pair are equally distributed among the observed unique routes between that OD pair. A complementary concept is the degree of dominance, when most trips are observed to have taken only one or a few unique routes.

To measure the degree of evenness in usage, the distribution of N trips among K different unique routes between a given OD pair may be characterized as a multinomial distribution, with each trip being allocated to any one of the K different unique routes. If the random variable X_k ($k = 1, 2, 3, \dots, K$) indicates the number of trips choosing route k and p_k is the proportion of trips allocated to route k , vector $X = (X_1, X_2, \dots, X_K)$ follows a multinomial distribution with parameters N and p , where $p = (p_1, p_2, \dots, p_K)$. The variance of such multinomial-distributed random variables is $Var(X_k) = N * p_k * (1 - p_k)$.

The variance of route usage between an OD pair is defined as the sum of variances of usage frequency for each route, as: $N * \sum_1^K p_k * (1 - p_k)$. This metric is influenced by three factors: (1) total number of unique routes between the OD pair (more routes, higher the variance), (2) total number of observed trips between the OD pair (more trips, higher the variance), and (3) evenness of the distribution of the observed trips among various unique routes. To measure solely the nature of trip distribution without being influenced by the number of observed trips (N) or unique routes (K), this metric may be standardized as follows. For a given OD pair with N observed trips and K unique routes, the maximum possible value of variance of route usage is: $N * K * (1/K) * (1 - 1/K) = N * (1 - 1/K)$, when all trips are evenly distributed among all unique routes.

Standardized variance of route usage is the ratio of the variance of usage to the maximum possible variance, defined as:

$$\frac{\sum_1^K p_k * (1 - p_k)}{(1 - 1/K)} \quad (3.3)$$

The closer this metric is to its maximum possible value 1, the more evenly-distributed the observed trips are among various unique routes. For example, if there are 100 trips using two unique routes in an OD pair, the standardized variance of usage for that OD pair would be 1 if 50 trips take the first route and the other 50 trips take the second route. The value of this metric would become 0.36 if 90 trips take the first route and the remaining 10 trips take the second route.

3.4.2 Standardized Shannon Entropy of Route Usage

Shannon entropy (Shannon, 2001) is a metric typically-used to measure the evenness of distribution of different entities among a given number of categories. Proposed in the field of information science, the concept of entropy has been applied widely by transportation researchers to quantify the degrees of geodiversity, etc., in a land use context (Brown et al., 2009; Li et al., 2016; Yabuki et al., 2009). The Shannon entropy of usage of K unique routes between an OD pair is defined as $\sum_1^K p_k \ln(p_k)$, where p_k is the proportion of trips taking the k^{th} unique route. The maximum value of the Shannon entropy of route usage is $K * (1/K) * \ln(1/K) = \ln(1/K)$ when all trips are equally distributed among the identified unique routes between an OD pair. To eliminate the effect of number of unique routes between an OD pair, the standardized Shannon entropy of route usage is computed as:

$$\frac{\sum_1^K p_k \ln(p_k)}{\ln(1/K)} \quad (3.4)$$

The standardized Shannon entropy of route usage has the maximum possible value of 1 when all trips are evenly distributed among all unique routes.

3.4.3 Illustrations of Route Dominance Metrics

Figure 3.4 presents the distributions of the two route dominance metrics across OD pairs as well as their correlations with the number of unique routes. The increasing trends of these two metrics for OD pairs with higher number of unique routes suggest that trips were distributed more evenly when there are more options to travel between an OD pair. These trends are compatible between long-haul and short-haul datasets.

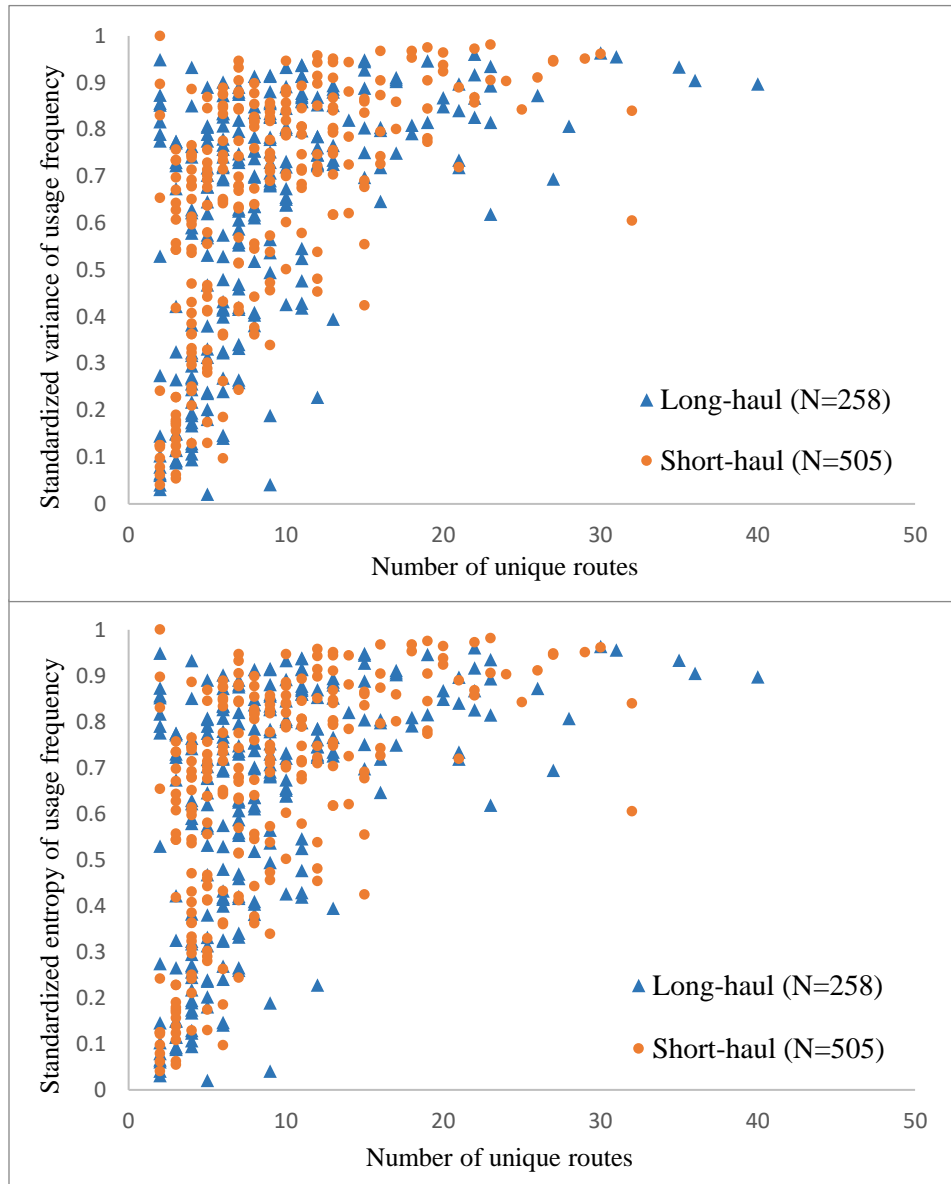


Figure 3.4 Correlation of a) Standardized Variance of Route Usage and b) Standardized Shannon Entropy of Route Usage with the Number of Unique Routes

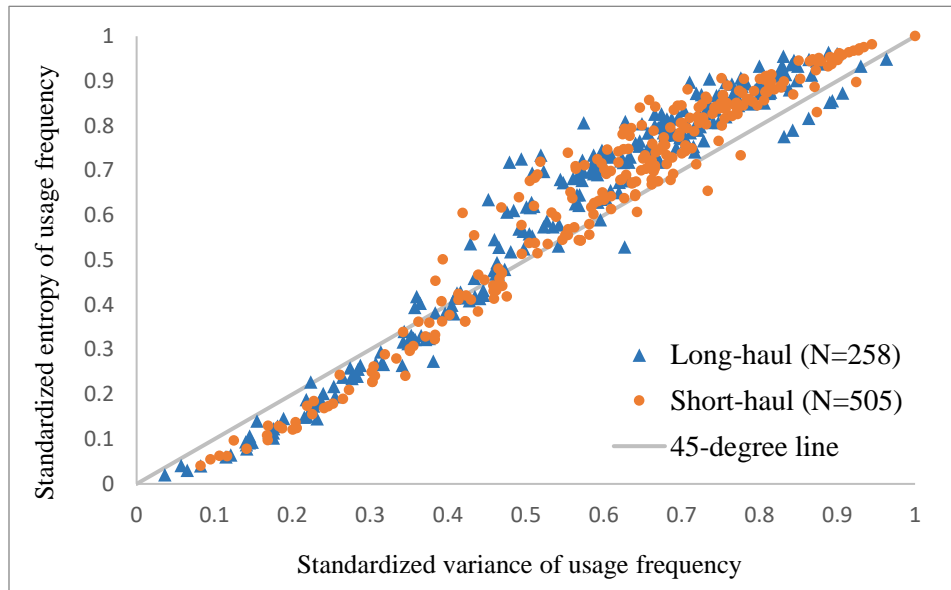
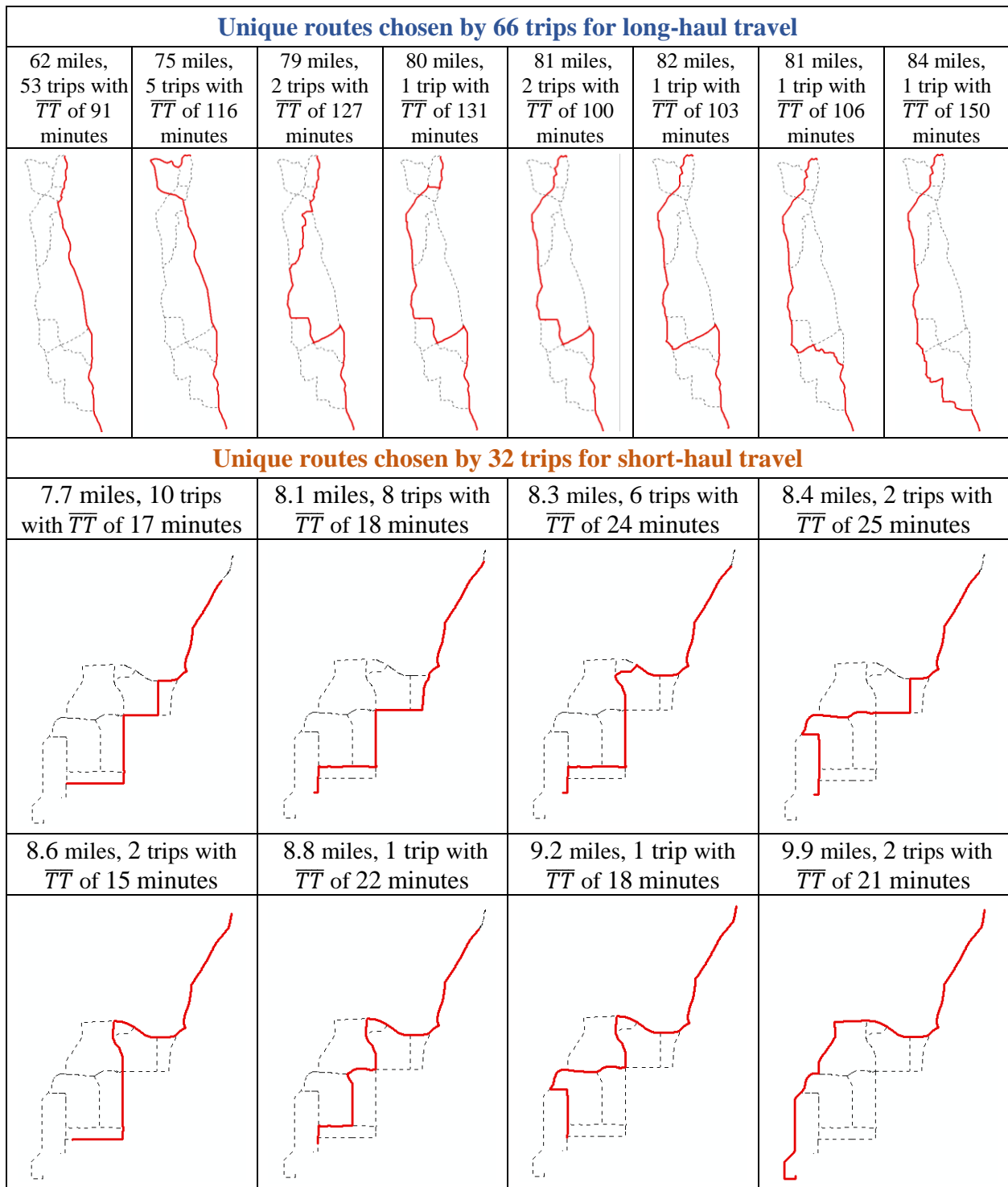


Figure 3.5 Correlation between Two Dominance Metrics

Figure 3.5 demonstrates a strong linear relationship between the two diversity metrics for both datasets. Among the two metrics capturing similar degree of route dominance, the standardized Shannon entropy of usage frequency as the dependent variable in fractional response models discussed in Chapter 6.

3.5 Illustration

To summarize the application of the above diversity metrics, Figure 3.6 presents examples of observed unique routes (indicated as the bold red lines) between two different OD pairs observed in the data. The long-haul OD pair has 8 unique routes that are 62 miles to 82 miles long. Note that many of the 8 unique routes overlap quite a bit with each other. Such overlap is measured by the average commonality factor and average path size. 53 out of 65 trips observed between this OD pair used the first unique route, indicating the dominance of the first unique route. The short-haul OD pair has 32 observed trips that are more evenly distributed among the different routes than those between the first OD pair. Such differences in dominance of route usage are measured by the standardized variance of usage and the standardized Shannon entropy of usage.



**Figure 3.6 Examples of Unique Routes
for a Long-haul OD Pair and a Short-haul OD Pair**

\overline{TT} denotes the mean travel time of all trips taking the same unique route.

CHAPTER 4: MODELING METHODOLOGY

4.1 Introduction

This section explains the statistical model structures used to analyze the determinants of the following three of the six metrics developed in this thesis—number of unique routes between and OD pair, average path size of unique routes between an OD pair, and standardized Shannon entropy of route usage between an OD pair.

4.2 Count Data Models for Number of Observed Unique Routes

Negative binomial (NB) regression (Washington et al., 2010) is an appropriate choice to model count data given by the number of observed unique routes in this research. Typically, Poisson regression is preferred if the mean of the count process is equal to the variance. If there is a significant difference between the mean and the variance of the count process, the data are said to be over-dispersed, and NB regression is preferred. Current empirical data supported the use of NB regression over Poisson regression, because of over-dispersion in the data.

In NB regression, the probability $P(y_i)$ of OD pair i having y_i number of unique routes is:

$$P(y_i) = \frac{\Gamma(1/\alpha + y_i)}{\Gamma(1/\alpha) y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left(\frac{y_i}{(1/\alpha) + \lambda_i} \right)^{y_i} \quad (4.1)$$

where $\Gamma(\cdot)$ is the gamma function, $\lambda_i = \exp(\beta X_i + \varepsilon_i)$, X_i is a vector of explanatory variables, β is a vector of parameters to be estimated, and $\exp(\varepsilon_i)$ is a Gamma-distributed disturbance term with unit mean and variance given by the dispersion parameter α . The model parameters can be estimated using a maximum likelihood estimation technique.

Depending on the count process being modeled, the regression can be right, left, or two-side truncated. To model the number of unique routes between an OD pair, count data models were left-truncated at 1, because any OD pair would have at least one unique route.

4.3 Fractional Response Models for Average Path Size and Standardized Shannon Entropy of Route Usage

It is worth noting that all diversity metrics proposed in this thesis, except the number of unique routes, ranged between 0 and 1. The fractional response model structure proposed by Papke and Wooldridge (1993) may be used to model such quantities whose values lie between 0 and 1. Before applying ordinary least squares regression, proportion data may be modeled by logit transformation of the dependent variable as:

$$\ln[y_i/(1 - y_i)] = \beta X_i \quad (4.2)$$

However, this transformation cannot be used when the dependent variable is at 0 or 1. This issue can be resolved with the fractional response model (Papke and Wooldridge, 1993) whose expected value of the dependent variable is $E(y_i|x_i) = G(x_i\beta)$ where $G(\cdot)$ is a known function with $0 < G(z) < 1 \forall z \in \mathbb{R}$. Two possible functional forms for $G(z)$ are (1) logistic function, $G(z) = \exp(z)/(1 + \exp(z))$ and (2) cumulative density function of a standard normal distribution. According to this model, the quasi likelihood of an OD pair with an observed value y_i is given by $\mathcal{L}_i(\beta) = y_i * \log[G(x_i\beta)] + (1 - y_i) * \log[1 - G(x_i\beta)]$. The parameter estimation is done using maximization of the quasi log-likelihood function. Note that since the expected value of the dependent variable modeled is always a fraction (ranging between zero and one), the expected value of the model prediction will also be a fraction.

CHAPTER 5: DESCRIPTIVE ANALYSIS

5.1 Introduction

This section provides a descriptive analysis of various metrics of diversity derived for both long-haul and short-haul datasets and the potential determinants of diversity.

5.2 Diversity Metrics

Table 5.1 summarizes the mean and standard deviation values of all diversity metrics calculated for the long-haul and short-haul datasets used in this thesis. There were 19 OD pairs in the long-haul data and 22 OD pairs in the short-haul data that had one observed unique route. Except for the number of unique routes metric, all other diversity metrics reported in the table were computed after excluding such OD pairs with a single unique route.

Table 5.1 Descriptive Statistics of Diversity Metrics

No.	Diversity Metrics	Long-haul		Short-haul	
		<i>Average</i>	<i>SD</i>	<i>Average</i>	<i>SD</i>
1	Number of unique routes	8.61	6.54	9.03	6.51
2	Average commonality factor	0.69	0.17	0.68	0.18
3	Average path size	0.28	0.12	0.29	0.14
4	Non-overlapping index	0.26	0.13	0.32	0.15
5	Standardized variance of usage	0.62	0.26	0.65	0.25
6	Standardized Shannon entropy of usage	0.57	0.21	0.61	0.21

From Table 5.1, a noteworthy pattern shows that the short-haul routes exhibit greater diversity than long-haul routes with higher average values of observed unique routes, non-overlapping index, standardized variance of route usage, and standardized Shannon entropy of route usage. The standard deviations are also higher for short-haul data, suggesting a greater incidence of higher values for this data. In other words, short-haul routes are more diverse than long-haul routes from the standpoint of lower overlap as well as lower dominance in their usage.

Figure 5.1 presents the histograms of the number of unique routes extracted for both datasets. The short-haul dataset has higher degree of diversity shown by a higher percentage of OD pairs with more unique routes. There were at most 42 unique routes found for one short-haul OD pair and 40 unique routes found for one long-haul OD pair. 88% and 85% of OD pairs have less than 20 unique routes for long-haul and short-haul data, respectively.

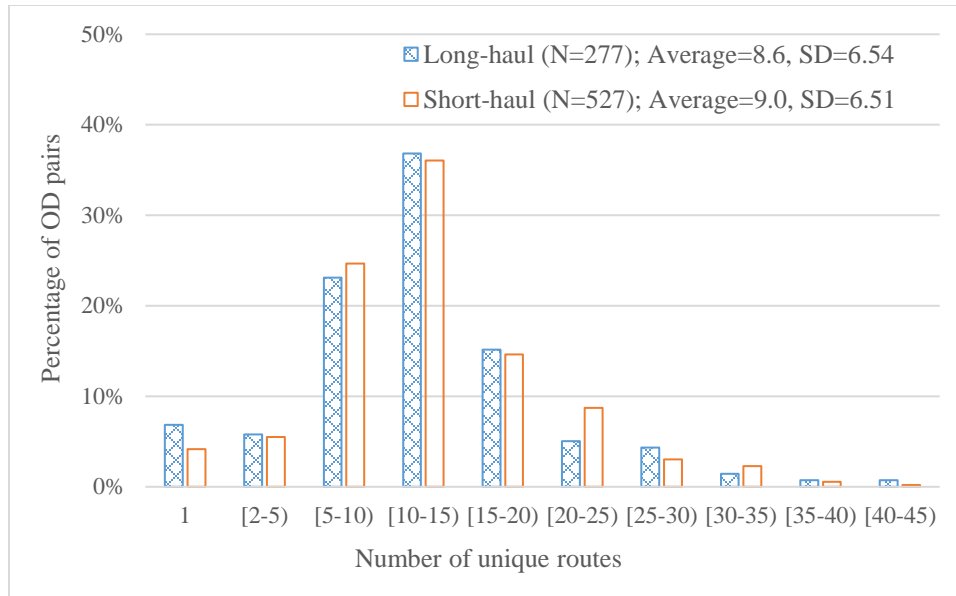


Figure 5.1 Number of Unique Routes Distribution

Figures 5.2 to 5.4 present the histograms of the three route variability metrics – average commonality factor, non-overlapping index, and average path size – calculated for both datasets. Note that, in all these figures, the numbers of OD pairs were reduced to 258 and 505 for long-haul and short-haul data, respectively, after excluding 19 and 22 OD pairs, respectively, with a single unique route. The short-haul dataset has higher degree of route variability shown by a higher percentage of OD pairs taking values closer to 1 for all three metrics. About 82% of OD pairs for both datasets have an average commonality factor greater than 0.5, which means most unique routes overlap considerably with each other.

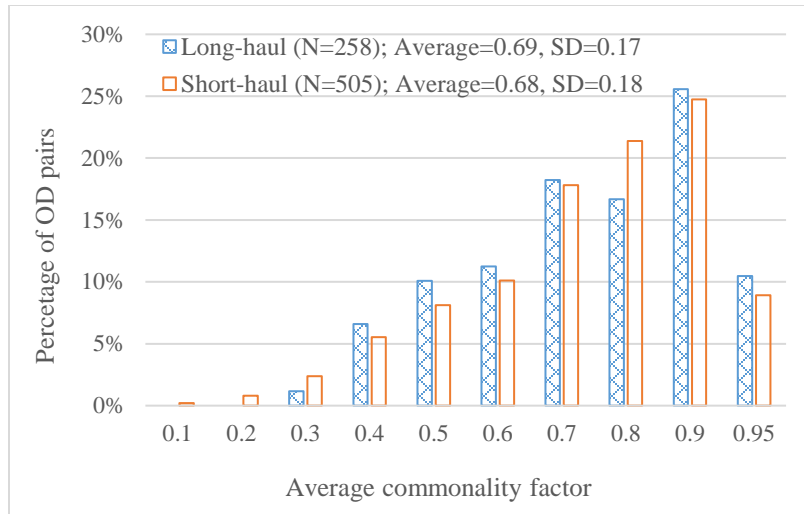


Figure 5.2 Average Commonality Factor Distribution

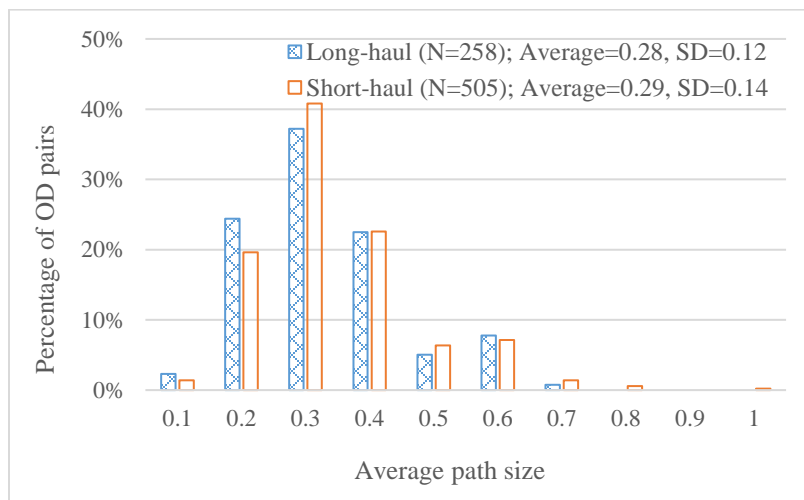


Figure 5.3 Average Path Size Distribution

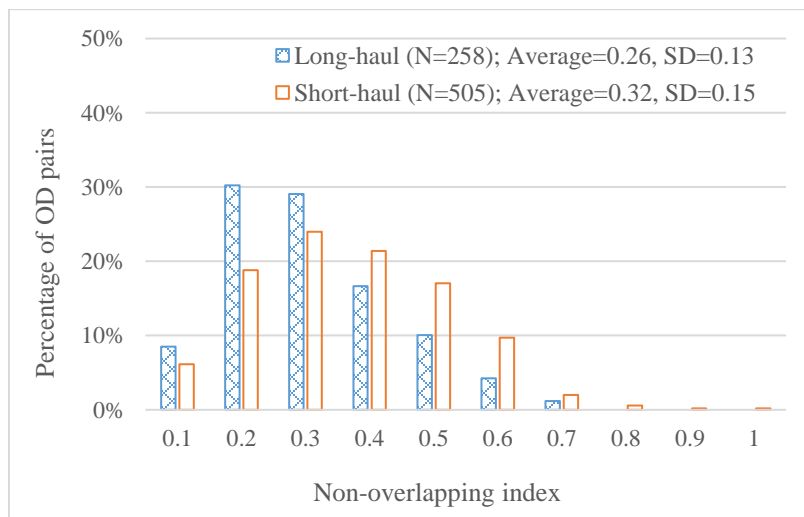


Figure 5.4 Non-overlapping Index Distribution

Figures 5.5 and 5.6 present the histograms of the two route dominance metrics calculated for both datasets. The short-haul dataset has lower degree of dominance of route usage shown by a higher percentage of OD pairs taking values closer to 1. A value of 1 for the standardized variance of usage means that each unique route in an OD pair has the same number of trips. For example, for an OD pair with 100 trips and 10 taking unique routes, standardized variance of usage of that OD pair will take the value of 1, if there are 10 trips taking each unique route.

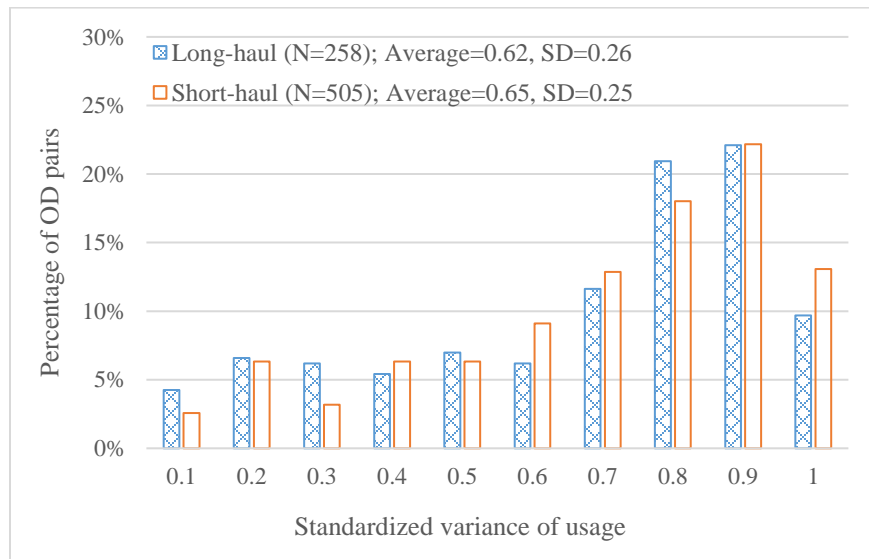


Figure 5.5 Standardized Variance of Usage Distribution

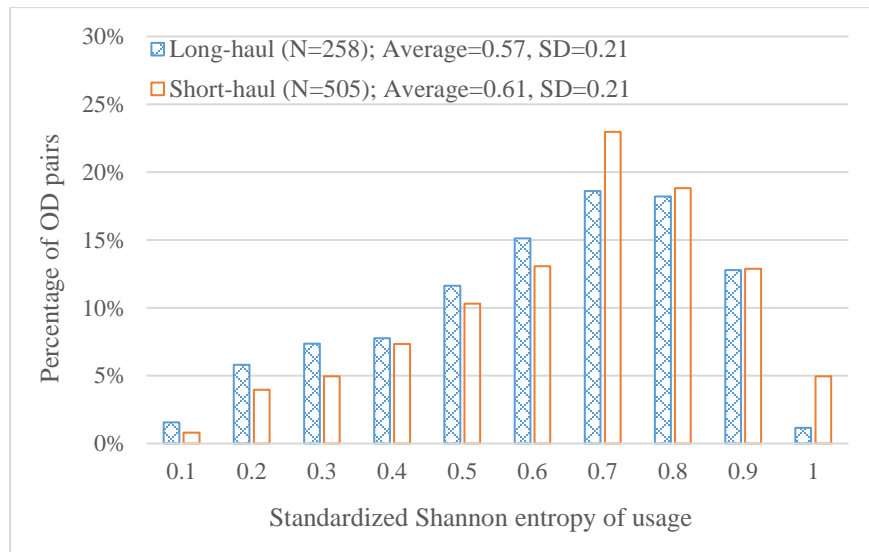


Figure 5.6 Standardized Shannon Entropy of Usage Distribution

5.3 Potential Determinants of Diversity

To explore the correlates of diversity in route choice between various OD pairs, a variety of factors describing observed travel demand, OD locations, and network structure between the OD pairs were extracted. These explanatory variables are presented in Table 5.2.

Table 5.2 Descriptive Statistics of Explanatory Variables for Route Diversity Analysis

No.	Potential Determinants of Diversity	Long-haul		Short-haul		
		Average	SD	Average	SD	
1	No. of trips observed for an OD pair	109.3	99.7	81.1	123.7	
2	No. of trucks observed for an OD pair	19.7	16.6	68.3	113.5	
3	OD airway distance (mi)	109.8	69.6	21.9	11.0	
4	Travel time of trips taking most used route (min)	SD	7.5	3.6	4.8	
5		Average	146.1	73.9	31.4	13.0
6		95 th percentile	158.5	77.4	35.9	14.3
7		5 th percentile	136.9	71.7	27.8	12.1
8	Ratio of most used route length to airway OD distance	1.2	0.1	1.2	0.4	
9	Employment density of OD TAZs (1000 jobs/sq. mi.)	All types	7.0	5.5	5.7	4.0
10		Industrial	1.5	1.1	0.9	0.7
11		Service	3.4	3.1	3.0	2.8
12		Commercial	2.1	1.9	1.8	1.3
13	Average area of OD TAZs (mi ²)	2.2	2.8	2.3	3.4	
14	Indicator if both OD TAZs are urban	0.8	0.3	0.9	0.3	
15	Average distance from centroid of all trip ends to each trip end (mi)	0.4	0.5	0.8	1.0	
16	Average distance from the TAZ centroid to major arterials (mi)	6.0	3.5	4.7	3.2	
17	Length of major arterials (mi)	Long ellipse	331.8	411.8	27.1	27.9
18		Short ellipse	274.2	386.1	15.7	17.5
19		Ending buffers	29.0	23.6	8.8	10.0
20	Length of minor arterials (mi)	Long ellipse	621.4	855.4	42.6	49.4
21		Short ellipse	502.6	800.1	18.2	25.9
22		Ending buffers	62.8	60.2	17.8	24.0
23	Length of collectors (mi)	Long ellipse	1276.6	1561.8	110.5	114.3
24		Short ellipse	1031.3	1446.2	48.7	62.2
25		Ending buffers	137.9	96.0	48.1	56.7
26	Length of local roads (mi)	Long ellipse	10155.1	14471.0	673.1	639.2
27		Short ellipse	9867.2	11496.5	311.3	349.2
28		Ending buffers	1212.1	768.9	259.8	279.2
29	Toll roads (mi)	81.4	117.6	5.8	8.2	
30	No. of rest stops	Long ellipse	9.2	12.1	1.0	1.4
31		Short ellipse	8.1	11.6	0.6	1.0
32		Ending buffers	0.3	0.9	0.4	0.8
33	No. of interchanges	Long ellipse	84.4	124.6	9.3	12.1
34		Short ellipse	60.5	107.6	4.0	6.7
35		Ending buffers	18.1	24.3	4.2	7.1
36	No. of traffic signals	Long ellipse	728.6	1043.8	59.2	86.3
37		Short ellipse	532.9	901.4	25.0	47.4
38		Ending buffers	136.2	168.4	29.1	53.0

5.3.1 Trip Characteristics

The first category of variables includes the number of trips observed for each OD pair and the number of trucks taking those trips (a measure of truck travel demand), spatial separation (straight-line distance or direct distance) between the OD locations, and travel conditions measured between the OD pair (particularly on the most used route). For the most used unique route, travel time variability and level of route circuitry (defined as the ratio of route length to the direct OD distance) were measured.

5.3.2 OD Location Characteristics

Characteristics of origin and destination TAZs include land-use descriptors (employment densities, TAZ size, urban/rural classification) and spatial dispersion of freight activity centers (calculated as the average distance of all trip end centroid to each trip end location).

5.3.3 Network Structure

To explore the impact of network structure on the diversity of observed routes, two different areas of influence between OD pairs were hypothesized, as illustrated in Figure 2.3. In the first hypothesis, the diversity of route choice between an OD pair was provided by the entire road network inside an elliptical area of influence connecting that OD pair, referred to as the long ellipse (see illustration on left side in Figure 2.3). The long ellipse's major axis was assumed to be the same distance and orientation of the straight line connecting the centroids of origin and destination TAZs. Its minor axis was set to be one-third of the major axis length. In the second hypothesis, the diversity of route choice between an OD pair was differentially impacted by two different areas of influence. The first area of influence was a circular area around the OD TAZ centroids, referred to as circular buffers. The buffer radii explored were 1, 2, and 5 miles for direct distances of 5–10, 10–20, and more than 20 miles, respectively. The second area of influence was

elliptically shaped, referred to as the short ellipse, with the major axis as the difference of straight-line distance and radius of the circular buffers on each end (see illustration on right in Figure 5.1).



Figure 5.7 Long Ellipse, Short Ellipse, and Circular Buffers

Within these hypothesized areas of influence for each OD pair, densities of various road types (major arterials, minor arterials, collectors, and local roads) were computed to characterize the network structure between the OD pair. In addition, other facilities along the roadway, such as traffic signals, intersections, interchanges, truck rest stops, were counted within long and short ellipses and circular buffers.

CHAPTER 6: ESTIMATION RESULTS

6.1 Introduction

Statistical models were estimated separately for long-haul and short-haul datasets to analyze the determinants of diversity metrics, including number of unique routes, average path size, and standardized Shannon entropy. This section presents the empirical model results.

6.2 NB Regression Model for Number of Unique Routes

Table 6.1 presents the NB regression estimation results for the number of unique routes between an OD pair, separately for long-haul and short-haul travel segments. Both model results indicate that OD pairs with a higher number of observed trucks are likely to have more unique routes. This was an expected result because more trucks traveling between an OD pair may lead to greater diversity in route choice due to heterogeneity in preferences of truck drivers, operators, and the businesses they serve. Similarly, OD pairs with more observed trips had more unique routes, in both long- and short-haul travel segments (specifically, when there are more than 150 trips in the short-haul segment). More trips represent a greater demand for travel and may lead to greater diversity in route choices as well.

The next variable in the long-haul model, indicating high travel time variability (when the difference between 95th and 5th percentile travel time on the most used route is greater than 15 minutes), suggests more unique routes since the variability in travel conditions or low reliability in travel time causes travelers to prefer alternative routes. Furthermore, in the long-haul model, deviation of the most used route from the straight-line OD distance (measured as the ratio of the most used route length to straight-line distance) had a positive influence on the number of observed

unique routes. When the most-used route is more circuitous, more available routes in the network may exist (or travelers may look for more alternatives), which decreases trucker preference for any particular route. Interestingly, neither travel time variability nor route circuitry had a significant influence in the short-haul segment.

In the context of OD location characteristics, OD pairs with larger OD TAZs were likely to have lower number of unique routes in both travel segments, perhaps because those TAZs were typically in areas with sparse network, population, and employment density and, therefore, had fewer network options to travel. For the same reason, both OD locations being in an urban zone is associated with a higher number of unique routes in the short-haul model. In the context of direct distance, OD pairs within 200 miles separation are likely to have more unique routes than those that are farther from each other. In the short-haul segment, the diversity of route choice appears to increase as spatial separation increases from small (<10 miles) to moderate (10–20 miles) and then decreases in the highest length segment. This may be because the network does not offer too many route options both for short-length (<10 miles) or long-length (>40 miles) travel. In the short-haul model, employment densities at the OD TAZs were positively correlated with the number of unique routes, perhaps because a greater employment density is a surrogate for the heterogeneity of businesses served by freight trucks, which leads to a greater diversity in route choice.

Similarly, the average distance between the TAZ-centroid of the trip ends and each trip's end coordinates (a measure of spatial dispersion of the freight activity generators in the OD TAZs) is positively associated with the number of unique routes observed between an OD pair (only in the short-haul model). The next variable in the short-haul model, average distance from TAZ-centroids to the nearest major arterial, is a surrogate for how quickly the trucks can reach a major arterial, which is negatively correlated with the number of observed unique routes.

Table 6.1 Estimation Results of Truncated Negative Binomial Regression of Number of Unique Routes

Variable Description	Long-haul Data		Short-haul Data	
	Coefficient	t-stat	Coefficient	t-stat
<i>Trip Characteristics</i>				
Logarithm of number of truck IDs	0.293	5.53	0.358	10.93
Logarithm of number of trips	0.379	4.97	--	--
Indicator (1 if more than 150 trips, 0 otherwise)	--	--	0.159	1.80
Travel time variability on most used route indicator (1 if difference of 95 th and 5 th percentile of travel time greater than 15 minutes, 0 otherwise)	0.172	1.86	--	--
Ratio of length of most used route to direct OD distance (mi/mi)	1.486	3.51	--	--
<i>OD Characteristics</i>				
Average area of OD TAZs (mi ²)	-0.040	-1.85	-0.036	-3.85
Indicator if both OD TAZs are urban zone	--	--	0.277	3.83
Indicator if direct OD distance indicator between 50 and 200 miles	0.381	3.59	N/A	N/A
Indicator if direct OD distance between 10 and 20 miles	N/A	N/A	0.128	2.84
Indicator if direct OD more than 40 miles	N/A	N/A	-0.671	-6.52
Industrial employment density (1000 jobs/mile ²)	--	--	0.215	6.41
Commercial employment density (1000 jobs/mile ²)	--	--	0.059	2.95
Average distance from centroid of all trip ends to each trip end (mi)	--	--	0.283	12.54
Average distance from TAZ centroids to nearest major or minor arterials (mi)	--	--	0.042	4.87
<i>Network Structure</i>				
Ratio of toll roads to major arterials in long ellipse (mi/mi)	1.283	3.09	--	--
Density of major and minor arterials in 5-mile buffers around both endings (mi/mi ²)	-0.583	-2.87	--	--
Density of collectors in 5-mile buffers around both endings (mi/mi ²)	0.340	2.29	--	--
Density of major, minor arterials and collectors in short ellipse (mi/mi ²)	0.231	2.12	--	--
Density of minor arterials and collectors in the long ellipse (mi/mi ²)	--	--	0.108	2.70
Proportion of major arterials to total length of major, minor arterials and collectors in long ellipse (miles/mile)	--	--	-0.534	-2.41
Proportion of minor arterials and collectors to total length of major and minor arterials and collectors in short ellipse (mi/mi)	1.375	2.24	--	--
Constant	-3.997	-5.06	-0.254	-1.41
Dispersion parameter	0.235	5.78	0.075	5.80
Number of observations (OD pairs)	277		527	
Log likelihood at convergence	-782.90		-1372.45	
Log likelihood for constant-only model	-842.92		-1617.90	
Adjusted ρ^2 with respect to constant-only model	0.056		0.141	

For variables that have significant influence in one model but not in other, "--" appears in place of parameter estimate and t-stat for that variable in latter model. "N/A" is used when variable not applicable to specific model.

In the context of network characteristics, long-haul OD pairs with a higher ratio of toll roads to major arterials captured in the long ellipse are likely to have more unique routes, because truck operators might look for alternative routes to avoid tolls. However, this variable is insignificant in the short-haul model mostly because the study region for the short-haul segment does not have many toll roads. In the long-haul model, OD locations with a higher density of major and minor arterials in circular buffers around trip ends likely are associated with a lower number of unique routes, whereas the OD locations with a higher density of collectors likely are associated with more unique routes. This may be because access to more major and minor arterials at the OD locations reduces the need to search for alternative routes. On the other hand, OD pairs with a higher density of major and minor arterials and collectors in the short ellipse are likely to have more unique routes, probably because of an increased number of route options.

For similar reasons, OD pairs with a greater proportion of minor arterials and collectors (with respect to major and minor arterials and collectors) in the short ellipse are likely to have a greater number of observed unique routes in the long-haul model. In the short-haul model, whereas the density of the minor arterials and collectors in the long ellipse has a positive influence on the number of observed unique routes, the influence of the proportion of major arterials (with respect to major, minor arterials and collectors) is negative. All these results highlight subtle but notable differences in the influence of network structure on the diversity of truck route choice between long-haul and short-haul travel segments.

Overall, modeling results were consistent between two datasets with a better overall fit for the short-haul model.

6.3 Fractional Response Models for Average Path Size

Table 6.2 presents the fractional response model estimation results for average path size estimated for OD pairs with at least two observed unique routes. OD pairs with only one unique route were removed because their average path size always takes a deterministic value of 1.

Table 6.2 Estimation Results of Fractional Response Models for Average Path Size

Variables in Average Path Size Model	Long-haul Data		Short-haul Data	
	Coefficient	t-stat	Coefficient	t-stat
Number of unique routes	-0.068	-9.66	-0.056	-9.74
Proportion of trips on the most used route	0.534	3.72	0.705	6.25
Direct OD distance (mi)	--	--	-0.012	-5.71
Direct OD distance indicator (1 if more than 200 miles, 0 otherwise)	-0.175	-1.95	--	--
Constant	-0.665	-4.73	-0.539	-4.70
Number of observations (OD pairs)	258		505	
Log pseudo likelihood at convergence	-101.08		-202.11	
Log pseudo likelihood for constant-only model	-106.37		-211.61	
Rho-square with respect to constant-only model	0.050		0.045	

Since the average path size metric measures the compounded correlations between unique routes through the degree of overlapping and usage frequency of a portion of each route, only a few explanatory variables were found to be statistically significantly correlated with the metric. Both models' parameter estimates exhibit consistent impact on the increase or decrease of average path size. As expected, OD pairs with a higher number of observed unique routes are likely to have lower average path size (i.e., greater overlap) in both models. OD pairs with a higher proportion of trips on the most used route are likely to have higher average path size (i.e., lower overlap) in both models. The presence of a dominant route may imply the presence of other longer routes that do not overlap much and are less preferable. A greater spatial separation of OD pairs is associated with a smaller value of path size (i.e., greater overlap) of the different unique routes in both models; perhaps because an increase in spatial separation may reduce the number of travel routes offered by the network. The negative constant in the model suggests that the average effect of unobserved factors on path size is likely to make the average path size to decrease toward zero, *ceteris paribus*.

6.4 Fractional Response Models for Standardized Shannon Entropy of Usage

Table 6.3 presents the fractional response model estimation results for standardized Shannon entropy of usage, estimated for OD pairs with at least two observed unique routes.

Table 6.3 Estimation Results of Fractional Response Models for Standardized Shannon Entropy of Usage

Variables in Standardized Shannon Entropy Model	Long-haul Data		Short-haul Data	
	<i>Coefficient</i>	<i>t-stat</i>	<i>Coefficient</i>	<i>t-stat</i>
Number of unique routes	0.059	5.76	0.052	6.37
Number of trips	-0.003	-4.83	-0.002	-3.74
Average path size	-1.345	-1.78	-1.667	-4.31
Average distance from centroid of all trip ends to each trip end (mi)	--	--	0.255	4.60
Constant	0.498	1.81	0.462	2.80
Number of observations (OD pairs)	258		505	
Log pseudo likelihood at convergence	-119.19		-224.05	
Log pseudo likelihood for constant-only model	-126.39		-241.98	
Rho-square with respect to constant-only model	0.057		0.074	

Similar to the lack of statistically significant explanatory variables, the fractional response models of the standardized Shannon entropy of usage frequency metric were mainly explained the number of trips and unique routes. As expected when modeling standardized Shannon entropy, OD pairs with a higher number of observed unique routes are likely to have a higher Shannon entropy (i.e., more even distribution of trips among unique routes) in both models. OD pairs with a higher number of observed trips are likely to have a more even usage of the routes in both models. OD pairs with a higher average path size (or lower overlap) among unique routes demonstrate a more uneven usage of different routes in both models. Such OD pairs with less overlapping routes are likely to have one or few dominant routes that are largely preferred over other routes. In the short-haul model, OD pairs with a greater average distance from the centroid of the trip end TAZs to all trip ends (i.e., greater spatial dispersion of freight activity generators) are likely to be associated with a more even distribution of trips among different unique routes. This suggests the influence of heterogeneity or spatial dispersion in trip ends on the heterogeneity of preferences for truck routes.

CHAPTER 7: CONCLUSIONS AND FUTURE RESEARCH

7.1 Summary of Research

This thesis presents a comprehensive exploratory analysis of truck route choice diversity in Florida for both long-haul and short-haul truck travel segments. First, six metrics were developed to measure three dimensions of diversity in truck routes between any given OD pair, as identified below:

1. The number of different routes traveled by trucks
2. The extent of overlap (or lack thereof) among the routes
 - i. Average commonality factor
 - ii. Average path size
 - iii. Non-overlapping index
3. The evenness (or, otherwise, the dominance) in usage of different routes
 - iv. Standardized variance of route usage
 - v. Standardized Shannon entropy of route usage

The above six metrics together provide a complete picture of the diversity in truck route choice between an OD pair. Using these metrics, the thesis conducted a detailed analysis of the the route diversity analysis as below:

1. Computed the diversity metrics for: (a) 277 long-haul OD pairs comprising 30,623 routes longer than 50 miles in Florida, and (b) 527 short-haul OD pairs comprising 42,884 routes between 5–50 miles long in the Tampa Bay region,

2. Performed descriptive analysis on the distribution of diversity metrics and their correlations with each other,
3. Compiled an extensive set of variables characterizing truck travel characteristics, OD location characteristics, and network structure characteristics, and
4. Estimated econometric models to understand the determinants of truck route diversity. Specifically, negative binomial regression models were estimated to explore the influence of various factors on the number of unique routes between OD pairs. Fractional response models were estimated to explore the determinants of average path size (route overlap) and standardized Shannon entropy (evenness) of route usage between OD pairs. And explored the differences in the influence of various factors on route diversity between short-haul and long-haul OD pairs.

7.2 Findings and Conclusions of Research

The analysis suggests that short-haul truck travel exhibits greater diversity in route choice than long-haul travel, in terms of number of unique routes observed, the extent of non-overlap between unique routes, as well the evenness of usage of different unique routes. Within the long-haul segment, OD pairs farther than 200 miles from each other exhibit lower diversity than those that are closer. Among the short-haul OD pairs, short distance (<10 miles) travel and long-distance travel (>40 miles) exhibit lower diversity than medium distance travel. OD pairs in urban zones are associated with a greater diversity in route choice, because urban areas offer wider network options for route choice. OD pairs with a greater number of trips and/or trucks observed (i.e., greater demand for travel) are associated with a higher number of unique routes. OD pairs with greater variability in travel conditions (travel time) and those with routes that deviate more from a straight-line have more diverse traveled routes. In addition, the network structure variables have a

considerable influence on the diversity of truck route choices. Another important finding is that the determinants and their extent of influence differs between short-haul and long-haul travel segments. For example, OD TAZ land-use (employment density and diversity of freight activity locations) has a significant influence on route choice diversity only in the short-haul segment. Furthermore, network structure variables have differential impacts on route diversity between the two segments. Finally, OD pairs with a higher number of observed unique routes have greater overlap (i.e., lower average path size) and lower dominance of route usage, whereas OD pairs with less overlapping routes exhibit greater dominance of usage.

The findings from this study can be used for improving the algorithms used in the literature for generating choice sets for truck route choice modeling; by customizing the algorithms based on the truck travel demand, OD location, and network structure characteristics found to be influential in this analysis. Specifically, the count models of number of unique routes can potentially be incorporated into route choice set generation algorithms to generate route choice alternatives for subsequent route choice analysis.

An enhanced understanding of truck route choice diversity observed in the field can also help improve truck routing policies and inform routing decisions during emergency situations. For example, as OD pairs with only one unique route tend to be more vulnerable in the event of roadway disruption, transportation planners can focus on these OD pairs to have a proactive plan for emergency routing options as well as better maintenance of critical road segments. Furthermore, as OD pairs with high average commonality factor and low number of unique routes indicate higher degree of route overlap, transportation planners can focus on these OD pairs to identify bottlenecks for both mitigating congestion as well as improving emergency response. Of

course, OD pairs with high travel demand but served by such critical links can potentially be considered for future roadway expansion to provide additional route options in the network.

7.3 Opportunities for Future Research

The current analysis can be complemented by conducting a survey of truck drivers and trucking companies to better understand the reasons behind truckers' variation in their chosen routes to travel between an OD pair. In addition, it is a fruitful avenue to examine roadway characteristics that might constrain trucks' travel route options, such as weight limit, turning radii, and vertical clearance. Further, it is of interest to explore aggregation units for trip ends other than TAZs to assess if the overall findings depend on the aggregation unit for trip ends. Finally, the diversity metrics and analysis methods used in this study can potentially be applied to understand route choice diversity for different travel markets, especially for commuting trips.

REFERENCES

- Bekhor, S., Ben-Akiva, M., Ramming, M., 2006. Evaluation of Choice Set Generation Algorithms for Route Choice Models. *Annals of Operations Research* 144 (1), 235–247.
- Ben-Akiva M., Bierlaire M., 1999. Discrete Choice Methods And Their Applications to Short Term Travel Decisions. *Handbook of Transportation Science*, Vol. 23, 5-33.
- Bliemer, M., Bovy, P., 2008. Impact of Route Choice Set on Route Choice Probabilities. *Transportation Research Record: Journal of the Transportation Research Board*. No. 2076, 10-19.
- Bovy, P., 2009. On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews*, 29 (1), 43–68.
- Brown, B.B., Yamada, I., Smith, K.R., Zick, C.D., Kowaleski-Jones, L., Fan, J.X., 2009. Mixed Land Use and Walkability: Variations in Land Use Measures and Relationships with BMI, Overweight, and Obesity. *Health & Place*, 15 (4), 1130-1141.
- Brown, D.T., Racca, D.P., 2012. Study & Calculation of Travel Time Reliability Measures.
- Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results for Interurban Networks. *International Symposium on Transportation and Traffic Theory*, 697-711.
- Hess, S., Quddus, M., Rieser-Schüssler, N., Daly, A., 2015. Developing Advanced Route Choice Models for Heavy Goods Vehicles using GPS Data. *Transportation Research Part E: Logistics and Transportation Review*, 77, 29-44.
- Hood, J., Sall, E., Charlton, B., 2011. A GPS-based Bicycle Route Choice Model for San Francisco, California. *Transportation Letters*, 3 (1), 63-75.

Jan, O., Horowitz, A., Peng, Z.-R., 2000. Using Global Positioning System Data to Understand Variations in Path Choice. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1725, 37-44.

Kamali, M., Ermagun, A., Viswanathan, K., Pinjari, A.R., 2016. Deriving Truck Route Choice from Large GPS Data Streams. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2563, 62-70.

Levinson., D., Zhu, S., 2013. A Portfolio Theory of Route Choice. *Transportation Research Part C: Emerging Technologies*, Vol. 35, 232-243.

Li, J., Qiu, R., Xiong, L., Xu, J., 2016. A Gravity-spatial Entropy Model for the Measurement of Urban Sprawl. *Science China. Earth Sciences*, 59 (1), 207.

Liao, C.-F., 2014. Generating Reliable Freight Performance Measures with Truck GPS Data: Case Study in Twin Cities Metropolitan Area, Minnesota. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2410, 21-30.

Papinski, D., Scott, D., Doherty, S., 2009. Exploring the Route Choice Decision-making Process: A Comparison of Planned and Observed Routes Obtained using Person-based GPS. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(4), 347-358.

Papke, L., Wooldridge, J., 1993. Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. National Bureau of Economic Research.

Pinjari A.R., Zanjani A.B., Thakur A., Irmania A.N., Kamali M., Short J., 2014. *Using Truck Fleet Data in Combination with Other Data Sources for Freight Modeling and Planning*. Report BDK84-977-20. Florida Department of Transportation.

http://www.fsutmsonline.net/images/uploads/reports/PINJARI_FDOT_BDK84_977_20_Final_Report_All_Chapters_Aug2014.pdf

Pinjari A.R., Kamali M., Luong T.D., Ozkul S., 2015. *GPS Data for Truck-Route Choice Analysis of Port Everglades Petroleum Commodity Flow*. Report BDK25-977-17. Florida Department of Transportation.

Shannon, C.E., 2001. A Mathematical Theory of Communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5 (1), 3-55.

Spissu, E., Meloni I., Snjust, B., 2011. Behavioral Analysis of Choice of Daily Route with Data from Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2230, 96-103.

Tahlyan, D., Luong T.D., Pinjari A.R., Ozkul S., 2017. *Development and Analysis of Truck Route Choice Data for the Tampa Bay Region using GPS Data*. Report BDK25-730-3. Florida Department of Transportation

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

Wang, Z., Goodchild, A., McCormack, E., 2016. Measuring Truck Travel Time Reliability using Truck Probe GPS Data. *Journal of Intelligent Transportation Systems* 20 (2), 103-112.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. *Statistical and Econometric Methods for Transportation Data Analysis*. CRC Press.

Woodard, D., Nogin, G., Koch, P., Racz, D., Goldszmidt, M., Horvitz, E., 2017. Predicting Travel Time Reliability using Mobile Phone GPS Data. *Transportation Research Part C: Emerging Technologies* 75, 30-44.

Yabuki, T., Matsumura, Y., Nakatani, Y., 2009. Evaluation of Pedodiversity and Land Use Diversity in terms of the Shannon Entropy. arXiv preprint arXiv:0905.2821.