March 2018

Performance Evaluation of Choice Set Generation Algorithms for Modeling Truck Route Choice: Insights from Large Streams of Truck-GPS Data

Divyakant Tahlyan
University of South Florida, dtahlyan@mail.usf.edu

Performance Evaluation of Choice Set Generation Algorithms for Modeling Truck Route Choice: Insights from Large Streams of Truck-GPS Data

by

Divyakant Tahlyan

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

Co-Major Professor: Abdul R. Pinjari, Ph.D.
Co-Major Professor: Fred L. Mannering, Ph.D.
Robert L. Bertini, Ph.D.

Date of Approval:
March 8, 2018

Keywords: truck travel behavior, discrete choice, BFS-LE, path size logit, error components logit

To my family

AKNOWLEDGEMENTS

First, I would like to thank my advisors Dr. Abdul R. Pinjari and Dr. Fred L. Mannering, who constantly supported me throughout my time at USF. Dr. Pinjari's guidance and motivation always kept me on track and encouraged me to go an extra mile in my research. Moreover, his courses on travel demand modeling and discrete choice modeling methods helped me to learn many new things and laid a strong foundation of my understanding of travel behavior research. Further, Dr. Mannering has been a source of constant inspiration to me. I greatly benefitted from his two courses on statistical and econometric methods, which were interesting, fun, and informational.

Besides my advisors, I would like to thank Dr. Robert L. Bertini, for participating in my thesis committee. I have always admired his calmness, and commitment to work. Further, I am also grateful to Dr. Michael Maness, for his constant support and guidance at various fronts.

The data used in this study 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 funded by the USDOT National University Transportation Center (UTC) consortium led by the Center for Advanced Infrastructure and Transportation (CAIT), Rutgers University. I am thankful to FDOT, USDOT, and CAIT as projects funded by them partially supported my education as USF. Further, I am grateful to another UTC named Teaching Old Models New Tricks (TOMNET), as it also partially supported my education at USF.

Parts of the results in this thesis were peer-reviewed by Transportation Research Board (TRB)'s standing committee on transportation demand forecasting and were consequently presented (Paper # 18-06000) at the TRB's 97th Annual Meeting held in Washington D.C. between 7 – 11 January 2018. I am grateful to three anonymous reviewers for their useful comments.

Lastly, I want to thank my graduate friends—Suryaprasanna Balusu, Parvathy Vinod Sheela, Trang Luong, Sashikanth Gurram, Chunfu Xin, MD Mokaddesul Hoque, Shihab Uddin, Lukai Guo, Manvitha Rajalingola, and Ashok Sampath—who supported me throughout my education at USF.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis evaluates truck route choice set generation algorithms and derives guidance on using the algorithms for effective generation of choice sets for modeling truck route choice. Specifically, route choice sets generated from a breadth first search link elimination (BFS-LE) algorithm are evaluated against observed truck routes derived from large streams of GPS traces of a sizeable truck fleet in the Tampa Bay region of Florida. A systematic evaluation approach is presented to arrive at an appropriate combination of spatial aggregation and minimum number of trips to be observed between each origin-destination (OD) location for evaluating algorithm-generated choice sets. The evaluation is based on both the ability to generate relevant routes that are typically considered by the travelers and the generation of irrelevant (or extraneous) routes that are seldom chosen. Based on this evaluation, the thesis offers guidance on effectively using the BFS-LE approach to maximize the generation of relevant routes. It is found that carefully chosen spatial aggregation can reduce the need to generate large number of routes for each trip. Further, estimation of route choice models and their subsequent application on validation datasets revealed that the benefits of spatial aggregation might be harnessed better if irrelevant routes are eliminated from the choice sets. Lastly, a comparison of route attributes of the relevant and irrelevant routes shed light on presence of systematic differences in route characteristics of the relevant and irrelevant routes.

CHAPTER 1: INTRODUCTION

1.1 Background

Route choice set generation is an essential precursor to analyzing travelers' route choice. Route choice set for a given origin-destination (OD) location pair is a subset of feasible alternative routes offered by the transportation network between that OD pair. However, the number of feasible routes in real life networks is typically very large, computationally difficult to enumerate, not readily distinguishable from each other (due to overlaps), unknown to travelers, and varies substantially from one OD pair to another (Bovy, 2009). Therefore, extraction of the set of routes known to and potentially considered by travelers (which comprises the consideration set) (Hoogendoorn-Lanser, 2005; Ton et al., 2017) is a challenging task. A variety of different choice set generation algorithms have been used in the literature to generate route choice sets (Ben-Akiva et al., 1984; Bovy and Fiorenzo-Catalano, 2007; de la Barra et al., 1993; Frejinger et al., 2009; Prato and Bekhor, 2006; Rieser-Schüssler et al., 2013; Schuessler and Axhausen, 2009). Most of these algorithms focus on generating alternative routes that are behaviorally realistic (for example, acyclic routes) and diverse (i.e., routes that do not overlap too much to become indistinguishable), with a primary goal to maximize the generation of relevant routes that are likely to be taken by travelers while reducing the generation of irrelevant routes that are not typically considered by travelers (for example, routes that involve large detours from shortest paths). As the composition of choice sets potentially can have a significant impact on route choice model estimation and prediction results (Bliemer and Bovy, 2008; Prato and Bekhor, 2007), evaluation of the generated choice sets is an important step prior to using them for route choice analysis.

A widely-used approach to evaluate route choice set generation algorithms is to measure the extent to which the generated choice sets include the observed travel routes. This approach operates at a trip level, where for each observed trip, it is assessed whether the generated route choice set includes the observed route within a certain tolerance level (Bekhor et al., 2006; Prato and Bekhor, 2007). The proportion of observed trips for which the generated choice sets include the observed routes is called coverage. Many studies in the literature report coverage ranging from 22% to 96.6% for tolerance levels ranging from 0% to 30% for various route choice set generation algorithms (Bekhor et al., 2006; Hess et al., 2015; Prato and Bekhor, 2006, 2007; Rieser-Schüssler et al., 2013; Ton et al., 2017). Using this evaluation approach, coverage can be improved by generating more routes (which may increase the computation time), improving the algorithm itself, using a better algorithm, or combining the choice sets from different algorithms. In doing so, however, one may end up with numerous irrelevant routes, which may not be considered by travelers and, therefore, potentially cause bias in estimation of choice model parameters and choice probabilities. In this context, a major drawback of the trip-level evaluation approach is that it does not offer a way to evaluate the generation of irrelevant routes, because the analyst cannot observe the travelers' consideration set from a single trip.

One way to overcome issues associated with trip-level evaluation is to perform the evaluation at an OD pair level. That is, if one can observe the routes of a sufficiently large number of trips between a given OD pair, one might get close to observing the travelers' consideration set for that OD pair. At the least, it is reasonable to assume that any feasible routes between an OD pair that are not used even after observing a sufficiently large number of trips are unlikely to be in the travelers' consideration choice sets and, therefore, need not be included in the choice sets used for analyzing route choice. With increasing availability of large data sources (such as GPS data), it is now possible to observe a substantial number of trips made by multiple travelers between a given OD pair. Therefore, using such data sources, analysts can compare observed choice sets with algorithm-generated choice sets at an OD pair level to evaluate the algorithm's ability to generate observed (i.e., relevant and/or considered) choice sets as well as the extent of generation of irrelevant routes. An evaluation of both aspects—the ability to generate relevant routes and the generation of irrelevant routes—can help improve choice set generation algorithms by increasing the capture of relevant routes while reducing irrelevant routes. Another appeal behind generating and evaluating choice sets at the OD pair-level is that typical application of route choice models for transport modeling and planning is anyway at some level of spatial aggregation in OD locations (such as traffic analysis zones).

There are a few practical issues associated with evaluating choice set generation algorithms at an OD pair level. First, for any given OD pair, a sufficiently large number of trips should be observed for an unbiased evaluation of the choice set generation algorithms. Using a small number of observed trips is likely to cause biased evaluation because those trips might provide only a censored view of the traveler's consideration choice sets. The natural question is, how many trips are necessary to observe the complete (or uncensored) consideration choice set between an OD pair? Conceptually, a rather substantial number of trips should be observed for each OD pair, but the data requirements may become prohibitively large to do so. Therefore, it may be pragmatic to determine a certain minimum number of trips that is, for practical purposes, sufficient to observe most of the consideration choice set.

The second practical issue is related to the spatial aggregation of trip ends (or OD locations). A disaggregate-level representation of OD locations for route choice analysis purposes is the link-level, where the OD pair is represented in the form of the network links at the trip ends; i.e., the first link of the route starting from the origin and the last link of the route ending at the destination. With such disaggregate spatial units, however, even with large data sources, it may not be easy to observe sufficient number of trips at the OD pair level. In addition, even if one observes a sufficient number of trips for a link-level OD pair, the observed route choices might not be diverse enough as these trips are typically made by only one or a few travelers (or, in case of freight travel, one or a few trucks belonging to only one or a few trucking companies). One way to overcome these issues is the consideration of spatially-aggregated OD pair locations, so it becomes easier to (1) observe sufficient number of trips for each (spatially) aggregated OD pair and (2) capture the diversity in route choices due to diversity in the travelers and their OD locations (or, in case of freight, diversity in the establishments trucks serve at the OD locations). Of course, spatial aggregation comes with its issues such as aggregation over large spatial units causing spurious diversity in route choices (due to the trip end locations being too far from each other) and aggregation over observed choices of multiple travelers (or trucks) masking individual-level heterogeneity in choice sets. The key lies in choosing spatial units that are neither too large to cause spurious diversity nor too small to censor true diversity in route choices between an OD pair. Carefully-selected spatial aggregation might help in observing routes that are different due to difference in the starting and/or ending network link for trips beginning and/or ending from same locations. Although aggregation leads to homogeneous choice sets for different travelers between the same OD locations, it is not inconceivable that route alternatives chosen by one traveler are relevant to (and potentially considered by) another traveler. In fact, application of route choice models for prediction purposes in transport model systems with spatially-aggregated OD pairs potentially will benefit from allowing such aggregated choice sets that are inclusive of differences in traveler and spatial characteristics (Hoogendoorn-Lanser and Van Nes, 2004).

In summary, evaluation of generated choice sets against observed choice sets from a sufficient number of trips between optimally aggregated spatial units potentially can provide insights on the strengths of choice set generation algorithms as well as ways to improve the quality of generated choice sets. The question to be addressed here is, what is the optimal combination of the spatial aggregation and the minimum number of trips to observe for each OD pair?

To improve choice set generation, a potentially effective approach that has not received much attention in the literature is to aggregate algorithm-generated choice sets over appropriately-defined spatial units or OD pairs (similar to aggregating observed routes for evaluation purposes). Doing so can help in gaining the diversity needed in generated choice sets without having to generate too many routes for each disaggregate-level trip in the spatially aggregated OD pairs. A relevant question to be addressed here is, which is a better approach—generation of a large choice set at a disaggregate OD pair level or aggregation of small choice sets generated at a disaggregate OD pair level to a spatially-aggregated OD pair? Also, how many routes should be generated at a disaggregate level, if they are aggregated to a spatially-larger OD pair, and how can irrelevant route alternatives be reduced while increasing the capture of relevant alternatives in the choice set? Addressing these questions potentially can lead to substantial improvements to and/or effective use of existing choice set generation algorithms for route choice analysis.

1.2 Current Research

The primary goal of this research is to evaluate truck route choice set generation algorithms and derive guidance on the use of such algorithms for effective and computationally efficient generation of choice sets for modeling truck route choice. Specifically, this study focuses on the evaluation (and effective use) of the breadth first search link elimination (BFS-LE) algorithm, proposed by Rieser-Schüssler et al. (2013), which has been gaining traction in the recent literature for generating route choice sets in high resolution transportation networks.

For evaluating route choice set generation algorithms, the study provides a carefully-designed evaluation approach that takes advantage of recently-emerging large data sources that enable analysts to observe a large number of trips between a given OD pair. The evaluation design is based on determining the optimal combination of (a) the spatial aggregation to represent trip OD locations and (b) the minimum number of trips to observe for each OD pair. Further, the evaluation uses metrics to assess the ability of route choice set generation algorithms to generate relevant routes (and the diversity therein) as well as the extent of generation of irrelevant (or extraneous) routes.

Based on findings from the evaluation, the study offers guidance on using the BFS-LE approach to maximize the generation of relevant routes for freight truck route choice modeling. Specifically, it is examined whether and to what extent spatial aggregation could help in reducing the need to generate large number of routes for each trip within a spatially-aggregated OD pair (and thereby reduce the computational burden of generating large number of diverse routes for each trip). Further, route choice models are estimated and applied (on validation datasets) using different choice sets to confirm the hypotheses discussed above on effectively using BFS-LE to generate truck route choice sets that maximize the capture of relevant routes. Finally, the attributes of the BFS-LE generated routes and observed routes are compared to understand the systematic differences between relevant routes and extraneous routes. An understanding of such systematic differences can assist in eliminating extraneous routes from the generated choice sets.

All the above explorations were conducted using truck route choice data derived from large streams of truck GPS traces (more than 96 million truck GPS records) from more than 110,000 trucks traveling in the Tampa Bay Region of Florida. The raw GPS traces were map-matched to a high-resolution transportation network to derive more than 200,000 truck trips and their routes for use in this analysis. Given that the majority of route choice studies, other than a few exceptions (Arentze et al., 2012; Feng et al., 2013; Hess et al., 2015; Knorring et al., 2005), are in the context of passenger car or bicycle route choice, this study contributes to a currently small body of literature on generating route choice sets for modeling freight truck route choice.

1.3 Thesis Organization

In the remainder of this thesis, Chapter 2 describes the data used. Chapter 3 discusses the BFS-LE algorithm for route choice set generation, its implementation in this research, and the design of the evaluation approach, including different combinations of spatial aggregations and minimum number of trips considered to generate and observe choice sets for each OD pair, and the metrics used to evaluate the algorithm-generated choice sets. Chapter 4 presents the performance evaluation results, findings, and guidance on generating high quality route choice sets. Chapter 4 also presents results of the estimated route choice models for different combinations of spatial aggregation and number of trips observed for each OD pair, results from application of such models to validation datasets to validate the findings, and results from comparison of route attributes of relevant routes and irrelevant (or extraneous) routes. Chapter 5 summarizes this thesis and identifies avenues for future research.

CHAPTER 2: DATA

2.1 Introduction

The primary data used in this research, provided by the American Transportation Research Institute (ATRI), is truck-GPS data of more than 96 million GPS traces from a large fleet of trucks carrying GPS receivers (see Tahlyan et al., 2017). Geographically, the data spanned six counties of the Tampa Bay region in Florida—Hillsborough, Pinellas, Polk, Pasco, Hernando, and Citrus—and 15 miles beyond the six-county region. Temporally, the data were obtained for the first 15 days in October 2015, December 2015, April 2016, and June 2016. Before truck-GPS data could be used for the analysis in this research, it needs to be processed to derive trip and route information. Procedures to derive trips, routes and its attributes are discussed in the next few sections of this chapter.

2.2 Converting Truck-GPS Data to Truck Trips

The raw data were first converted into a database of truck trips using GPS-to-trip conversion algorithms developed by Thakur et al. (2015) and refined by Pinjari et al. (2015). Specifically, the algorithm identifies trip ends by detecting potential stops (based on travel speed) of a certain minimum duration (five minutes in our case) and using detailed land-use information to eliminate traffic stops and stops at rest areas. More than 1 million truck trips were generated along with the information on the OD location of each trip and other attributes such as trip start and end times and travel time. Subsequently, validation procedures were used to eliminate potentially problematic trips (due to GPS error or algorithmic error), highly circuitous trips with large detours potentially due to the algorithm missing a stop in between (detected by the ratio between direct OD distance and trip length less than 0.7), and trips less than five miles in length (as short truck trips would not have many route options). This resulted in more than 650,000 trips.

For the trips generated above, the traveled routes were not necessarily readily-observable in the form of network links and nodes traversed between the OD locations. The raw GPS data of those trips had to be map-matched to the roadway network to derive the traveled routes. In this study, we used a high-resolution NAVTEQ roadway network, available from the Florida Department of Transportation (FDOT), comprising more than 1.8 million links and 6.9 million nodes in the state. The network was thoroughly checked for missing links, topological and directional consistency, and strong connectivity (i.e., every node is reachable by every other node) and was converted into a directed weighted graph for later use in choice set generation.

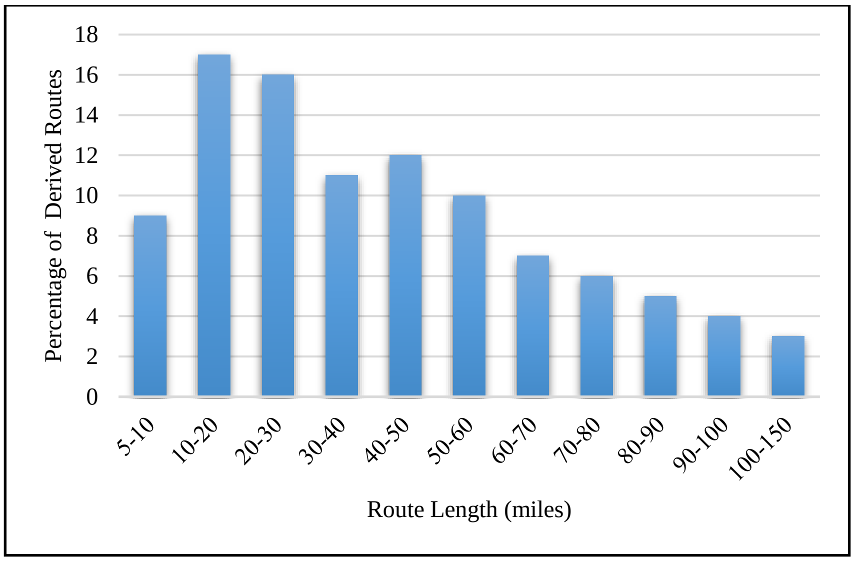


Figure 2.1 Trip Length Distribution of Derived Routes

2.3 Deriving Truck Routes from Truck Trips

To derive traveled routes for the truck trips generated from the GPS data, the GPS data were map-matched to the roadway network employing the procedures used in Kamali et al. (2016) and refined later by Tahlyan et al. (2017). High-frequency (i.e., closely spaced) GPS data are necessary for accurately deriving the traveled routes. GPS data for only about 50% of the derived truck trips were sufficient and spaced closely enough to avoid missing links in the routes derived from map-matching. For another 10% of the trips, some GPS data points could not be map-matched to an accurate network link, because the GPS data was not close to any link. After eliminating all such trips, traveled routes were derived for more than

228,000 trips. For all these derived routes, an algorithm was developed and implemented to identify loops (or cycles) and routes that were too far from the original GPS data. Routes with loops and those that spatially deviated considerably from the raw GPS data were not considered for further analysis. Of the remaining 212,800 trips, 300 randomly-selected routes were validated for consistency in the direction of travel, feasibility, and presence of large detours by evaluating the sequence of links in the route and visualizing the routes on Google Earth. The validation exercise indicated high accuracy in the derived traveled routes. Such derived traveled routes were considered as observed routes against which route sets generated using choice set generation algorithms are evaluated. Trip length distribution of derived trips is presented in Figure 2.1.

2.4 Deriving Route Attributes

For each derived trip, the derived route included information on the trip OD coordinates, corresponding TAZs defined in Florida's statewide travel demand model (FLSWM), and all the network links traversed by the truck between the OD locations. In addition, for each trip, several route attributes were computed, including route length, free flow travel times (from link-level speed limit information), travel costs (derived using the procedures by Torrey et al., 2014), number of intersections, left turns, right turns, and exit/entry ramps (each of these attributes was also computed per mile and per minute of travel), proportion of toll road length, and proportion of roads of several types (interstate highways, major arterials, minor arterials, collectors, local roads). For most of these computations, R codes were written to extract the information for each route from the network. In addition, to account for the similarity (or degree of overlap) of a route with other routes in the choice set for that same OD pair, a path-size attribute (Ben-Akiva and Bierlaire,1999) was computed as: $PS_i = \sum_{a \epsilon \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \epsilon C_n} \delta_{aj}}$, where Γ_i is the set of all links in path i between the OD pair n, l_a is the length of link a, L_i is the length of path i, C_n is the choice set of routes between the OD pair n, and δ_{aj} is equal to 1 if a route $j \epsilon C_n$ uses link a, 0 otherwise. The value of path-size for a route ranges between 0 and 1 (excluding zero), where a greater path-size value indicates smaller extent of overlap (and no overlap if path-size = 1)

CHAPTER 3: CHOICE SET GENERATION AND EVALUATION METHODOLOGY

3.1 Introduction

This chapter first discusses BFS-LE algorithm for route choice set generation and its implementation in this research. Next, design of the evaluation approach, including different combinations of spatial aggregations and minimum number of trips considered to generate and observe choice sets for each OD pair, and metrics used to evaluate the algorithm-generated choice sets are discussed.

3.2 BFS-LE Algorithm and Its Implementation

The BFS-LE approach for route choice set generation belongs to the class of algorithms based on repeated least cost path search and is well-suited for extracting routes from large-scale, high-resolution networks. It is a link elimination approach (Azevedo et al., 1993) based on a repeated least cost path search, where links on the current shortest path are eliminated, one by one, to find subsequent least cost paths.[1] What distinguishes BFS-LE from other link elimination approaches is its use of a tree structure in which each node is a network. Beginning with the original network (which is the root node of the tree), any unique network obtained after the elimination of a link from a current least cost path is a node of the tree, as long as the network offers at least one feasible route for the OD pair under consideration. The nodes are arranged at various depths (d) in the tree based on the number of links eliminated. That is, d = 1 for a network obtained after removing any one link from the first least cost path between the OD pair in the root node (i.e., the original network), d = 2 for a network obtained after removing a link from the current least cost

[1] Other variants of repeated least cost search algorithms are (1) simulation (Bierlaire and Frejinger, 2005; Prato and Bekhor, 2006; Ramming, 2001), where stochasticity in travelers' perceptions of travel costs and/or their preferences is simulated to generate multiple least cost routes, (2) path labeling (Ben-Akiva et al., 1984), where several least cost paths are obtained based on different criteria/labels for the cost function, and (3) link penalty (de la Barra et al., 1993), where links in the current shortest path are penalized with additional impedance before searching for the next least cost path.

path between the OD pair in any of the nodes (or networks) at depth 1, and so on. For each node (network) at each depth, the links on the current shortest path between the OD pair under consideration comprise the breadth. The breadth first approach finishes the search for the next least cost path within a depth level, by removing links (one by one) on the current shortest paths in all nodes at that depth (i.e., across all breadths in that depth), before proceeding to the next depth level. The algorithm is aborted when a certain pre-defined number of routes are generated, a pre-defined time threshold is reached, or there are no more feasible routes to be found. The choice of the cost function to use (for least cost path search), the maximum number of routes to generate, and the time threshold are at the discretion of the analyst. To improve the computational performance of BFS-LE, Rieser-Schüssler et al. (2013) employ a topologically-equivalent network reduction technique in which nodes that are not junctions of more than two links or dead-ends are eliminated and the corresponding links are merged to form a reduced (yet topologically equivalent) network for use in choice set generation. In addition, they use the A-star landmarks routing algorithm (Lefebvre and Balmer, 2007) instead of Dijkstra's algorithm (Dijkstra, 1959) for a quicker search of the least cost path.

In this study, the original network was coded and reduced to a topologically-equivalent network, and the BFS-LE algorithm was implemented in the Python programming language.[2] For the least cost path search, the free flow travel time was used as a cost function. Following Dhakar and Srinivasan (2014), to avoid premature termination of the algorithm in situations with fewer than two outgoing links at the origin of a trip, the BFS-LE least cost search was started from the next junction or intersection in the route that had at least two outgoing links. The BFS-LE generates routes that are different from each other even by one small network link. Since travelers may not consider routes with small deviations from each other as distinct, we considered a generated route to be a *unique* route (and, therefore, a part of the choice set) only if it is different from previously generated routes by at least 5%. Specifically, for a given OD pair, *unique* routes are determined (on the fly) using the commonality factor metric proposed by Cascetta et al. (1996), which determines the degree of similarity between two routes. Commonality factor (C_{ij}) between two routes

[2] The Python code written for implementing BFS-LE in this study is available upon request.

i and j is: $C_{ij} = l_{ij}/\sqrt{L_i L_j}$, where l_{ij} is the length of shared portion between two routes and L_i and L_j are the lengths of the routes i and j, respectively. For a given OD pair, at every instance a route was generated from the BFS-LE algorithm, we considered it *unique* (and a part of the choice set) only if the commonality factors between that route and all previously generated unique routes were less than or equal to 0.95.

3.3 Evaluation Design

To evaluate choice sets generated from the BFS-LE approach, we compared them to the observed route choice sets derived from large streams of GPS data. An important aspect of this evaluation was aimed at finding the appropriate combination of spatial aggregation and minimum number of trips to be observed for each OD pair. These aspects are discussed first, followed by a discussion of the metrics used to evaluate how well the generated choice sets capture observed choice sets while not generating irrelevant routes that are not present in the observed choice sets.

3.3.1 Spatial Aggregation and Minimum Number of Trips to be Observed

1. Link-level aggregation: For all observed trips and their routes derived from the GPS data, the OD locations were represented in the form of network links at the trip ends; i.e., the first link of the route starting at the origin and the last link of the route ending at the destination. Such a link-level aggregation comprises the most disaggregate representation of OD locations.
2. XY-level aggregation: The GPS locations of trip ends were aggregated by simply rounding off the longitude and latitude values from five decimal places to two decimal places. All trips with the OD coordinates matching up to the second decimal place were combined into a single XY-level OD pair. Such rounding leads to a spatial aggregation of roughly 1 km^2 at each of the trip ends.
3. TAZ level aggregation: The observed trips were aggregated based on the TAZs defined in the Florida Statewide Travel Demand Model (FLSWM), in which the state is divided into 5,403 TAZs. The size of these TAZs vary from 0.0067 km^2 to 232.45 km^2 depending on their population and employment densities. Most of the large-size zones covered large waterbodies and/or rural locations. To avoid spurious diversity in the generated routes due to large-sized zones, we did not consider TAZ-level OD

pairs with O/D TAZ sizes beyond 10 km^2. Further, we considered TAZ-level OD pairs with the following three levels of maximum TAZ size: 2 km^2, and 5 km^2, and 10 km^2.

4. Spatial clusters: Since large TAZs potentially cause spurious diversity in routes, spatial clustering was used to aggregate trip ends in larger (than 10 km^2) TAZs into smaller spatial clusters. After preliminary experimentation with different clustering techniques, the leader clustering technique (Hartigan, 1975) was used to divide the trip ends belonging to large TAZs into smaller clusters of radius 2 km while retaining the TAZ boundaries. An advantage of the leader clustering technique over the commonly used k-mean clustering technique is that the number of clusters need not be defined a priori but an output of the algorithm.
5. Minimum number of trips to be observed: As discussed earlier, it is necessary to observe a sufficiently large number of trips for an uncensored view of route choice sets in the data. Therefore, only OD pairs that have at least a minimum number of observed trips should be considered for a fair evaluation of choice set generation algorithm. To determine the minimum required number of trips, for each of the above-discussed aggregations, we considered OD pairs with the minimum number of trips of 20, 30, 50, and 100.

3.3.2 Observed and Generated Unique Routes for Each Combination of Spatial Aggregation and Minimum No. of Trips

For each OD pair in each of the above categories, the observed routes of all trips (derived from the GPS data) were reduced to a set of unique routes using Cascetta et al.'s (1996) commonality factor formula described earlier and applying an overlap threshold of 0.95. In the unique route set for each OD pair, the commonality factor of a given route with respect to all other routes was less than 0.95. In addition to deriving the set of observed unique routes for each OD pair, the number of trips observed to have taken each unique route was also recorded.

Next, to generate route choice sets at different spatial resolutions, the BFS-LE algorithm was run to generate unique route choice sets at the link-level first. Specifically, for each link-level OD pair corresponding to all observed trips, the BFS-LE algorithm was run (in a high-performance computing cluster) up to a maximum of 15 unique routes generated or for 1 hour, whichever was earlier, unless the algorithm stopped earlier due to completion of the tree. Such link-level generated choice sets were aggregated into larger spatial units discussed above using the commonality factor formula with an overlap threshold of 0.95. For example, unique routes for different link-level OD pairs in a same TAZ-level OD pair were aggregated to generate a set of unique routes for the TAZ-level OD pair (similarly for other spatial aggregations). The hypothesis is that such aggregation, if done at a carefully-selected spatial aggregation, can potentially help in better capturing the observed routes.

3.3.3 Evaluation Metrics

Let the set of observed unique routes for an OD pair n be $O_n = \{o_1, o_2, \dots, o_i, \dots, o_{I_n}\}$ and the set of generated unique routes for that OD pair be $G_n = \{g_1, g_2, \dots, g_j, \dots, g_{J_n}\}$, where i is the index for an observed unique route, j is the index for a generated unique route, I_n is the number of observed unique routes in the n^{th} OD pair and J_n is the number of generated unique routes for that OD pair. Let k_i be the number of trips observed to have taken the unique route i (i.e., all observed trips between that OD pair whose routes have a commonality factor greater than 0.95 with the unique route i). To measure the performance of BFS-LE-based choice set generation implemented in this study, we devised three metrics to compare the observed and generated unique route sets at an OD pair level—(1) false negative error, (2) weighted false negative error, and (3) false positive error—each of which is discussed next.

1. False negative error (ε_n^-) for an OD pair n is the proportion of observed unique routes that are not generated by the choice set generation algorithm (i.e., not present in the generated unique routes set). Mathematically, $\varepsilon_n^- = 1 - \frac{\sum_{i=1}^{I_n} \delta_i}{I_n}$, where $\delta_i = 1$ if the commonality factor C_{ij} between the observed unique route i and any of the generated unique routes $j \in G_n$ is greater than 0.95, zero otherwise. ε_n^- ranges between 0 and 1; the most desirable value is 0 (when all observed routes are generated) and least desirable value is 1 (when none of the observed routes is generated).
2. Weighted false negative error (ε_{wn}^-) is the proportion of observed trips (not unique routes) whose observed unique routes are not generated by the choice set generation algorithm. It is a weighted version

of the false negative error, where the capture (by the choice set generation algorithm) of each observed unique route is weighted by the proportion of trips taking that route. Specifically, $\varepsilon_{wn}^{-} = 1 - \frac{\sum_{i=1}^{I_n} k_i \delta_i}{\sum_{i=1}^{I_n} k_i}$. It is observed in the data that only a few of the observed unique routes are used by majority of the trips. The ε_n^{-} metric equally penalizes the choice set generation algorithm for not capturing any observed unique route, regardless of the usage of that route. The weighted metric overcomes this shortcoming by penalizing an uncaptured route based on the extent of its usage.

3. False positive error (ε_n^{+}) for an OD pair n is the proportion of generated unique routes that are not presented in the observed unique routes set. This metric provides a measure of the irrelevant (or extraneous) routes generated that are not observed to have been chosen by the traveler. Specifically, $\varepsilon_n^{+} = 1 - \frac{\sum_{j=1}^{J_n} \delta_j}{J_n}$, where $\delta_j = 1$ i if the commonality factor C_{ji} between the generated unique route j and any of the observed unique routes $i \in O_n$ is greater than 0.95, zero otherwise. ε_n^{+} ranges between 0 and 1; the most desirable value is 0 (when all generated routes are observed) and least desirable value is 1 (when none of the generated routes are observed). As discussed earlier, a trip-level evaluation of the choice set generation algorithms doesn't allow one to evaluate false positives (i.e., the generation of extraneous routes).

3.3.4 Performance Evaluation

First, to evaluate the performance of the implemented BFS-LE approach, the above discussed error metrics were compared at various levels of spatial aggregation and minimum number of trips per OD pair. The same metrics were used to determine the appropriate combination of spatial aggregation and minimum number of trips for the performance evaluation. Second, for OD pairs with the determined spatial aggregation and minimum number of observed trips, the error metrics were recomputed by reducing the threshold value of commonality factor between the observed and generated choice sets from 0.95 to 0.90, 0.85, and 0.80 to assess how much the error measures would decrease. Third, for various spatial aggregations ranging from link-level to TAZ-level, we recomputed the error metrics for generated choice sets constructed out of implementing BFS-LE with the following limits on the maximum number of routes generated for each link-level OD pair: 5, 10, 15, 20, and no limit. The time limit to abort the algorithm was set to 1 hour in all cases. The resulting error metrics were analyzed to determine which is a better approach – generation of a large choice set at a disaggregate OD pair level or aggregation of small choice sets generated at a disaggregate OD pair level to a spatially aggregated OD pair? To further examine this, choice models were estimated and applied (on validation datasets) using choice sets constructed at link-level and TAZ-level aggregations; constructed from a maximum of 5 and 15 BFS-LE routes generated at the link-level. Finally, various attributes of routes that were observed as well as algorithm-generated (i.e. relevant routes) were compared with those of the extraneous (or irrelevant) routes that were generated but not observed. The comparison shed light on identifying extraneous routes for eliminating them in a post-processing step after choice set generation.

CHAPTER 4: EVALUATION RESULTS

4.1 Introduction

This chapter first presents the results and finding from the performance evaluation. Next, results of the estimated route choice models for different combinations of spatial aggregations and number of trips observed for OD pairs, and results from application of such models to validation datasets to validate the findings are presented. Lastly, results from the comparison of route attributes of relevant and irrelevant (or extraneous) routes are presented.

4.2 OD Pair-level Evaluation of Choice Set Generation Algorithm at Different Combinations of Spatial Aggregation and Minimum Number of Observed Trips

Table 4.1 presents the evaluation results for each combination of spatial aggregation and minimum number of observed trips considered at an OD pair level. Altogether, this table represents a total of 82,738 truck trips extracted from the initial set of 212,800 trips for which we had derived (and validated) routes. These 82,738 trips belong to 23,112 link-level OD pairs, which were in-turn aggregated to different spatial levels, while considering the minimum number of trips available for each spatially-aggregated OD pair. Various observations and inferences can be made from this table, each of which are discussed next.

First, the columns titled “No. of OD Pairs” and “No. of Trips” present the observed data available for each combination of spatial aggregation and minimum number of observed trips. For example, at least 20 trips were observed for 615 OD pairs at the link-level. In addition, a total of 29,003 trips were observed between these 615 OD pairs. As expected, for a given spatial aggregation, the number of OD pairs with available data decreased as the minimum number of trips increased from 20 to 100. Likewise, for a given minimum number of trips, the number of OD pairs with available data increased from a finer spatial resolution to a higher spatial aggregation.

The column titled "No. of Observed Unique Routes" reports the average number of observed unique routes (and the standard deviation) across all OD pairs in each combination of spatial aggregation and minimum trips. One can infer from this column that the number of observed unique routes per OD pair increased with increase in spatial aggregation and/or with increase in the minimum number of trips observed. In the context of spatial aggregation, a visual inspection of trip ends in different OD pairs suggested that increasing the TAZ size beyond 2 km^2 led to a spurious increase in unique routes due to the trip ends within a TAZ becoming too far from each other. In the context of the role of minimum number of trips observed, the number of unique routes observed did not stabilize even after observing a minimum of 50 trips per OD pair, suggesting a possibility that one may have to observe many more trips per OD pair to get an uncensored view of the actual route choice set. However, it can be noted that the increase of the number of observed unique routes with respect to the minimum number of observed trips occurred at a decreasing rate, with the lowest increase in the number of additional observed unique routes per unit increase in the minimum number of trips observed occurring between 50 to 100 minimum trips per OD pair. Besides, there were some outlier OD pairs (which have very high number of observed unique routes) among those with a minimum of 100 trips that skewed the reported average values in Table 4.1. Therefore, for pragmatic reasons (such as not to lose substantial amount of data), we determined that observing a minimum of 50 trips per OD pair was sufficient to derive an observed route choice set for evaluation purposes.

The column titled "No. of Generated Unique Routes" reports the average number of generated unique routes (and standard deviation) across all OD pairs for each combination of spatial aggregation and minimum number of trips. It can be observed from comparing this column to the preceding column that the number of generated routes was generally greater than the number of observed routes for an OD pair. Further, as expected, the number of generated unique routes increased with increase in spatial aggregation, but at a higher rate than the increase in the number of observed unique routes.

The error metrics—false negative error, weighted false negative error, and false positive error—are reported in the last three sets of columns in Table 4.1. These columns report the average and standard

Table 4.1 Comparison of Number of Observed Unique Routes, Generated Unique Routes, and Errors in OD Pairs with at Least 20, 30, 50, and 100 Observed Trips at Various Levels of Aggregation

Aggregation Level	**Minimum Number of Trips**	**No. of OD Pairs**	**No. of Trips**	**No. of Observed Unique Routes**		**No. of Generated Unique Routes**		**False Negative Error**		**Weighted False Negative Error**		**False Positive Error**	
				Mean	***S.D.***	***Mean***	***S.D.***	***Mean***	***S.D.***	***Mean***	***S.D.***	***Mean***	***S.D.***
Link level	20	615	29,003	2.6	2.3	9.2	4.4	0.34	0.34	0.17	0.32	0.81	0.19
	30	335	22,327	2.8	2.4	8.9	4.5	0.38	0.35	0.19	0.35	0.81	0.19
	50	145	15,315	3.0	2.9	8.3	4.4	0.43	0.35	0.19	0.36	0.81	0.19
	100	48	8,995	3.4	2.8	7.2	4.5	0.53	0.33	0.26	0.41	0.79	0.2
XY cluster	20	1071	51,556	4.0	3.3	17.7	10.7	0.39	0.31	0.19	0.29	0.87	0.10
	30	615	40,654	4.6	3.6	18.3	11.2	0.44	0.29	0.18	0.28	0.87	0.10
	50	282	28,266	5.0	4.2	18.9	12.7	0.45	0.30	0.17	0.27	0.86	0.10
	100	80	15,008	6.2	5.4	19.9	14.3	0.55	0.24	0.19	0.29	0.86	0.09
Spatial cluster	20	966	58,774	5.5	4.3	26.0	20.1	0.41	0.29	0.18	0.25	0.87	0.09
	30	574	49,491	6.4	4.9	26.7	20.3	0.45	0.29	0.18	0.25	0.86	0.09
	50	294	39,001	7.4	5.7	28.0	19.8	0.49	0.27	0.18	0.26	0.86	0.10
	100	111	26,417	9.4	7.4	29.6	22.1	0.52	0.24	0.17	0.25	0.84	0.11
TAZ level (max. 2 km^2)	20	373	16,851	6.0	4.1	32.2	22.1	0.38	0.27	0.15	0.21	0.89	0.07
	30	205	12,989	6.8	4.5	32.6	22.6	0.43	0.26	0.14	0.19	0.88	0.07
	50	84	8,211	7.6	5.2	33.0	28.5	0.47	0.23	0.11	0.15	0.88	0.07
	100	28	4,336	8.3	6.2	33.4	28.4	0.54	0.21	0.11	0.18	0.88	0.08
TAZ level (max. 5 km^2)	20	723	40,229	6.8	4.7	36.9	28.4	0.38	0.26	0.17	0.22	0.88	0.07
	30	423	33,181	7.8	5.1	38.8	29.6	0.41	0.26	0.16	0.20	0.88	0.07
	50	196	24,602	8.9	5.8	39.2	27.1	0.44	0.23	0.14	0.19	0.87	0.07
	100	74	16,307	11.0	6.5	43.3	34.0	0.48	0.21	0.15	0.19	0.86	0.08
TAZ level (max. 10 km^2)	20	1152	70,494	7.7	5.8	41.4	33.2	0.38	0.25	0.18	0.23	0.88	0.08
	30	697	59,726	9.0	6.6	44.1	36.5	0.41	0.25	0.18	0.24	0.87	0.09
	50	336	46,047	10.7	7.8	47.6	38.0	0.44	0.24	0.17	0.23	0.87	0.09
	100	132	31,986	13.1	9.6	51.1	42.5	0.47	0.22	0.16	0.22	0.85	0.11

S.D. = standard deviation

deviation of the OD pair-level error measures across all OD pairs. Several observations can be made from these columns. First, the weighted false negative errors, ranging from 11% to 26%, were smaller than their unweighted counter parts, which range from 34% to 55%. As discussed earlier, the unweighted metric did not take into consideration the extent of usage of a route; whereas the weighted metric computes the errors based on usage of routes, with the errors on more (less) used routes carrying a greater (lower) weightage. In fact, the average weighted false negative errors were under 20% for most combinations of spatial aggregation and minimum number of observed trips. Therefore, one can infer that the BFS-LE performs well in capturing the more frequently-used routes than the less frequently used routes.

Second, for any given minimum number of trips between an OD pair, the weighted false negative errors were lowest at a spatial aggregation of TAZs of up to 2 km^2. This suggests that choice sets created by aggregating the generated routes over a spatial resolution of TAZs of up to 2 km^2 can help in improving the capture of observed routes. Interestingly, the improvement in weighted false negative errors was lost when larger-sized TAZs were included, perhaps because the observed routes between larger TAZs would have spurious diversity due to the trip ends being too far from each other. Also, the error rates for spatial aggregations of XY-level and spatial clusters were higher than those of smaller-sized TAZs. This is likely because TAZs are typically created keeping in view the transportation network structure around (as opposed to the other aggregations we created) and that small-sized TAZs provided an optimal mix of diversity in trip-starting and trip-ending links (which results in diverse routes between the TAZs), while keeping the trip ends within a concentrated area to avoid spurious diversity. It is also interesting to note that the standard deviations of weighted negative errors were smallest for the spatial aggregation of TAZ-level of up to 2 km^2. All these results suggest that route choice sets created out of aggregating routes generated between different trip-end links of small-sized TAZ pairs can potentially capture a large share of observed routes.

Third, as can be observed from the column titled "False Positive Error", the proportion of extraneous/irrelevant routes in the generated choice sets increased from the link-level to any other spatial aggregation considered in this study. As expected, increasing the capture of relevant routes (i.e., decreasing weighted false negative error rates) through spatial aggregation comes with an increase in extraneous routes as well. Interestingly, however, the average false positive error rates were not very different across different spatial aggregations other than the link-level.

Overall, the above-discussed results suggest the potential benefits of OD pair-level evaluation of choice set generation algorithms over the traditionally used trip-level evaluation. As importantly, aggregating the generated choice sets over carefully-defined spatial units (which happens to be TAZs of up to 2 km^2 in this empirical analysis) can help improve the capture of relevant routes for subsequent route choice modeling and prediction. However, it should be noted that spatial aggregation also results in an increase in the number of irrelevant routes.

4.3 Comparison of OD Pair-level Evaluation Results to Trip-level Evaluation Results

Note that the errors reported in Table 4.1 are OD pair level errors, as opposed to trip-level errors typically reported in the literature, which is simply the proportion of observed routes of all trips not captured in the generated routes[3]. The trip level error computed out of all 82,738 trips used in this study is 0.25—i.e., observed routes for 25% the trips were not present in the generated choice sets. When we examined only those trips belonging to OD pairs with a minimum of 20 trips at various spatial aggregations, the corresponding trip-level errors ranged from 0.18 for all 16,851 trips between TAZs of up to 2 km^2 size to 0.28 for all 58,774 trips between spatial clusters. These errors are not reported in the tables, but their OD-pair level counterparts are reported as weighted false negative errors in Table 4.1, which range from an average value of 0.15 for 373 OD pairs at the TAZ-level (of up to 2 km^2 size) to an average value of 0.18 for 966 OD pairs at the spatial cluster level. It is interesting to note that both the trip-level errors and OD pair-level average errors are smallest for the spatial aggregation of TAZs (of up to 2 km^2 size).

The trip-level errors from various studies in the literature that use repeated shortest path based choice set generation methods, including those from the current study, are reviewed in Table 4.2. This table presents trip-level false negative errors reported in the literature for different tolerance thresholds on the difference between observed and generated routes—0%, 5%, 10%, and 20%—along with salient features

[3] To be precise, most studies in the literature report trip-level coverage, which is 1 minus trip-level error.

of the choice set generation algorithms in the literature. Although it is difficult to compare errors reported in different studies due to differences in the modes of travel, the choice set generation algorithms, and the specifics of implementation, one can observe from the reported errors of the current study and those in another truck route choice study by Hess et al. (2015) that the use of BFS-LE approach to generate route choice sets for truck travel seems to result in relatively small trip-level errors compared to that for other modes of travel. To examine this further, we analyzed (for all 82,738 trips used in Table 4.1) how different are the observed routes from their corresponding shortest time routes and shortest distance routes on the network, again using the commonality factor metric between each observed route and the corresponding shortest route. Interestingly, more than 80% of the observed routes had commonality factors above 0.9 with respect to their corresponding shortest time route. On the other hand, only about 70% of the observed routes had commonality factors above 0.9 with respect to their corresponding shortest distance route. It appears that the BFS-LE approach based on repeated shortest time search performs well for truck route choice set generation because the chosen routes are not very different from the shortest time routes. Another plausible reason the current study had a small error rate (when compared to that in other studies) is perhaps because we generated up to a maximum of 15 *unique* route alternatives that were different from each other by at least 5% (using a commonality factor threshold of 0.95). Most (if not all) other studies consider generated routes as different from each other even if they are different from each other by a small link and generate up to a maximum of 15 or 20 such routes (which are not very different from each other). This limits the diversity of generated routes and, therefore, limits the capture of diverse observed routes.

4.4 Evaluation of Generated Choice Sets at Different Thresholds of Overlap between Observed and Generated Choice Sets

In all the analysis above, the generated unique choice sets were compared to the observed unique choice sets using a threshold value of 0.95 for the commonality factor. That is, an observed unique route was considered to be captured in the set of generated unique routes if the commonality factor between the observed route and any of the generated routes was at least 0.95. Table 4.3 provides false negative and weighted false negative errors computed for OD pairs with a minimum of 50 trips at the spatial aggregation

Table 4.2 False Negative Errors for Various Choice Set Generation Algorithms

Algorithm	Study	Mode	Max. Number of Alternatives	Important Features of Used Generation Algorithm	False Negative Error (%) Tolerance (%) 0	10	20
Breadth-first-search link elimination	Present study	Truck	15**	Use of free-flow travel time as cost function to generate routes that are at least 5 percent different from each other.	25 (at 5% tolerance)		
	Rieser-Schüssler et al. (2013)	Car	20*	Use of free-flow travel time as cost function	37	N.T.	N.T.
			100*		27	N.T.	N.T.
	Hess et al. (2015)	Truck	15*	Use of generalized cost function that includes penalties that reflect other sources of inconvenience occurring on minor rods	26	N.T.	N.T.
	Halldórsdóttir et al. (2014)	Bicycle	20*	Use of generalized cost function taking into account road types, cycle lanes, and land use	34	28	22
	Ton et al. (2017)	Bicycle	20*	Use of distance as travel cost	99	98	97
	Dhakar and Srinivasan (2014)	Car	20**	Use of commonly factor to generate routes that are at least 5% different from each other	N.T.	51	N.T.
Link elimination	Bekhor et al. (2006)	Car	N.R.	Elimination of links on shortest path (in sequence) to generate new routes	40	37	29
	Prato and Bekhor (2007)	Car	10*	Elimination from shortest path of links that takes driver farther from destination and closer to origin or compels driver to turn from high hierarchical road to low hierarchical road	42	42	30
`Labeling	Bekhor et al. (2006)	Car	3*	Generation of routes to minimize distance, free-flow time. and time	61	56	48
			16*	Use of 16 different labels to generate various routes	28	24	15
	Prato and Bekhor (2007)	Car	4*	Generation of routes to minimize distance, free-flow time, travel time, and delay	60	60	60
	Broach et al. (2010)	Bicycle	9*	Use of 11 different labels to generate various routes but still making sure that no generated route deviate from shortest path by more than 100%	80	75	65
	Ton et al. (2017)	Bicycle	N.R.	Use of various labels to generate routes	99	98	96
Calibrated labeling	Broach et al. (2010)	Bicycle	20*	Generation of routes using multiple labels and cost function parameters, calibrated using observed distribution of shortest path deviation	78	71	58
Link penalty	Bekhor et al. (2006)	Car	40*	Shortest route generation after gradual increase of impedance of all links on shortest path	43	33	20
			15*		44	34	22
	Prato and Bekhor (2007)	Car	15*	Iterative shortest route generation after increasing impedance of shortest path by factor of 1.05	46	46	38
Simulation (low variance)	Prato and Bekhor (2007)	Car	N.R	Generation of shortest path by drawing link impedances from truncated normal distribution with mean travel to travel time, variance equal to 20% of mean, left truncation limit equal to free-flow travel time, right truncation limit equal to time for speed of 10km/h	51	51	46
Simulation (high variance)	Prato and Bekhor (2007)	Car	N.R	Generation of shortest path by drawing link impedances from truncated normal distribution with mean travel to travel time, variance equal to 100% of mean, left truncation limit equal to free-flow travel time, right truncation limit equal to time for speed of 10km/h	39	38	29
Doubly stochastic generation function	Fiorenzo-Catalano et al. (2004)	Multi-modal	1600*	Repeated shortest path generation by considering stochasticity in travelers' perception of network attributes and preferences for different trip components	22	N.T.	N.T.

N.R: Maximum number of generated alternatives not reported in study.
N.T: Tolerance level not tested in study.
* Generated route alternatives were elemental alternatives (i.e. two route alternatives considered separate alternatives even if they differ from each other by one link.)
** Generated alternatives were unique alternatives (i.e. two route alternatives considered separate alternatives if they differ from each other by a certain minimum non-overlap.

of TAZ-level (of up to 2 km^2) for different thresholds values of commonality factors—0.95, 0.90. 0.85, and 0.80. It can be observed that the weighted false negative error values decreased substantially as the threshold value decreased – an average false negative error of 0.11 at 0.95 threshold value to an average false negative error of 0.04 at 0.90 threshold value. The false positive error values also decreased substantially with a decrease in the threshold value. Admittedly, threshold values of 0.90 or more are a bit too high for trips of mid-rage to long distance. However, the results do suggest that most uncaptured observed routes (with a 0.95 threshold value) are not substantially different from the generated routes, highlighting the performance of the BFS-LE algorithm implemented in this thesis.

Table 4.3 Comparison of Errors at Various Overlapping Thresholds in OD Pairs with at Least 50 Trips at TAZ Level (Max. Area = 2 km^2) Aggregation

Overlapping Threshold	Measure	False Negative	Weighted False Negative	False Positive
0.95	Mean	0.47	0.11	0.88
	S.D.	0.23	0.15	0.07
0.9	Mean	0.16	0.04	0.79
	S.D.	0.19	0.08	0.14
0.85	Mean	0.09	0.02	0.76
	S.D.	0.16	0.07	0.17
0.8	Mean	0.06	0.01	0.74
	S.D.	0.12	0.03	0.20

S.D. = standard deviation

4.5 Which is Better: Spatial Aggregation of a Limited Number of Generated Routes or Increasing the Number of Routes Generated from BFS-LE?

Findings from Table 4.1 suggested that spatial aggregation of generated routes can potentially help in increasing the capture of observed routes. Now, we examine if one can increase the capture of observed routes by generating a small number of routes at the link-level OD pairs and then spatially aggregating them to TAZ-level (instead of generating large number of routes at the link level). The hypothesis is that generating a smaller number of unique routes at the link-level and aggregating them spatially (to a TAZ-level, in this case) will lead to sufficient diversity in the generated choice sets. In doing so, we can reduce the computational burden of generating a large number of unique routes at the disaggregate level.

Table 4.4 Comparison of Errors at Various Limits on Maximum Number of Routes to Generate in OD Pairs with at Least 50 Trips at TAZ Level (Max. Area = 2 km^2) and Link Level Aggregation

Limit on No. of Unique Routes	Measure	TAZ Level (max. 2 km^2)				Link Level			
		No. of Generated Unique Routes	***False Negative***	***Weighted False Negative***	***False Positive***	***No. of Generated Unique Routes***	***False Negative***	***Weighted False Negative***	***False Positive***
5	Mean	21.10	0.49	0.11	0.83	4.50	0.45	0.20	0.75
	S.D.	10.23	0.23	0.16	0.09	0.97	0.35	0.37	0.20
10	Mean	27.90	0.47	0.11	0.86	7.04	0.43	0.19	0.80
	S.D.	16.75	0.23	0.15	0.07	2.91	0.35	0.36	0.19
15	Mean	32.16	0.47	0.11	0.88	8.28	0.43	0.19	0.81
	S.D.	22.11	0.23	0.15	0.07	4.44	0.35	0.36	0.19
20	Mean	36.19	0.46	0.11	0.88	8.59	0.42	0.19	0.81
	S.D.	25.19	0.23	0.15	0.07	4.98	0.35	0.36	0.19
No limit	Mean	37.56	0.46	0.11	0.89	8.68	0.42	0.19	0.81
	S.D.	26.69	0.23	0.15	0.07	5.24	0.35	0.36	0.19

Table 4.4 presents error measures for choice sets generated from different limits on the maximum number of generated unique routes at the link-level—5, 10, 15, 20, and no limit—for two different spatial aggregations—TAZ-level (of up to 2 km^2 size) and link-level. The columns under "TAZ level (max. 2 km^2)" show the metrics for the unique routes aggregated to the TAZ-level and the columns under "Link-level" show the metrics for unique routes at the link-level. It is remarkable to note that the average weighted false negative values (and the corresponding standard deviations) for the TAZ-level aggregation did not vary from choice sets constructed out of a maximum of 5 unique BFS-LE routes to those generated out of 20 or more (see the column titled "Weighted False Negative" under the TAZ-level columns). The same can be observed for the link-level aggregation as well (see the column titled "Weighted False Negative" under the link-level columns).

The results also suggest that route choice sets constructed out of aggregating (to a TAZ level) unique routes from running BFS-LE (at the link level) for a maximum of 5 unique routes provide a better capture of observed routes than those generated from running BFS-LE (at the link-level) for a maximum of 20 or more unique routes. This is probably because the BFS-LE algorithm may not consistently generate up to a maximum of 20 unique routes within a time span of one hour (recall that we had set a time limit of one hour per link-level OD pair); see column titled "No. of Generated Unique Routes" under the "Link Level" column, where the average number of generated routes does not increase beyond 8.68. Since our search was for unique routes that are different from each other by at least 5%, the BFS-LE would not generate as many routes as needed within one hour. Also notice that while the average number of generated unique routes at the link level increased from 4.50 to 8.68 when the maximum limit increased from 5 routes to no limit, the average weighted false negative error did not decrease discernably (it decreased from 0.20 to only 0.19), but the false positive errors increased from 0.75 to 0.81. Therefore, an effective and computationally-efficient alternative to increase the diversity of generated choice sets (and thereby increase the coverage of observed routes) is to aggregate a limited number of link-level choice sets generated from close by locations. In the current empirical context, it was sufficient to generate up to a maximum of only 5 unique routes at the link level and then aggregate all such choice sets from trip ends in a same TAZ pair (of up to 2 km^2 size). Of course, false positive errors increase with spatial aggregation. Therefore, one must estimate and apply route choice models to compare the prediction ability using different choice sets.

4.6 Estimation and Validation of Route Choice Models with Different Choice Sets

To further evaluate the hypothesis that aggregating a limited number of BFS-LE routes leads to better choice sets than generating a large number of routes from the BFS-LE without aggregation, we estimated and applied a series of route choice models from choice sets at link-level and TAZ-level aggregations constructed from up to a maximum of 5 or 15 BFS-LE alternatives. All the models were estimated on a sample of 6,453 trips and were applied on a validation sample of 1,758 trips (~20 % of the total sample) to evaluate the impact of choice set composition on route choice prediction.

Three different empirical specifications were used: path size logit (Ben-Akiva and Bierlaire, 1999), error components logit (Frejinger and Bierlaire, 2007) and error components logit with random coefficients on route attributes. The path size logit (PSL) model structure employs the theory of aggregation of alternatives (see Ben-Akiva and Lerman, 1985) to recognize that a route that overlaps with another may not be perceived as a distinct alternative. To do so, the utility of a route is corrected by including natural logarithm of a path size (PS) attribute. The utility associated with a route i for observation n is written as $U_{in} = \beta' X_{in} + \beta_{PS} ln PS_{in} + \varepsilon_{in}$, where X_{in} is a vector of observed attributes of route i, β is a corresponding vector of parameters, PS_{in} is the path size variable for route i, β_{PS} is a parameter corresponding to the path size variable, and ε_{in} is the random utility component assumed to be to IID Gumbel. The probability (P_{in}) of choosing a route i by a truck in observation n facing a choice set C_n is written as:

$$P_{in} = \frac{\exp(\beta' X_{in} + \beta_{PS} lnPS_{in})}{\sum_{j \epsilon C_n} \exp(\beta' X_{jn} + \beta_{PS} lnPS_{jn})} \quad (1)$$

The path size logit formulation accommodates correlations between route alternatives due to physical overlap between routes. However, correlations between route alternatives might also arise due to unobserved factors that are not attributable to physical overlap. To capture such correlations, we use the error components logit (ECL) model structure proposed by Frejinger and Bierlaire (2007) for route choice models. Specifically, ECL model captures the perceptual correlations among route alternatives. For example, two routes passing through different sections of a major named road (say interstate 4 (I-4) in the state of Florida) may share unobserved effects due to some specific (but) unobserved characteristics of that named road. Alternatively, two routes that have some portion of their lengths labeled as *"scenic route"* might also have shared unobserved effects. As an illustration, according to the error components logit model structure, the utilities of the routes i, j and k in a choice situation faced by a truck in observation n are written as:

$$U_{in} = \beta' X_{in} + \beta_{PS} lnPS_{in} + \sigma_a \sqrt{L_{in,a}} \xi_{n_a} + \sigma_b \sqrt{L_{in,b}} \xi_{n_b} + \varepsilon_{in} \quad (2)$$

$$U_{jn} = \beta' X_{jn} + \beta_{PS} lnPS_{jn} + \sigma_a \sqrt{L_{jn,a}} \xi_{n_a} + \varepsilon_{jn} \quad (3)$$

$$U_{kn} = \beta' X_{kn} + \beta_{PS} lnPS_{kn} + \sigma_a \sqrt{L_{kn,a}} \xi_{n_a} + \sigma_b \sqrt{L_{kn,b}} \xi_{n_b} + \varepsilon_{kn} \quad (4)$$

where $L_{in,a}$, $L_{jn,a}$, and $L_{kn,a}$ are the distances covered by routes i, j, and k, respectively, on the named road/label a. Similarly, $L_{in,b}$, and $L_{kn,b}$ are the distances covered by routes i, and k, respectively, on the named road/label b. Further, ξ_{n_a} and ξ_{n_b} are independent random variables, assumed to be standard normal and distributed independently and identically across observations. The variance-covariance matrix (Ω) of the error components in the illustration above can be written as:

$$\Omega = \begin{bmatrix} \sigma_a^2 L_{in,a} + \sigma_b^2 L_{in,b} & \sigma_a^2 \sqrt{L_{in,a} L_{jn,a}} & \sigma_a^2 \sqrt{L_{in,a} L_{kn,a}} + \sigma_b^2 \sqrt{L_{in,b} L_{kn,b}} \\ \sigma_a^2 \sqrt{L_{in,a} L_{jn,a}} & \sigma_a^2 L_{jn,a} & \sigma_a^2 \sqrt{L_{jn,a} L_{kn,a}} \\ \sigma_a^2 \sqrt{L_{in,a} L_{kn,a}} + \sigma_b^2 \sqrt{L_{in,b} L_{kn,b}} & \sigma_a^2 \sqrt{L_{jn,a} L_{kn,a}} & \sigma_a^2 L_{kn,a} + \sigma_b^2 L_{kn,b} \end{bmatrix} \quad (5)$$

As evident from the variance-covariance matrix (Ω), the correlation between two routes increases as a function of the distance two routes cover on a named road/label, regardless of whether these routes overlap or not. In addition to such ECL models, to account for unobserved heterogeneity in sensitivity to route attributes, we allowed random (normally distributed) parameters for the coefficients of route characteristics.

The PSL model estimation was carried out using the maximum likelihood estimation technique. The ECL and ECL with random parameters models were estimated using the maximum simulated likelihood estimation approach where 400 Halton draws (Bhat, 2003) were used to evaluate the multi-dimensional integral of the likelihood function. The choice sets used for all model estimations were augmented with the chosen routes (if the chosen routes were not already generated).

4.6.1 Estimation Results of Route Choice Models

Table 4.5 presents estimation results of the ECL model with a random coefficient on the travel time variable – estimated with TAZ-level choice sets built out of up to 5 BFS-LE generated routes at the link level. Estimation results suggest that routes with a lower travel time, lower travel cost, smaller proportion (in length) of tolled routes, smaller number of turns and ramps per minute, and those with a higher proportion of road length on major highways were preferred over other routes. However, there is significant unobserved heterogeneity in the sensitivity to travel time, as evidenced by the random coefficient on the travel time variable. Further, out of a total of nine different error components that were tested, those corresponding to the following four named roads turned out to be statistically significant: Interstate 4 (I-4), Interstate 75 (I-75), Polk parkway (also known as Florida's state road 570), and United States Route 19 (US-19).

Including the model reported in Table 4.5, a total of 16 models were estimated whose estimation results are not reported here to conserve space. Table 4.6 reports the following model fit measures on the estimation data for 15 of these models: log-likelihood value at convergence ($\mathcal{LL}_C$), log-likelihood value for equal shares model ($\mathcal{LL}_{ES}$), adjusted rho-square ($\overline{\rho^2}$), Akaike information criterion (AIC), and Bayesian information criterion (BIC).[4] For any given choice set, models with error components and random coefficients show better fit to estimation data. These results align with intuitive expectations and support the results reported by other studies (Frejinger and Bierlaire, 2007). Of course, one should not use such model fit measures for comparing the performance of models with different choice sets. Therefore, the next sub-section compares measures of route choice predictions using different choice sets.

Table 4.5 Route Choice Model Estimated with TAZ Level (Max. Area = 2 km^2) Choice Sets Aggregated from up to 5 BFS-LE Alternatives at Link Level

Variable Description	**Error Components Logit with Random Parameter on Travel Time Variable**	
	Parameter Estimate	***t-stat***
Travel cost ($)	-0.1261	-6.513
Travel time (min)	Mean = -0.0970 Std. Dev =0.6034	-3.003 30.635
Proportion of tolled portion of a route	-17.4014	-25.905
No. of turns per minute	-0.3996	-4.989
No. of ramps per minute	-0.2453	-2.489
Proportion of interstate portion of a route$^{\varphi}$	36.3844	36.552
Proportion of major arterial portion of a route	22.3101	22.372
Proportion of minor arterial portion of a route	12.5747	15.432
Proportion of collector portion of a route	6.2076	8.089
Natural log of path size	-2.8777	-40.71
σ_{I-4}	2.3289	17.512
σ_{I-75}	2.2604	13.956
σ_{Polk}	1.3970	9.986
σ_{US-19}	2.9823	2.72
No. of cases	6,453	
Log-likelihood at convergence	-9,681.31	
Log-likelihood for equal shares model	-19,327.52	
Rho-square	0.4991	
Adjusted rho-square	0.4983	

$^{\varphi}$ Each link in the network was classified into one of five categories: interstate, major arterial, minor arterial, collector, and local road.

[4] Interestingly, the ECL model with random parameter on travel time variable, estimated with choice sets build out of up to 15 BFS-LE generated routes at link level, did not converge.

Table 4.6 Model Fit Measures for Various Models Estimated Using Different Choice Sets

Model Specification	Model Fit Measures	Choice Set at Link Level with up to 5 BFS-LE Alternatives	Choice Set at Link Level with up to 15 BFS-LE Alternatives	Choice Set at TAZ Level (max. area = 2 km2) Aggregated from up to 5 BFS-LE Alternatives at Link Level	Choice Set at TAZ Level (max. area = 2 km2) Aggregated from up to 15 BFS-LE Alternatives at Link Level
Path Size Logit	$\mathcal{LL}_C$	-5,332.06	-6,915.12	-10,775.18	-11,970.52
	$\mathcal{LL}_{ES}$	-10,590.42	-15,951.96	-19,327.52	-21,674.72
	$\overline{\rho^2}$	0.496	0.566	0.442	0.447
	AIC	10,682.12	13,848.24	21,566.36	23,961.04
	BIC	10,672.89	13,839.01	21,559.13	23,949.81
Error Components Logit	$\mathcal{LL}_C$	-4,789.81	-6,303.43	-10,067.51	-11,331.78
	$\mathcal{LL}_{ES}$	-10,590.42	-15,951.96	-19,327.52	-21,674.72
	$\overline{\rho^2}$	0.546	0.604	0.478	0.477
	AIC	9,607.62	12,634.86	20,163.02	22,691.56
	BIC	9,588.39	12,615.60	20,143.79	22,672.33
Error Components Logit with Random Parameter on Travel Cost Variable	$\mathcal{LL}_C$	-4,727.12	-6,129.55	-9,994.44	-11,229.98
	$\mathcal{LL}_{ES}$	-10,590.42	-15,951.96	-19,327.52	-21,674.72
	$\overline{\rho^2}$	0.552	0.615	0.482	0.481
	AIC	9,482.24	12,287.10	20,018.88	22,487.96
	BIC	9,463.01	12,267.87	19,997.65	22,468.73
Error Components Logit with Random Parameter on Travel Time Variable	$\mathcal{LL}_C$	-4,609.86	--	-9,681.31	-10,810.01
	$\mathcal{LL}_{ES}$	-10,590.42	--	-19,327.52	-21,674.72
	$\overline{\rho^2}$	0.564	--	0.498	0.501
	AIC	9,245.72	--	19,392.62	21,648.02
	BIC	9,228.49	--	19,371.39	21,628.79

$\mathcal{LL}_C$ = log-likelihood value at convergence
$\mathcal{LL}_{ES}$ = log-likelihood value for equal shares model
$\overline{\rho^2}$ = adjusted rho-square
AIC = Akaike information criterion
BIC = Bayesian information criterion

4.6.2 Validation Results with Route Choice Models

As indicated earlier, a validation sample of 1,758 trips was used to evaluate the impact of choice set composition on route choice prediction. For all these cases, the choice sets used for prediction included the chosen route only if it was generated (so that the prediction results can be used to evaluate the generated choice sets). The number of cases for which the chosen route was not generated were 303 and 223 for link-level choice sets built out of up to 5 and 15 BFS-LE routes, respectively. And the number of cases for which the chosen route was not generated were 183 for both the choice sets at TAZ-level aggregation. Although the generated choice sets used to predict route choice for such observations did not include the chosen route, we noticed that many routes in the generated choice sets overlap substantially with the chosen route. Therefore, the metric used for validation of route choice predictions (on the validation dataset) is based on expected overlap of route choice predictions with the observed route. Specifically, for a trip (or observation) n with route choice set $\{1, \dots, 2, \dots, i, \dots, I\}$ and chosen route r, the expected overlap was $E(O)_n = \sum_{i=1}^{I} p_i C_{ir}$, where p_i is the probability of choosing route i from the choice set and C_{ir} is the proportion of route i common with the chosen route r. The average value (and standard deviation) of expected overlap across all trips in the validation data was used to evaluate route choice predictions by different models.

Table 4.7 Comparison of Average (and Standard Deviation) Values of Expected Overlap Across Various Choice Sets and Model Specifications

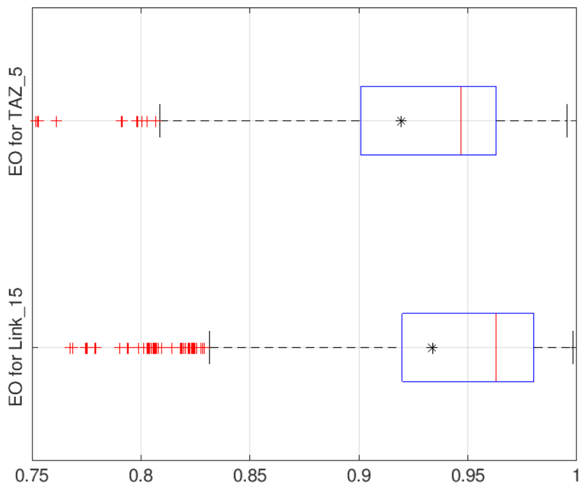
Model Specification	**Measure of expected overlap**	**Choice Set at Link Level with up to 5 BFS-LE Alternatives**	**Choice Set at Link Level with up to 15 BFS-LE Alternatives**	**Choice Set at TAZ Level (max. area = 2 km^2) Aggregated from up to 5 BFS-LE Alternatives at Link Level**	**Choice Set at TAZ Level (max. area = 2 km^2) Aggregated from up to 15 BFS-LE Alternatives at Link Level**
Path Size Logit	Mean (std. dev)	0.9290 (0.0741)	0.9340 (0.0737)	0.9192 (0.0722)	0.9190 (0.0743)
Error Components Logit	Mean (std. dev)	0.9130 (0.0734)	0.8203 (0.1878)	0.8018 (0.2752)	0.7913 (0.3530)
Error Components Logit with Random Parameter on Travel Cost Variable	Mean (std. dev)	0.9135 (0.0735)	0.8204 (0.1880)	0.8017 (0.2751)	0.7914 (0.3487)
Error Components Logit with Random Parameter on Travel Time Variable	Mean (std. dev)	0.9136 (0.0746)	--	0.8016 (0.2752)	0.7914 (0.3527)

Table 4.7 reports the validation results. Interestingly, the results suggest that the models estimated with choice sets at link-level aggregations build out of up to a maximum of 5 or 15 BFS-LE generated routes have, on average, better expected overlap (hence, better predictive ability) than the models estimated with choice sets at TAZ-level aggregations. Figures 4.1 (a) and 4.1 (b) present a closer comparison of

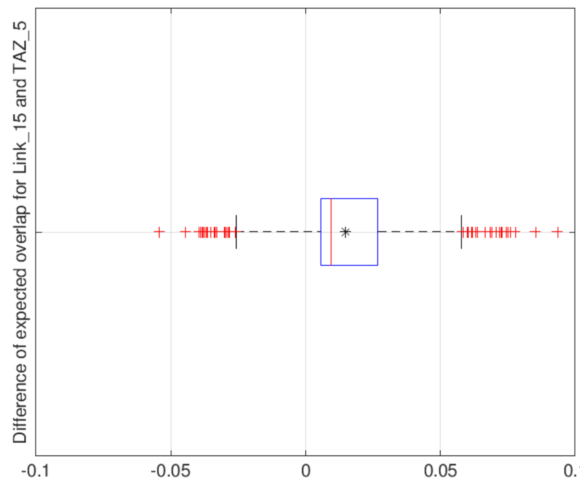
expected overlap values obtained from applying the estimated PSL model on the corresponding validation dataset for link level choice sets build out of up to 15 BFS-LE alternatives and TAZ (max. area = 2 km^2) level choice sets build after aggregating link level choice set with up to 5 BFS-LE alternatives. These figures underscore the finding presented in Table 4.7. The pattern that the models estimated with choice sets at link-level aggregations build out of up to a maximum of 5 or 15 BFS-LE generated routes have, on average, better expected overlap holds for all model specifications – PSL, ECL, and ECL with random coefficients. A possible explanation to poor predictive performance of the models estimated using aggregated choice sets (i.e., choice sets aggregated from link-level to TAZ-level) is the greater presence of irrelevant (or extraneous) routes in these choice sets. As discussed earlier, spatial aggregation of choice sets increases the diversity of generated routes and thereby improves the coverage of relevant routes. At the same time, spatial aggregation increases the presence of irrelevant routes whose overlap with the chosen route is much smaller than that of the relevant routes. This is likely a reason for a lower value of average expected overlap for TAZ-level choice sets than those for link-level choice sets. Although not in favor of the proposed spatial aggregation approach to building choice sets, this finding is not totally unexpected as the adverse effect of the presence of irrelevant routes on the prediction capability of the route choice models has been pointed out by other studies as well (Bliemer and Bovy, 2008). The results also suggest that the prediction benefits of spatial aggregation approach to choice set building (which helps in increasing the coverage of relevant alternatives) can potentially be harnessed if irrelevant routes are eliminated from aggregated choice sets.

Another interesting result is that advanced model structures, such as ECL and ECL with random parameters exhibit inferior prediction capabilities (as measured by average expected overlap) when compared to a simpler, path size logit model. This is in contrast with the model fit trends discussed earlier in the context of estimation data, where advanced model structures were associated with better fit to the estimation data. In addition, this finding appears to contrast with those of Frejinger and Bierlaire (2007) who demonstrate better predictive likelihood values for ECL models over simple path size logit models. It is worth noting, however, given our focus on the role of choice set composition in predictions, that we did not include the chosen alternative in the choice set used for prediction unless it was generated by the BFS-

LE.[5] Therefore, it is our conjecture that prediction abilities of different model structures might depend considerably on the choice set composition.



(a)



(b)

Figure 4.1 (a) Comparison and (b) Difference of Values of Expected Overlap (Obtained Using PSL Model) for Link Level Choice Set Build Out of up to 15 BFS-LE Alternatives and TAZ (max. area = 2 km^2) Level Choice Set Build Out of up to 5 BFS-LE Alternatives at Link Level Aggregation.

4.7 Comparison of the Characteristics of Observed and Generated Choice Sets

Table 4.8 presents a comparison of characteristics of the routes that were observed as well as generated (i.e., relevant routes captured in generated choice sets) to routes that were generated but not observed (i.e., extraneous routes). This comparison suggested that extraneous routes were generally longer, have a greater proportion of tolled roads and involve a greater proportion of the route through smaller roads (such as minor arterials, collectors, and local roads), more network links per mile, and more intersections and turns than relevant routes captured by the choice set generation algorithm. This is reasonable because trucks typically do not consider routes that involve going through many smaller roads and turns. A visual examination of the extraneous routes suggested that many such routes involve getting off an interstate highway to smaller roads and then getting back on to the interstate highway.

[5] Therefore, we did not use a predictive log-likelihood metric (which would be indeterminate if the chosen route did not exist in the choice set).

Table 4.8 Comparison of Route Characteristics of Observed and Generated Routes in OD Pairs with at Least 50 Trips at TAZ Level (Max. Area = 2 km^2) Aggregation

Route Characteristics	**Relevant Routes Captured in Generated Choice Sets (i.e., Observed and Generated)**		**Irrelevant/Extraneous Routes (i.e., Generated but not Observed)**	
	Mean	***S.D.***	***Mean***	***S.D.***
Length (mi)	43.350	22.360	45.050	22.640
Proportion of ramps	0.037	0.039	0.049	0.034
Proportion of tolled roads	0.000	0.062	0.028	0.063
Proportion of interstate highways and major arterials	0.784	0.284	0.667	0.255
Proportion of minor arterials	0.137	0.222	0.173	0.190
Proportion of collectors	0.061	0.105	0.131	0.101
Proportion of local roads	0.018	0.040	0.0290	0.047
No. of links	214.90	123.920	253.200	119.100
No. of links per mile	5.750	3.070	6.460	2.820
No. of intersections	89.770	77.010	119.300	72.510
No. of intersections per mile	2.580	2.070	3.220	1.960
No. of right turns	1.950	1.520	4.750	2.260
No. of left turns	1.920	1.290	4.850	2.480
Average path size	0.29*(0.09)#	0.19(0.06)	0.140	0.060

*Pathsize of observed relevant routes with respect to observed routes.
#Pathsize of generated relevant routes with respect to generated routes.
S.D. = standard deviation

A potential use of the comparison presented above is in devising strategies to remove extraneous routes in a post-processing step. For example, further analysis may be conducted to identify thresholds (either deterministic or probabilistic) on selected route attributes such as maximum number of turns/intersections per mile. Once such thresholds are identified, generated routes that do not meet the threshold criteria may be eliminated from the choice set. Another approach is to device a probabilistic approach that corrects route choice probabilities based on how likely a route is to be extraneous. Exploration of such strategies is an avenue for future research and has been discussed in detail in Chapter 5.

CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH

5.1 Summary

This study evaluated truck route choice set generation algorithms and derived guidance on using the algorithms for effective generation of choice sets for modeling truck route choice. Specifically, route choice sets generated from the breadth first search link elimination (BFS-LE) algorithm were evaluated against observed truck routes derived from large streams of GPS traces of a sizeable truck fleet in the Tampa Bay region of Florida. A carefully-designed evaluation approach was presented to arrive at an appropriate combination of spatial aggregation and minimum number of trips to be observed between each OD location for evaluating algorithm-generated route choice sets. The evaluation was based on both the ability to generate relevant routes that are considered by travelers and the generation of irrelevant (or extraneous) routes that are seldom chosen. Based on the evaluation, the study offered guidance on effectively using the BFS-LE approach to maximize the generation of relevant truck routes. Further, route choice models were estimated and applied on validation datasets to confirm findings from the above evaluation. Lastly, a comparison of route attributes of relevant and irrelevant routes was done to understand systematic differences in route characteristics of the relevant and irrelevant routes.

5.2 Conclusions

The results demonstrate the benefit of evaluating algorithm-generated choice sets against observed choice sets from large datasets at a spatially-aggregated OD-pair level (instead of performing trip-level evaluations). Doing so helps in evaluating the ability to generate relevant routes as well as the generation of irrelevant routes. Based on the evaluation results, it was found that a carefully-chosen spatial aggregation (of generated routes) can help improve the coverage of relevant routes while also reducing the need to generate substantial number of routes for each trip. In the current empirical context of truck route choice, it was found that generating up to a maximum of 5 routes at the link-level and then aggregating such routes to a TAZ-level spatial aggregation (or up to 2 km^2) provided better coverage of observed routes than that from generating more than 20 routes for each trip without spatial aggregation. The implication is that an effective and computationally-effective use of the BFS-LE algorithm for generating truck route choice sets is to generate a small number of routes at the disaggregate-level and then aggregate such routes from nearby OD locations.

The spatial aggregation approach is not without its disadvantages. Specifically, the percentage of irrelevant routes is higher in spatially aggregated route choice sets than that in disaggregate choice sets. A greater presence of irrelevant routes might offset (or even outdo) the benefits of increased coverage of relevant routes in the context of route choice prediction. For these reasons, our empirical results with data from Florida showed a poorer predictive ability of route choice models with spatially aggregated choice sets than those with disaggregate choice sets. It is likely that the prediction benefits of spatial aggregation approach to choice set building (which helps in increasing the coverage of relevant alternatives) can be better harnessed by eliminating irrelevant routes from aggregated choice sets. Exploration of alternative ways to explore irrelevant routes is a potentially fruitful avenue for near-future research.

The findings of this study also suggest that extraneous routes generated by the BFS-LE are generally longer, have a greater proportion of tolled roads, and involve a greater proportion of the route through smaller roads (such as minor arterials, collectors, and local roads), more network links per mile, and more intersections and turns than observed truck routes in Florida. Using such results, future research can focus on the development of approaches to eliminate extraneous routes from generated choice sets prior to embarking on route choice modeling.

5.3 Avenues for Future Research

Use of large streams of GPS data in route choice modeling provides an opportunity to observe the set of all relevant routes in an OD pair. As shown in thesis, the ability to observe the set of relevant and irrelevant routes makes it possible to evaluate the performance of route choice set generation algorithms in a better way. Further, it also allows a systematic comparison of attributes of relevant and irrelevant routes.

As shown in Chapter 4 of this thesis, presence of irrelevant routes in choice sets used for route choice modeling can significantly affect the prediction capability of the route choice models. A fruitful avenue for future research is to devise strategies to identify these irrelevant routes. This will not only help in improving the quality of the parameter estimates of the route choice models, but also improve the prediction results.

One possible strategy to address this issue is to identify deterministic thresholds on select route attributes such as maximum route length, maximum travel time, or maximum number of turns/intersections per mile. Routes with attributes exceeding these deterministic thresholds can be removed from the choice sets and route choice models can be estimated using the remaining route alternatives. This approach is similar to *rule-based* choice set reduction technique presented in Schuessler and Axhausen (2009). But unlike *rule-based* approach where the determination of thresholds is left to the analyst's judgement, with the ability to compare the route attributes of the relevant and irrelevant routes, analyst can make data-driven decisions to determine these thresholds.

Another possible direction is of discrete choice models with implicit choice set generation, where latent choice set models are used to associate consideration probabilities with each alternative in the universal choice set (set of all feasible alternatives in our case). These consideration probability values are used to adjust the utility of the alternatives in the universal choice set. Discrete choice models with implicit choice set generation essentially try to approximate the model proposed by Manski (1977), where the analyst's inability to observe true consideration choice set is alleviated by explicitly modeling the probability of each possible choice set and then conditionally modeling the probability of choosing an alternative. However, as the number of alternatives in the universal choice set increases, estimation of the Manski's (1977) model become extremely difficult due to large number[6] of possible choice sets from the universal choice set. Fairly recently, Martínez et al. (2009) combined the ideas from Cascetta and Papola

[6] Specifically, the number of possible choice sets with N alternatives in the universal choice set is equal to $2^N - 1$. For universal choice sets with 5, 6, and 7 alternatives, possible number of choice set are equal to 31, 63, and 127, respectively.

(2001) and Swait (2001) to propose the constrained multinomial logit (CMNL) as an approximation to the model proposed by Manski (1977). Though Bierlaire et al. (2010) showed that CMNL is not an accurate approximation of the model proposed by Manski (1977), Paleti (2015) showed that CMNL is a first order approximation of the Manski's (1977) model and also proposeed higher order approximations that provide accurate results. Even though there are a few attempts (Cascetta et al., 2002; Zhang et al., 2017) to apply first order approximation of the Manski's (1977) model in the route choice context, a thorough analysis of the higher order approximations in route choice context has not been done yet. Likely, use of these higher order approximations will reduce the impact of irrelevant routes on route choice model estimation and prediction results.

Apart from addressing the presence of irrelevant routes in the choice sets, another dimension of interest is of improving the route choice models presented in the Section 4.6 by using better exploratory variables. Specifically, incorporation of better measures of travel time (actual travel time instead of free flow travel time) and travel time variability is of interest. It is expected that incorporation of these exogenous variables will significantly improve the model fit measures and will provide extra insights on the route choice behavioral process. Lastly, application of the estimated route choice models to calculate traffic equilibrium is also of interest.

REFERENCES

1. Arentze, T., Feng, T., Timmermans, H., Robroeks, J., 2012. Context-dependent influence of road attributes and pricing policies on route choice behavior of truck drivers: results of a conjoint choice experiment. *Transportation* 39 (6), 1173-1188.
2. Azevedo, J., Costa, M.E.O.S., Madeira, J.J.E.S., Martins, E.Q.V., 1993. An algorithm for the ranking of shortest paths. *European Journal of Operational Research* 69(1), 97-106.
3. Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2006. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research* 144(1), 235-247.
4. Ben-Akiva, M., Bergman, M., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour, *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*. VNU Science Press Utrecht, The Netherlands, pp. 299-330.
5. Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. *Handbook of transportation science* 23, 5-33.
6. Ben-Akiva, M.E., Lerman, S.R., 1985. *Discrete choice analysis: theory and application to travel demand*. MIT press.
7. Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological*, 37 (9), 837-855.
8. Bierlaire, M., Frejinger, E., 2005. Route choice models with subpath components, *Swiss Transportation Research Conference*.
9. Bierlaire, M., Hurtubia, R. and Flötteröd, G., 2010. Analysis of implicit choice set generation using a constrained multinomial logit model. *Transportation Research Record: Journal of the Transportation Research Board*, (2175), pp.92-97.

10. Bliemer, M., Bovy, P., 2008. Impact of route choice set on route choice probabilities. *Transportation Research Record: Journal of the Transportation Research Board* (2076), 10-19.

11. Bovy, P.H., 2009. On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews* 29 (1), 43-68.

12. Bovy, P.H., Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica* 3 (3), 173-189.

13. Broach, J., Gliebe, J., Dill, J., 2010. Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. *Transportation Research Record: Journal of the Transportation Research Board* (2197), 89-97.

14. 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, *Internaional symposium on transportation and traffic theory*, pp. 697-711.

15. Cascetta, E. and Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, *9*(4), pp.249-263.

16. Cascetta, E., Russo, F., Viola, F.A. and Vitetta, A., 2002. A model of route perception in urban road networks. *Transportation Research Part B: Methodological*, *36*(7), pp.577-592.

17. de la Barra, T., Perez, B., Anez, J., 1993. Multidimensional path search and assignment, *PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom*.

18. Dhakar, N., Srinivasan, S., 2014. Route choice modeling using GPS-based travel surveys. *Transportation Research Record: Journal of the Transportation Research Board* (2413), 65-73.

19. Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numerische mathematik* 1 (1), 269-271.

20. Feng, T., Arentze, T., Timmermans, H., 2013. Capturing preference heterogeneity of truck drivers' route choice behavior with context effects using a latent class model. *European journal of transport and infrastructure research* 13 (4), 259-273.

21. Fiorenzo-Catalano, S., Van Nes, R., Bovy, P.H., 2004. Choice set generation for multi-modal travel analysis. *European journal of transport and infrastructure research, 4 (2)*.

22. Frejinger, E., Bierlaire, M., Ben-Akiva, M., 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological* 43 (10), 984-994.

23. Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological* 41 (3), 363-378.

24. Halldórsdóttir, K., Rieser-Schussler, N., Axhausen, K.W., Nielsen, O.A., Prato, C.G., 2014. Efficiency of choice set generation methods for bicycle routes. *European journal of transport and infrastructure research* 14 (4), 332-348.

25. Hartigan, J.A., 1975. *Clustering algorithms*. Wiley New York.

26. 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.

27. Hoogendoorn-Lanser, S., 2005. *Modelling travel behaviour in multi-modal networks*. *Ph.D. Thesis, Delft Institute of Technology.*.

28. Hoogendoorn-Lanser, S., Van Nes, R., 2004. Multimodal choice set composition: Analysis of reported and generated choice sets. *Transportation Research Record: Journal of the Transportation Research Board* (1898), 79-86.

29. 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* (2563), 62-70.

30. Knorring, J., He, R., Kornhauser, A., 2005. Analysis of route choice decisions by long-haul truck drivers. *Transportation Research Record: Journal of the Transportation Research Board* (1923), 46-60.

31. Lefebvre, N., Balmer, M., 2007. Fast shortest path computation in time-dependent traffic networks. *7th Swiss transport research conference, Ascona*, September 2007.

32. Manski, C.F., 1977. The structure of random utility models. Theory and decision, 8(3), pp.229-254.
33. Martínez, F., Aguila, F. and Hurtubia, R., 2009. The constrained multinomial logit: A semi-compensatory choice model. *Transportation Research Part B: Methodological*, *43*(3), pp.365-377.
34. Paleti, R., 2015. Implicit choice set generation in discrete choice models: Application to household auto ownership decisions. *Transportation Research Part B: Methodological*, *80*, pp.132-149.
35. Pinjari, A.R., Kamali, M., Luong, T., Ozkul, S., 2015. *GPS Data for Truck-Route Choice Analysis of Port Everglades Petroleum Commodity Flow*. Report BDV25-977-17. Florida Department of Transportation.
36. Prato, C., Bekhor, S., 2006. Applying branch-and-bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board* (1985), 19-28.
37. Prato, C., Bekhor, S., 2007. Modeling route choice behavior: how relevant is the composition of choice set? *Transportation Research Record: Journal of the Transportation Research Board* (2003), 64-73.
38. Ramming, M.S., 2001. Network knowledge and route choice. *Ph. D. Thesis, Massachusetts Institute of Technology*.
39. Rieser-Schüssler, N., Balmer, M., Axhausen, K.W., 2013. Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science* 9 (9), 825-845.
40. Schuessler, N., Axhausen, K.W., 2009. Accounting for route overlap in urban and suburban route choice decisions derived from GPS observations, *12th International Conference on Travel Behaviour Research*, Jaipur.
41. Swait, J., 2001. A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, *35*(10), pp.903-928.
42. 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.
43. Ton, D., Duives, D., Cats, O., Hoogendoorn, S., 2017. Evaluating a data-driven approach for choice set identification, *International Choice Modeling Conference*, Cape Town, South Africa.

44. 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.

45. Torrey, I., Ford, W., Murray, D., 2014. An analysis of the operational costs of trucking: 2014 Update, *American Transportation Research Institute*.

46. Train, K.E., 2009. *Discrete choice methods with simulation*. Cambridge university press.

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

48. Zhang, Y., Yao, E., Wei, H. and Zheng, K., 2017. A constrained multinomial Probit route choice model in the metro network: Formulation, estimation and application. *PLoS ONE*, *12*(6), p.e0178789.