MODELLING LONG HAUL TRUCK ROUTE CHOICE IN ONTARIO

SYED UBAID ALI

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

GRADUATE PROGRAM IN CIVIL ENGINEERING
YORK UNIVERSITY
TORONTO, ONTARIO

September 2020

Abstract

A route choice model is developed in this thesis to explain and predict long-haul truck vehicle movements in Ontario. An algorithm is devised to process approximately 58,000 observed trips from GPS data in ArcGIS to establish variable choice sets based on an optimal commonality factor that measures route overlap. Novel implications of the commonality factor add to the existing literature.

Route characteristics are next used to estimate a C-logit model. Results indicate that truck drivers are more likely to select routes exhibiting lower minimum travel times, more freeway and Highway 401 usage, more diesel stations, and fewer intersections. The travel time is the most dominant variable based on measurements of elasticity. Two scenarios are tested using the final model to determine routing changes due to increased travel time on Highway 401 and other freeways. Further detailed scenarios can be used to predict long-haul trucking patterns for future transportation planning purposes.

to my parents

Acknowledgments

I am grateful to Dr. Kevin Gingerich for inspiring this thesis and his continuous guidance throughout my MASc. program. Through his instruction and advice, I learned the art of conducting research and solving interesting problems. I also thank him for teaching me how to use ArcGIS, a software that I always feared and avoided until now. He also got me involved with the ITE York University Student Chapter, which has given me great exposure and contributed to me getting the most out my graduate program.

I would also like to thank Dr. Peter Park and Dr. Mehdi Nourinejad for their support and providing the critique necessary to ensure the quality and thoroughness of this work.

I thank Rob Reynolds for setting up my workstation with the software necessary for this research and providing immediate solutions for any issues I encountered. I am also thankful to Sindy Mahal for her timely reminders and support.

I thank my colleagues (friends) for their encouragement and help throughout my time at York University: Seun Oluwajana, Erik Nevland, Ravichandra Rampure, Crystal Wang, Ahmad Mohammadi, Soha Saiyed, MD Tanvir Chowdhury, Yashar Zarrin Zadeh, Danelle Bishoff, Elanakayon Annalingam, Ucchas Saha, Pushpendra Sharma, Nora Liu, and Khlood Azam.

I am grateful to my parents for encouraging me to enroll in this program and pursue higher level education

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

AADT	Average Annual Daily Traffic
AADTT	Average Annual Daily Truck Traffic
AHP	Analytical Hierarchy Process
ALOGIT	Software used to+ estimate Discrete Choice Models
ArcGIS	Software used for Geospatial Analysis
CD	Census Division
CF	Commonality Factor
C-Logit	Logit Model with Commonality Factor Term
GIS	Geographic Information System
GPS	Global Positioning System
HAZMAT	Hazardous Materials
IIA	Independence of Irrelevant Alternatives
Link	A portion of a route
MLE	Maximum Likelihood Estimation
MNL	Multinomial Logit
MROR	Maximum Route Overlapping Ratio
MSA	Metropolitan Statistical Area
MTO	Ministry of Transportation Ontario
NLOGIT	Software used to estimate Discrete Choice Models
OD	Origin and Destination
OD-Pairs	Origin and Destination Pairs
ORN	Ontario Road Network

Route	A path used to complete a trip between a given OD-Pair
Trip	An individual observation in the GPS dataset
RL	Recursive Logit
SQL	Structured Query Language
VMT	Vehicle Miles Travelled

List of Appendices

Chapter 1: Introduction

1.1 Freight Transportation in Canada

The truck volumes in Canada have linearly increased over the past several decades in response to the rising global population and level of international trade. A 242% growth has been observed in the number of registered freight vehicles in Ontario between 1990 and 2014 (Statistics Canada, 2018b).

There are now an estimated 112,000 trucking companies operating in Canada, with 47,500 trucking companies in Ontario alone (Statistics Canada, 2018a). In 2015, the mode split for domestic freight movements in Canada included 72% truck, 21% rail, and 7% marine (Transport Canada, 2017).

More recently, the rapid growth of e-commerce and shifting consumer demands for time-sensitive deliveries have further increased demand for truck freight, which is the most flexible mode of travel for goods movement. As shown in Figure 1-1, retail e-commerce sales have increased in Canada between 2016 to 2019 at an average rate of almost 22% per year (Statistics Canada, 2020).

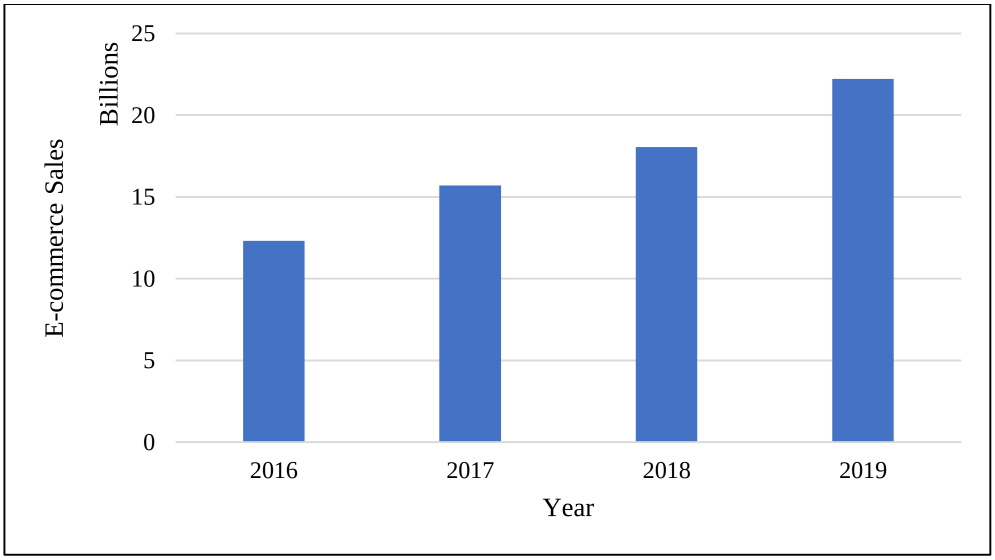


Figure 1-1: E-Commerce Retail Sales in Canada

Source: Adapted from (Statistics Canada, 2020)

The primary corridors for truck movements in Ontario are well known to the transportation industry. For example, most trucks in Southern Ontario will use major roadways such as the provincial highways and 400-series freeways administered by the Ontario Ministry of Transportation (MTO). In Northern Ontario, the trucking industry heavily relies on the federally designated Trans-Canada Highway.

Historical values of the annual average daily truck traffic (AADTT) catalogued by the MTO provide further details on the pattern of Ontario truck movements. Their iCorridor online open source database (MTO iCorridor, 2019) includes truck volume data from 1988 to 2016. Figure 1-2 shows the increasing truck volume trend on major Ontario provincial highways during this time period. The overall AADTT on Provincial Highways steadily increased at a rate of approximately 2% per year. As previously stated, this observed growth can be attributed to the growing Canadian economy and population along with the increased consumer demands of e-commerce shipping from around the world.

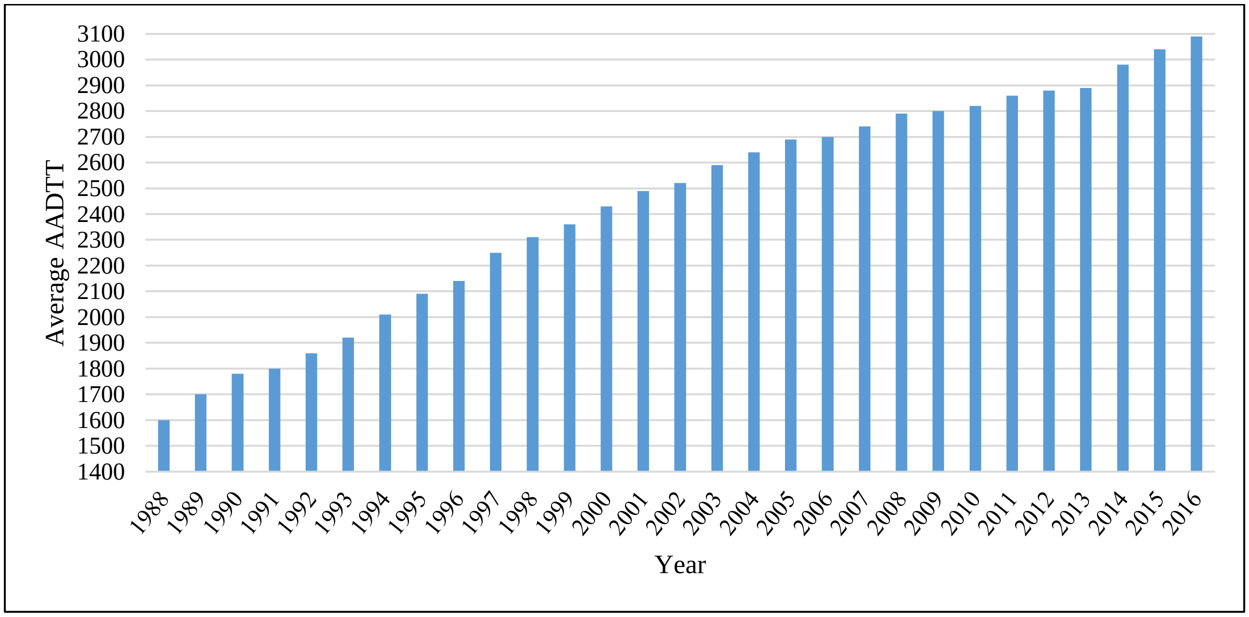


Figure 1-2: Historical Average Annual Daily Truck Traffic on Ontario Provincial Highways
Source: (MTO iCorridor, 2019)

Figure 1-3 shows a visual comparison of provincial highways between 1988 and 2016. The spatial pattern remains relatively consistent while the overall volume of trucks on the provincial road network has increased. For example, in both 1988 and 2016, the 400-series freeways exhibited heavier utilization compared to other roadways in Southern Ontario. In addition, the Trans-Canada Highway observed more AADTT in Northern Ontario compared to other alternative highways to connect freight between the northern towns and cities. In both time periods, the freight corridors in Southern Ontario exhibit greater traffic volumes proportionally compared to their counterparts in Northern Ontario. This contrast in AADTT is explained by the higher concentration of population and industry in Southern Ontario.

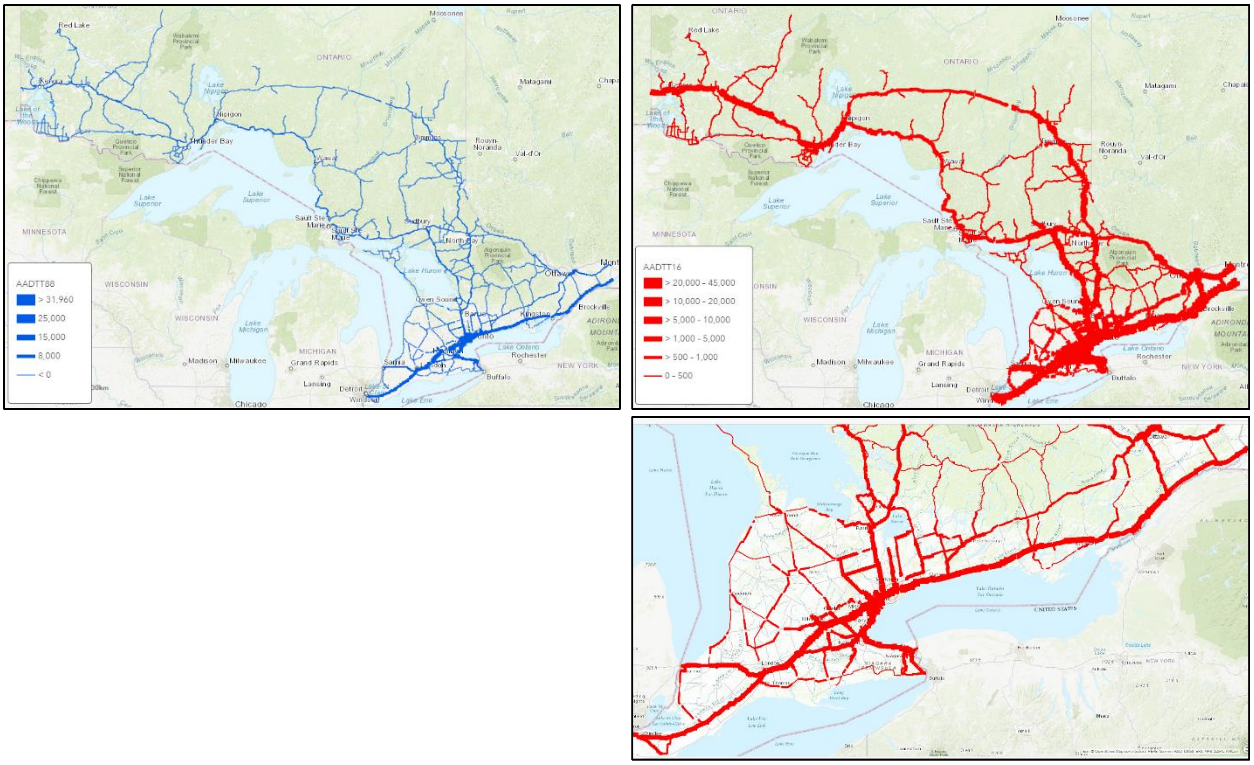


Figure 1-3: Average Annual Daily Truck Traffic on Provincial Highways in 1988 and 2016
Source: (MTO iCorridor, 2019)

The steady and visible increase in truck traffic volumes across Ontario’s road network raises several transportation planning implications. These concerns relate to changes in travel time and network equilibrium which subsequently incur changes to route choice. Aggregate patterns of routing decisions influence infrastructure usage and deterioration, traffic congestion, and availability of viable alternatives between major freight centres. It is therefore necessary to understand the factors that give rise to existing truck routing decisions in order to forecast future trucking patterns and effectively plan for future infrastructure.

1.2 Research Goal

It is necessary to plan for future freight transportation patterns on roadway facilities given the general growth trend of the freight industry along provincial highways. The existing patterns of long haul truck movements can be utilized to develop route choice models capable of: (i) explaining the factors giving rise to route choice decisions; and (ii) predicting future truck patterns given expected changes in truck volumes and infrastructure investments. This type of research can therefore be used to generate data driven policies and plan freight corridors.

The primary goal of this thesis is to develop a route choice model that explains and predicts long-haul truck vehicle movements in Ontario, Canada. Observations of Canadian truck trips during a one-week period in March, 2016 were obtained for this task. ArcGIS is used to process these truck trip observations through custom algorithms and geospatial analysis to develop variables used to estimate route choice models. The final route choice model will suggest the key factors and measures that influence the decision making of truck drivers. To achieve this goal, the thesis is organized into three objectives provided below in Section 1.3.

1.3 Research Objectives

The first objective of this thesis is to undertake a literature review of existing model types of route choice and measures of route quality such as travel time, reliability, and route overlap. These attributes will be later adapted into the route choice model. Moreover, the degree of overlap (commonality factor) between various routes will be investigated since routes typically share some road segments with each other. This is necessary to differentiate unique routes and build a suitable choice set.

The second objective is to determine the existing patterns of truck trips and preferred routes in Ontario. This is achieved through a custom algorithm in ArcGIS that uses the commonality factor to establish choice sets between origins and destinations in Ontario and assigns truck trips to routes within these choice sets. A sensitivity analysis is later conducted using this algorithm to determine the impact of the commonality factor on choice set generation and subsequent model results. The results of this analysis are novel and provide new information not widely available.

The third objective is to estimate C-logit models and test their suitability using scenario analyses. The C-Logit is a type of multinomial logit model that has an extra term capturing the commonality factor. The model estimation is performed using variables identified from the literature review in objective one, and the choice sets developed in objective two. The results from this model should reveal insights into the behaviour of trucks for route choice and forecast the probability of a truck selecting each possible route on the road network as the quantity of traffic volume and quality of routes change in the future.

The above described objectives of this thesis are summarized as follows:

1. Conduct a literature review identifying current route modelling techniques and measures of route performance.
2. Devise an algorithm to generate route choice sets based on the commonality factor for various combinations of origin-destination locations. The impact of the commonality factor is explored using a sensitivity analysis of the threshold factor.

3. Process variables and estimate a route choice model for long-haul truck route choice in Ontario. Future scenarios are tested to explore the functionality of this model.

1.4 Research Scope

While the estimated route choice model is intended to explain existing route choice in Ontario and predict future impacts on the road network, this thesis is subject to the spatial, temporal, and data scopes of work as discussed below.

The truck trips analysed in this thesis are related to the province of Ontario with an emphasis on provincial highways which are predominantly used for long-haul trucks. Urban routing within a city is not considered due to the varying behaviours of such trips. In addition, trips used in this thesis extend beyond Ontario into other Canadian provinces and American states, but the estimated model is intended to explain route choice for Ontario. For example, a variable pertaining to Highway 401 would be inapplicable for other study areas.

The truck trip dataset used in this study was collected from the GPS data logging devices of 313 freight trucking companies during one week of March, 2016. While this dataset yields an adequate cross-section of Ontario's present route choice behaviour, it was processed specifically for long-haul truck trips and therefore ignores short urban travel. There are also biases in this dataset that under-represent some industries, but this is not expected to impact the results of the model since industry is not explicitly considered. The truck trip dataset included characteristics about each trip in terms of the carrier ID, trip distance, and trip time. Other specifics are further detailed in Section 3.2.1.

Other types of geospatial datasets were obtained from open sources to analyze route characteristics in ArcGIS. These include Road Network data, Gas Stations, Weigh Stations, and HERE data through ArcGIS Network Analyst. These datasets are collected and processed such that they pertain to the boundaries of the truck trip dataset.

1.5 Thesis Outline

In terms of organization and structure, the remainder of this thesis has been outlined below:

- Chapter 2: Literature review is performed on the measures and factors used in route choice modelling including the commonality factor. Modelling techniques and tools used by researchers in this field of study are also explored.
- Chapter 3: A description of the study area and prominent data types used in this thesis are provided for contextual and analytical purposes.
- Chapter 4: The truck trip data obtained for this study will be analyzed for overlaps and commonality using an algorithm developed in ArcGIS. Each trip occurring between a given origin and destination will be grouped into a unique route based on its measured overlap with other trips. The final set of unique routes becomes the choice set for a given pair of origin-destination zones.
- Chapter 5: After determining the choice set, each route is evaluated for performance by analyzing the trips assigned to those routes. Physical characteristics and facilities along routes are also determined.
- Chapter 6: Estimation of a route choice model is performed based on the choice set from Chapter 4 and variables derived in Chapter 5. NLogit software is used to estimate the route choice of long haul trucks in Ontario. This further includes a discussion of modelling methods, results, and interpretation.
- Chapter 7: A sensitivity analysis of the commonality factor is conducted. Increasing or decreasing the threshold directly affects the number of route alternatives in a choice set and

the subsequent model results. The most appropriate model based on the optimal commonality factor is then used to test the performance of the model.

- Chapter 8: Future scenarios are designed to test the optimized model and predict changes to truck route patterns based on changes to travel time on Highway 401 and other freeways.
- Chapter 9: Conclusions and recommendations are provided on the research along with a discussion of relevant route choice model applications.

Chapter 2: Literature Review

Historically, research in long haul freight route behavior has been limited compared to other activities influenced by routing. For example, passenger travel and urban freight routing analysis can be found more frequently by comparison. This literature review comprises of publications related to freight routing with emphasis on safety, route choice factors, data, and route choice modelling. Each publication is summarized herein, with commentary on important factors and measures, data sources, methods, findings, and directions for further research arising from the existing literature.

All studies listed in Appendix A provide valuable insight to inform this thesis. The literature showed how many route choice models were developed to analyze and predict routes used by vehicles and freight service providers. The models also showed which factors and route characteristics influenced the utilities of a route and their probabilities of being selected.

The objectives of this literature review are threefold. First, safety considerations are explored to provide an understanding of the need for adequate routing and the conditions required for this routing to be safe. Second, the methods used to develop route choice models will be discussed to determine the best approach for this thesis. An important consideration here is the best approach to derive a suitable choice set available to the decision maker for each observation in the route choice model. Third, the factors included in previous route choice models will be analyzed to determine appropriate model inputs to estimate a discrete choice model for long-haul truck route choice in Ontario that will help understand the current routes used by long haul truck routes and predict how truck traffic will utilize Ontario’s road network in the future. These factors are summarized in this thesis by using the categories of geometric design, route performance, and facilities.

2.1 Safety Considerations

Previous research has been conducted to understand the safety implications of truck route. Their objective was to determine the primary contributors to long haul truck related collisions and propose suggestions for selecting routes such that the probability of a truck related accident is minimized. A general measure of truck route safety is the frequency of truck collisions. This is often transformed into a rate by using an exposure variable such as average annual daily truck traffic (AADTT), or vehicle miles travelled (VMT) (Dijkstra & Drolenga, 2008).

Truck related collisions also tend to be more severe than other type of roadway accidents (Dong et al., 2013). The 2019 Ontario Road Safety Annual Report showed that collisions involving large trucks accounted for approximately 23% of all fatalities on Ontario’s roadways, which was the highest compared to other type of fatal collisions involving pedestrians (20%), drinking and driving (10%), distracted driving (16%), speeding (15%), motorcycles (10%) and unbelted occupants (7%) (Ministry of Transportation, 2019).

Using collision data, models have been developed to predict crash frequency along routes based on significant variables such as physical road features, traffic exposure, payload, and driver conditions (Golob & Regan, 2003). Collision modelling is frequently performed using regression analyses and the obtained significant variables become important elements of route safety to hopefully reduce future truck crashes (Bauer & Harwood, 1998).

2.1.1 Geometric Features

The Wright & Burnham (1985) analysis of collision data revealed that tractor-semitrailer trucks have substantially higher collision rates compared to passenger vehicles and other types of trucks. They determined substandard vertical curvature to be the most influential factor affecting truck route safety. Apronti et al. (2019) applied safety performance functions to mountain passes and similarly concluded that routes with multiple vertical curves, long grades, and sharp horizontal curves have greater safety implications for trucks.

Similarly, Daniel & Chien (2004) found vertical curvature, pavement width and the number of signalized intersections to be significant influencers of truck crashes along urban arterial roadways, and noted that higher posted speed limits on a roadway have a negative relationship with truck crash frequency. Aultman-Hall et al. (1999) defined a route safety evaluation method through developing adequacy criteria for lane width, shoulders, railway crossings, safe speed on horizontal curves, curb radii, and stopping sight distance.

Ontario's diverse road network consists of several urban and rural facilities with varying geometric features that trucks use to complete trips. By considering such physical characteristics, safety measures can be developed to evaluate and compare observed truck routes to explore the importance of route geometry on route choice.

2.1.2 Volume Characteristics

Since vehicle collisions are often a result of both risk and exposure (Dijkstra & Drolenga, 2008), the latter implies that we generally expect more frequent collisions on a given route if there is a larger volume of traffic. Numerous studies have indicated that the traffic volumes on routes and their interaction with trucks are indeed the most influential factors resulting in truck collisions (Dijkstra & Drolenga, 2008; Golob & Regan, 2003; Joshua & Garber, 1990; Wright & Burnham, 1985).

Robertson & Aultman-Hall (2001) analyzed historical truck collision data obtained from police reports to develop a truck route safety model that showed the only significant predictor of collision rates to be the Annual Average Daily Traffic (AADT). It was therefore concluded that truck routes with higher AADT should be given priority for safety improvements. Meyer & Chu (2009) showed that heavy vehicle percentages, vehicle miles travelled, and volume to capacity ratios were also influential to truck crash frequency.

A simulation study performed by Mussa (2004) found that truck lane restrictions on freeways reduce collisions caused by frequent lane changes and hence recommended truck lane restrictions along routes with high truck traffic. Similarly, Samuel et al. (2002) argued that truck operators are willing to pay for tolled truck lanes as they improve safety in terms of reduced volume exposure for the trucks and increased separation from other road users. Conversely, Meyer & Chu (2009) claimed that truck drivers avoid paying tolls, even after concluding that usage of tolled truck routes can reduce truck collision frequency by 6-8%. They also explained that by avoiding tolled routes, truck drivers increase arterial congestion and truck collisions along non-tolled routes.

Using these results and recommendations, future research in truck route planning can be conducted to study and propose restrictions or tolled truck lanes along freight corridors in order to reduce corresponding volumes of passenger vehicles. These findings can also be used in route choice modelling to define variables that account for exposure or the provision of such facilities.

2.1.3 Payload

Sometimes, the nature of a vehicle’s payload is an important factor in route choice. Abkowitz & Cheng (1988) and Harwood et al. (1993) studied route choice for Hazardous Materials (HAZMAT) and developed risk assessment models to predict direct and indirect damages of a truck collision and HAZMAT discharge. The models accounted for the affected population, weather conditions, groundwater and surface water impacts, area of influence, and collision rates. Such considerations are also made for train movements, which are planned such that impacts to communities are minimized in the event of a collision. The Lac-Megantic rail bypass project is an example of

HAZMAT route planning that followed the disaster of July, 2013 incurring substantial damages to the adjacent urban core of Lac-Megantic, Quebec (Transport Canada, 2020).

Oversized and overweight trucks are also common in the freight industry and face route selection challenges as they can incur severe deterioration to transportation infrastructure and safety concerns to the public. Y. Li et al. (2012) and proposed a procedure to identify the safest routes for oversized vehicles by considering geometric limitations including prohibitive turns, bridges with weight limits, narrow roads, and limited-height underpasses. Their algorithm determined that routes that may be operationally inefficient are preferred over the shortest routes if they accommodate the safety concerns of oversized trucks. If route security ever becomes a governing factor in route choice, the monetary value and sensitivity of transported goods should also be considered (Private Motor Truck Council of Canada, 2019).

From a policy-perspective, payload and HAZMAT would be valuable concerns to ensure truck routes safety, but have not been considered in this thesis as routes observed in the dataset have already been deemed safe and viable by the carriers utilizing them.

2.1.4 Driver Fatigue

Driver fatigue has become as increasingly popular issue in the trucking industry. To address this, the Government of Canada has established hours-of-service regulations for commercial vehicles with enforcement through electronic logging devices. Ongoing research is also being conducted to determine other appropriate mitigations for driver fatigue such as truck parking supply and informed route choice (Nevland et al., 2020). By understanding road characteristics and the route facilities, it is possible to incorporate driver fatigue to route choice and identify optimal truck routes to reduce the probability of truck collisions due to driver fatigue.

The Oron-Gilad & Ronen (2007) driver simulator study showed that fatigue can be induced by route characteristics and route experience. A route with long monotonous segments leads to boredom and subsequently increased speeds, poor lane maintenance, and inattentiveness. Similarly, a difficult route with sharp turns and speed fluctuations is physically and mentally exhausting, also encouraging speeding through frustration. Based on the study, one can incorporate the road characteristics into route choice to address fatigue. While route familiarity is intuitively a safety feature, it can also lead to boredom and inattentiveness. Therefore, it was advised that a driver should introduce curiosity and attention throughout their trips, especially along unfamiliar routes.

Ensuring adequate rest areas is one method to combat driver fatigue. However, not all routes provide rest areas frequently enough depending on the trip length. For a decision maker, it is necessary to note the supply of rest areas for truckers along a set of route options, and compare each route to determine which one provides adequate opportunities for breaks with reference to regulation driver service hours (Fleger et al., 2002). In addition, safety and amenity factors such as appropriate lighting and shower facilities need to be considered for rest stops to ensure that the locations are desirable for drivers.

2.2 Modelling Techniques

Route choice has typically been studied using a mathematical approach which will be discussed in the next sections. However, there are some studies that view route choice with the purpose of obtaining abstract or qualitative results. Although such approaches are not usually compatible with mathematical methods, they are useful for identifying factors and developing expectations of their influences.

For example, Kunchev (2017) proposed a route choice methodology using the analytical hierarchy process (AHP) based on pre-determined weighted route evaluation criteria. Each main criterion was further characterized by several sub-criteria.

Two main approaches have been identified for quantitatively modelling route choice based on a review of the literature. First, some route models are derived from hypothetical scenarios. For example, stated preference surveys are often utilized to explore situations with very limited existing observations such as the choices based on future technology. Rowell et al. (2014) used surveys to identify truck routing priorities of freight companies from the perspectives of shippers, carriers, and receivers. Item response theory and latent class analysis was used to highlight common and differing priorities among respondents.

Alternatively, the observations may exist but the routing alternatives are hypothetically estimated based on travel measure such as the fastest potential routes. The second option is to develop the model based fully on observed decisions. The latter is becoming more common as GPS-based trace data becomes more frequent. It can be observed that more recent studies have been moving towards the latter approach.

Regardless of the selected approach above, studies focusing on route choice rely on discrete choice models to explain the routing behavior. The rest of this section subsequently discusses several variations of common route choice models along with their benefits such as result stability, computational performance, and applicability for a given dataset. In addition, methods used to generate the choice set for a given model are discussed along with typical software that is used for estimating these models.

2.2.1 Discrete Choice Model Types

Discrete choice models are a class of predictive modelling with many applications. These applications include, but are not limited to, route choice (Prato, 2009) discussed in this thesis, mode choice (Hasnine & Habib, 2018), and destination choice (Morley, 1994). Discrete choice models are a mathematical approach for determining the probability that a decision maker would select an alternative from a choice set.

This probability is calculated based on the relative utility of the alternatives and the assumption that a rational decision maker selects the alternative that offers the most utility. The utility of an alternative is function of its descriptive attributes or factors (Train, 1993). This is analogous to the routing decision process, where a decision maker (truck driver), selects the route with most utility from a choice set of route alternatives. Therefore, the application of discrete choice modelling is useful for defining choice sets and determining the factors or attributes of routes that contribute to their probabilities of being selected. Various types of discrete choice models and their usage have been described in this section.

Logit models were the most popular tool for route choice selection due to their versatility and simple format. Authors from Appendix A used various forms of logit models in the interest of accuracy, computational time, and interpretability of results. The most basic type of logit model is the Multinomial Logit (MNL) model, which has been applied by authors such as Gingerich & Maoh (2019), Li & Lai (2019) and Knorring et al. (2005) to model truck route choice for their respective datasets. The general form of the MNL model is shown in Equation 2-1.

$$P_{k,i} = \frac{\exp(V_k)}{\sum_{l \in C_d} \exp(V_l)} \tag{2-1}$$

Where

- P_k is the probability of a route *k* being selected
- V_k and V_l are the utilities of routes *k* and *l* respectively
- C_d represents the possible routes between an origin and destination for a given decision maker *d*

As discussed by Cascetta et al. (1997) and Li & Lai (2019), the MNL model in its original form cannot interpret the degree of overlap between routes and leads to enlarged probabilities for routes with shared links. This is referred to as the independence of the irrelevant alternative's property (IIA). To address this, Cascetta et al. (1997) introduced the Commonality Factor (CF) to the MNL model and coined it as the C-Logit model. Equation 2-2 shows the general form of the C-Logit proposed by Cascetta et al. (1997).

$$P_k = \frac{\exp(V_k + \beta_{CF}.CF_k)}{\sum_{l \in C_i} \exp(V_l + \beta_{CF}.CF_l)} \tag{2-2}$$

Where

- P_k is the probability of a route *k* being selected
- V_k and V_l are the utilities of routes k and l, respectively
- CF_k and CF_1 are commonality factors for routes k and l, respectively
- β_{CF} is a calibrated parameter for CF

The utility function for a given route *k* is shown in Equation 2-3, and the general form of the commonality factor is shown in Equation 2-4.

$$V_k = \sum\nolimits_n (\beta_n X_{kn}) \tag{2-3}$$

Where

- V_k is the utility of route k

- *n* represents the number of explanatory variables included in the model
- X_{kn} represents the explanatory variables
- β_n are parameter coefficients belonging to input variable X_{kn}

$$CF_i = \beta_0 \log\left[\sum_j \frac{C_{ij}}{\left(C_i C_j\right)^{\gamma}}\right] \tag{2-4}$$

Where

- CF_i is the commonality factor for route i

  C_i and C_j represent the cost of links i and j respectively
- C_{ij} represents the cost of links shared by routes i and j
- β and γ are calibrated parameters
- $\sum_j$ Represents the summation of all shared portions of route

2.2.2 Simplified Commonality Factor

As shown by Cascetta et al. (2002), and Luong et al. (2018), a simplified version of the commonality factor equation based on route length can be used to study route choice by assuming β and γ as 1 and ½ respectively with a log base of 1. This results in the simpler and applicable form of the Commonality factor shown in Equation 2-5. Other forms of the commonality factor have been discussed in Prato (2009).

$$\boldsymbol{CF_i = \sum_j \frac{l_{ij}}{\sqrt{L_i L_j}}} \tag{2-5}$$

Where

- i and j are given routes in the dataset;
- Li and Lj = the lengths of routes i and j, respectively;
- lij = the length of the shared portion between the two routes.

The lower the commonality factor between two given set of routes, the more those routes can be considered as unique alternatives. A visual demonstration of this variable is shown in Chapter 4.

Other examples of the C-Logit model can be found in the works conducted by (Prato & Bekhor, 2007; Freitas, 2018; Telgen, 2010; Luong et al., 2018). There are many other types of models that address the IIA issue. Li & Lai (2019) introduced a utility correction term based on the concept of equivalent impedance to the MNL model, while Oka et al. (2017) applied the maximum route overlapping ratio (MROR) model that considered the overlap of the shortest length alternative to the observed. All authors in Appendix A showed that their modified models produced better and more consistent results than those that did not account for IIA.

The commonality factor (CF) form introduced in the Equation 2-5 provides a method for determining the choice set based quantitatively on the uniqueness of a pair of routes. This metric will become the defining parameter for unique independent and feasible routes from map-matched trip data. If two routes for a decision maker exhibit a high CF, it is reasonable to conclude that they are not unique. Conversely, pairs of routes exhibiting low CF can be considered distinct.

To further clarify, three hypothetical scenarios have been illustrated to show the commonality factor or degree of overlap between routes connecting the Region of Peel to the City of Ottawa. Figure 2-1 shows an example pair of routes having no overlap with each other and hence a commonality factor of 0%. Figure 2-2 shows two partially overlapping routes which have a commonality factor of approximately 45%. The third example, Figure 2-3, shows two routes which heavily overlap with each other and therefore have a commonality factor exceeding 90%. Sample calculations for the pairs of routes shown in each figure are provided in Appendix B.

--- Route 1 --- Route 2

Figure 2-1: Commonality factor equal to 0%

--- Route 1 --- Route 2 --- Overlap between Route 1 and Route 2

Figure 2-2: Commonality factor approximately 45%

--- Route 1 --- Route 2 --- Overlap between Route 1 and Route 2

Figure 2-3: Commonality factor greater than 90%

2.2.3 Choice Set Generation

The generation of a suitable choice set is an important aspect of the route choice decision model. This choice set includes an observation for a decision maker along with other potential alternatives that they could have feasibly selected. This is a much more difficult task for route choice compared to some other problems such as mode choice due to the sheer number of potential alternatives. For example, there may be hundreds of route possibilities confronting someone with a given origin and destination. Moreover, there may be hundreds more possibilities for the decision makers that have different origins or destinations.

The issue of IIA presented earlier indicates that some of the many potential routes may be problematic in a model given the potential overlap with other routes. Two approaches to this issue are discussed below. First, models intended to add hypothetical route alternatives to the observed choice are discussed. This is followed by a discussion on the generation of a choice set based solely on observed alternatives.

Models Creating a Hypothetical Choice Set

More complex models tend to be used for larger road networks due to the number of available alternatives. For example, Kaneko et al. (2018) developed a Recursive Logit (RL) model for the Tokyo Metropolitan Area which was used for modeling large-scale route choice behavior. As opposed to calculating the utility of a pre-established route like in an MNL model, the RL model constructs routes by iteratively joining links that offer the most utility at a given node. The overall route choice is then based on the utilities of individual links. To restrict the number of alternatives in large networks to a more feasible choice set, Kaneko et al. (2018) introduced a driver awareness term to the RL model, which was based on road classification, implying that drivers tend to be more aware of major expressways and arterials as opposed to local roads and small streets.

Oka et al. (2017) predicted travel pattern changes of freight trucks by developing a maximum route-overlapping ratio (MROR) type model suitable for the large road network of Tokyo. This type of model is different from the discrete choice models described above that are based on random utility theory. Instead, it focuses only on the route with the lowest cost or greatest utility and compares it to the route that was actually used by a vehicle in the data by observing the overlapping ratio between them. The advantage of such a model are that it can incorporate the same route choice decision factors like in other discrete choice models while evading the choice set generation process altogether, which can be problematic in large road networks.

Models Building a Choice Set from Observations

Generating a set of feasible route options between origin and destination zones on a given road network is a prerequisite for route choice analysis. If suitable observations exist, an assumption can be made that the possible choices include those previously selected.

In a route choice model, the possible options for a decision maker are supposed to be comprehensive to cover all feasible choices. All routes in a choice set should be unique and independent. Their uniqueness is measured using a commonality factor which represents degree of overlap between two given routes Cascetta et al. (1997). The general form of the commonality factor was previously shown in Equation 2-3.

Several authors from Appendix A including (Luong et al., 2018; Prato & Bekhor, 2007; Telgen, 2010; Ramming, 2002; Freitas, 2018; Li & Lai, 2019) used the commonality factor for choice set generation.

The theory behind the threshold value for the commonality factor was rarely specified. Tahlyan, et al. (2017a) claimed it the "discretion of the analyst" and arbitrarily used a value of 95% in Tahlyan et al. (2017b). Ramming (2002) and Prato & Bekhor (2007) presented the same experiment of three candidate values of the CF, 100%, 90%, and 80%, to conclude the optimal CF threshold to be 80%.

Based on the above discussion, an MNL model is preferred for study. The model will be modified to include the commonality factor and hence become a C-logit model as proposed by Cascetta et al. (1997) and used by other studies such as (Prato & Bekhor, 2007; Freitas 2018; Telgen 2010).

Given the large range of CF thresholds, this thesis will initially apply a middle-ground value of 85% and later conduct a sensitivity analysis to determine the best CF value. This sensitivity analysis will explore the claim made by Cascetta that the C-logit model performs better when used with a limited number of alternatives (Telgen, 2010).

This novel analysis will help future studies understand the impact of the selected threshold value for choice set generation.

2.2.4 Software

ArcGIS was the most commonly used software to visualize and manipulate observed truck trip data due to the spatial tools available. The Model Builder feature in ArcGIS is especially useful for creating custom algorithms to perform repetitive tasks efficiently for large datasets as the ones (Dalumpines & Scott 2011; Gingerich & Maoh 2019; Tahlyan et al. 2017a) worked with to conduct map-matching and data processing and visualization for their trip datasets.

Cascetta et al. (2002) estimated their route choice model using the ALogit software while Gingerich & Maoh (2019) estimated their multinomial logit model in the popular NLogit software. Both software offer maximum likelihood estimation (MLE) techniques to determine parameter values that provided best fit. Prato & Bekhor (2007) estimated C-logit and other types of models using Biogeme software and Gauss matrix programming language, Tahlyan et al. (2017a) used Python programming language, and Oka et al. (2017) utilized FOTRAN language for applying the MROR model. Rowell et al. (2014) used statistical software R to identify items that distinguished between their survey respondents.

This thesis will subsequently develop the route choice model primarily using ArcGIS and NLogit software.

2.3 Truck Route Choice Factors

Decision makers, such as truck drivers, have varying priorities and preferences that need to be considered in any route choice model. The primary objective of the route choice models discussed in the previous section is to estimate the factors giving rise to the route selection process of a driver by assuming that they are rational decision makers intending to maximize the utility of their choice. This utility is determined in a discrete choice model from a set of observable factors and the additional assumption of an error distribution representing unobservable factors.

A review of freight route choice publications revealed four recurring themes under which routes were assessed for utility. Routes were evaluated and selected based their geographic features, their performance, and the facilities they offer. These categories mirror the safety categories discussed earlier in the literature review. The factors and measures relating to these themes are further described below.

2.3.1 Geometric Design

Route length is a commonly studied factor when discussing the physical description of routes (Dalumpines & Scott, 2011; Kaneko et al., 2018; Li & Lai, 2019; Oka et al., 2017). The length or trucking distance of a route has usually been a preferred metric over travel time since it is more consistent and easier to obtain than travel-time, a highly correlated equivalent. All studies showed route length to have a negative influence on route choice. Kaneko et al. (2018) and Rowell et al. (2014) noted the width of links in a route as another positive factor, usually quantified by the number of lanes in the travel direction. Another physical feature observed in routes was the number of left and right turns throughout the route (Dalumpines & Scott, 2011; Kaneko et al., 2018; Kunchev, 2017).

Li & Lai (2019) showed arterial road ratio to be a positive variable in route choice, while the number of unsignalized and signalized intersections were negative variables. Kunchev (2017) also considered passage through settlements although it held lower relative significance.

Routes that offer fewer turns, flatter grades, and fewer controlled intersections tend to be more attractive to drivers (Kunchev, 2017; Knorring et al., 2005). Knorring et al. (2005) explained that driver awareness of the geographic conditions of a given route heavily influenced the perception of its utility, and consequently its probability of being selected.

2.3.2 Route Performance

Oka et al. (2017) developed a freight route choice model heavily based on the value of travel time of a given route. Most publications including (Dalumpines & Scott, 2011; Kaneko et al., 2018; Knorring et al., 2005; Rowell et al., 2014) considered travel time and perceived speeds in an observed route as the main measures of route performance. Several authors quantified route performance in terms of observed congestion (Gingerich & Maoh 2019; Knorring et al., 2005; Kunchev, 2017; Oka et al., 2017). Oka et al. (2017) used dummy variables to designate roadways as congested based on observed average speeds from GPS data (40km/h on expressways, and 20 km/h on general roads). While speed has not been explicitly analyzed in the literature, other metrics such as road classification and number of turns have been found to be influential to route choice due to their impact on speed (Oka et al., 2017; Kunchev, 2017). Knorring et al. (2005) did however show that the perception of higher speeds along a given route makes it more attractive to drivers.

Route reliability is another metric of route performance as shown by Kunchev (2017), who captured it through considering the types of roadway (freeways, highways, and intercity arterials) based on how consistently traffic moves along them. Gingerich & Maoh (2017) described reliability in terms of the proportion of a trip consisting of no delay, expected delay, and unexpected delay beyond the averages experiences by a typical user. Popular performance metrics for roadways also include the travel time index, planning time index, and the buffer time index (Russell, 2014). Each index represents additional travel time as a portion of the minimum trip time to account for delays when planning a trip. These indexes are further explained in Section 5.2.

Route reliability was a heavily weighted criteria expressed in terms of roadway facilities according to Kunchev (2017) giving positive influence to route choice. However, Wang & Goodchild (2014) showed that route reliability was highly correlated with travel time and was only significant and positive in the absence of travel time in their estimated model. Therefore, in route choice modelling, it may be valuable to initially consider both travel time and route reliability but retain only one of them in the final model.

2.3.3 Facilities

Available facilities along a given route also influence the perceived utility for a route (Kunchev, 2017). Kunchev (2017) and Rowell et al. (2014) listed the number of rest areas, number of refueling stations, and adequate truck parking as the most desired facilities along routes for trucks. Cargo size and weight regulations were another recurring criterion for truck companies according to Knorring et al. (2005) and Rowell et al. (2014). Therefore, certain routes are not feasible alternatives for trucks depending on the vehicle size and cargo. Knorring et al. (2005) also considered hazards avoided in a given route and the availability of support in case of problems. Environment and scenery were non critical but noted features of routes.

Knorring et al. (2005) concluded that truck drivers are primarily time minimizers and estimate trip duration based on past experience, time of day, current traffic conditions, and knowledge of the route. Rowell et al. (2014) found that truck routing factors are not constant but depend on trip length. For example, while the general route objectives may be the same for long haul and short haul trucking companies, long haul carriers give more importance to refueling locations, parking availability, and size and weight limits, while parcel delivery companies are more concerned with service hours. Appendix A shows how factors affecting route choice vary across different study groups, road networks, jurisdictions, data, and research methods, confirming the theory of how truck route priorities vary with the nature of the study group (Rowell et al., 2014).

The most significant factors discussed above will be applied in this thesis to create a suitable model for the province of Ontario.

Chapter 3: Study Area and Data

The Province of Ontario has been selected as the study area for this thesis with long haul trips either starting or finishing within its boundary. Ontario is Canada's most populous province and contains almost half of the country's registered trucking companies. Ontario also includes numerous major commercial and industrial cities such as Toronto, Hamilton, Niagara, Brampton, London, Milton, Thunder Bay, and Waterloo, all of which are major truck trip generators as shown by the processed truck trip data described in Section 3.3.

The top five highways by AADTT in 2016 are provided in Table 3-1 along with the segment location where these volumes have been recorded. While Ontario is a geographically large province in Canada with a diverse highway network, its highest truck traffic volume has been observed in the municipalities surrounding Toronto's Pearson International Airport. This area located in the Region of Peel is a central hub for the Ontario freight industry. In addition to serving local economic demands, the 400-series highways in Southern Ontario also serve as major trade routes between Canada and the U.S. by connecting Canadian cities such as Montreal, Toronto, Brampton, Sarnia, and Windsor to nearby States including Michigan, Illinois, Indiana, New York, and Pennsylvania.

Table 3-1: Top 5 Provincial Highways in Ontario by AADTT volume

Source: Adapted from (MTO iCorridor, 2019)

Rank	Highway	City	2016 AADTT
1	401	Toronto	41,650
2	427	Toronto	27,552
3	410	Mississauga/Brampton	21,480
4	QEW	Toronto/Mississauga	20,292
5	400	Toronto	17,584

The presence of substantial trucking activities in Ontario makes it a suitable candidate study area for this thesis since this leads to abundant data. The large number of major roadways on the Ontario's road network also helps the quality of data and scales this thesis to a manageable and meaningful area of analysis for long-haul truck trips. Selecting a larger area like Canada is feasible but presents a heterogeneous pattern of road networks from coast to coast. For example, authorities such as the Ministry of Transportation (MTO) are directly responsible for major roadways at the provincial level. Conducting a study for Canada itself would additionally be computationally cumbersome and was therefore not selected.

3.1 Data Requirements

3.1.1 GPS Data

Developing a useful route choice model often requires large and detailed datasets on infrastructure and decision maker behavior. Large datasets help improve the statistical confidence in modelled results. In the case of route choice modelling, this means that there is more confidence in the explanatory variables that predict the likelihood of a truck selecting a given route. Moreover, having a large number of observations provides suitable information on the choice set available to a decision maker. For example, this thesis identifies all possible routes observed between two cities. Without this information, the alternatives must be estimated by alternative sources of data such as surveys or theoretical estimates (Rowell et al., 2014). More information on these methods can be found in the literature review provided in Chapter 2.

The convenience of global navigation satellite systems has made it possible for more recent studies to utilize big-data to study freight transportation behavior (Dalumpines & Scott, 2011; Gingerich & Maoh, 2019; Luong et al., 2018; Oka et al., 2017). GPS technology provides individual points containing time and location information but can also lead to erroneous data as it is heavily reliant on available satellite connections, device functionality, and vehicle operations. Additionally, this information typically needs to be translated from unconnected points in space into a more useful form such as polylines for each defined trip along a specified route with appropriate end points. Hence, large datasets often require substantial processing before meaningful information can be extracted.

Appendix A demonstrates many studies utilizing large datasets from GPS pings tracking observed vehicles. For example, Dalumpines & Scott (2011) used a data set containing 47 million GPS points collected over 2 days with a spatial resolution of 10 meters, Gingerich & Maoh (2019) used a dataset consisting of 100 million GPS pings collected for a one-month period. Luong et al. (2018) obtained a dataset of 145 million GPS pings over 4 months and 96 million GPS pings over 2 months for long haul and short haul truck trips respectively.

GPS pings typically provide time and location information along with a vehicle ID, but many datasets also include other data. For example, the GPS data provided by Shaw Tracking used by Gingerich (2017) also included an anonymized carrier ID. GPS Data may also include variables pertaining to speed and data accuracy. These datasets are obtained either directly from GPS service providers, Gingerich & Maoh (2019), and freight companies Oka et al. (2017), or provided indirectly through governmental organizations (Dalumpines & Scott, 2011).

The successful application of large GPS datasets in prior route choice modelling research works by (J. Li & Lai, 2019; Luong et al., 2018; Oka et al., 2017; Gingerich & Maoh, 2019; Dalumpines & Scott, 2011; Sobhani et al., 2019; Tahlyan et al., 2017a; Jan et al., 2000; Ben-Akiva et al., 2015; Wang & Goodchild, 2014) and the availability of recent truck data for the province of Ontario make GPS data the primary input for observations used in the route choice model developed in this thesis.

3.1.2 Road Network Data

Another necessary input for developing a route choice model is the road network dataset. These are typically GIS-based files that include information on the features of a road network. A road network dataset comprises of links that each represent a segment of road. The attributes of each road link are provided in the dataset, indicating important characteristics such as link length, speed, width, capacity, intersection control and roadside features (Dalumpines & Scott, 2011; Oka et al., 2017). If intersection attributes are required, such as the type of control features used at the given location, they can be represented using point-based representation. Dalumpines & Scott (2011) obtained the road network dataset from DMTI Spatial Inc. that provides a detailed road and highway network for Canada while Oka et al. (2017) used data from the 2013 Tokyo Metropolitan Area Freight survey, road network data from an on-board unit manufacturer in Tokyo. Kaneko et al. (2018) obtained a digital road map for Tokyo from the Japan Digital Road Map Association. Luong et al. (2018) used road network data obtained from the Florida Department of Transportation.

3.1.3 GPS Data Map-Matching

GPS data originally exists as a point over space that is unaffiliated with any representation of the road network. These pings belonging to a trip must be superimposed onto the digital road network data set through a method called “Map-Matching”. Dalumpines & Scott (2011) developed a map-matching algorithm for route choice modeling using Python scripting and the built-in GIS Network Analyst Tool using ESRI’s ArcMap software. An example of the original pings and map-matched trip is illustrated in Figure 3-1.

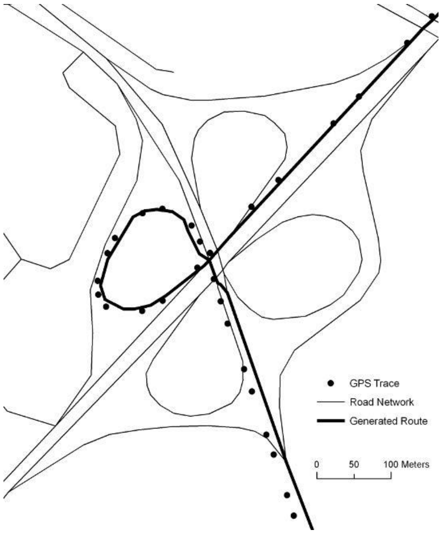


Figure 3-1: Map-Matching Process

Source: (Dalumpines & Scott, 2011)

While map-matching has also been used in many studies (Oka et al., 2017; Knorring et al., 2005; Luong et al., 2018), this thesis utilizes data originally processed and map-matched by Gingerich (2017). The latter process was performed using the algorithm and script developed by Dalumpines & Scott (2011). This script creates a buffer connecting GPS pings for a given trip and assigns the given route within this buffer. The previously map-matched GPS data is applied in this thesis to trace the routes taken by vehicles relative to a road network map and transfer truck trips into a set of observations that can be used as inputs in a route choice model.

3.2 Thesis Data

Based on the information from the previous section, the two key data types necessary for estimating a route choice model include the road network data and truck trip observations map-matched onto this network.

To add appropriate variables to a route choice model, the road network data should include information such as the routes length, number of lanes, speed limit, and facilities along major roadways. The trip data should typically include details of each trip made by a truck such as the origin and destination, trip direction, overall distance, travel time, and departure and arrival times.

The map-matched GPS data represent individual trips observed for long-haul trucks. These trips are discussed in the next section. In addition, the methods used to convert these trips into a choice set of routes is presented later in Chapter 4, followed by an analysis of these routes presented in Chapter 5.

3.2.1 Truck Trip GPS Data

Originally, GPS data of long-haul truck trips between multiple origins and destinations in Ontario were obtained from Shaw Tracking (now Omnitracs LLC) in the form of GPS pings from their trucks' onboard GPS devices. These GPS pings were processed into trips and map-matched onto the road network of Ontario while retaining some measured attributes of the completed trips. Each trip used in this thesis includes the following information: route name; anonymized truck ID; anonymized carrier ID; origin latitude; origin longitude; origin departure time; origin zone; destination latitude; destination longitude; destination arrival time; destination zone; trip duration (hours); and trip distance (km).

The original dataset processed by Gingerich (2017) and used as a starting point for this thesis includes 58,325 truck trips occurring between 3,926 unique origin-destination (OD) pairs over the span of one week in March 2016. These selected trips have been filtered such that they all either initiate or conclude the trip within Ontario. Zones in this thesis are based on the census divisions in Canada and census metropolitan areas or counties in the U.S. Within this dataset there are 13,700 trips that originated or concluded in America.

The trip data obtained for this study is illustrated in Figure 3-2, with Ontario zones (census divisions) highlighted. From this figure, the extent of Ontario's trucking industry, can be seen from

Vancouver in the West to Halifax in the East, and Southern most border of the United States in Laredo.

Figure 3-3 shows the same map emphasizing Ontario. It is apparent that freight transportation in this context is not limited to provincial highways and freeways but also includes major arterials, and regional highways throughout the province. This is particularly true for the denser road network of southern Ontario.

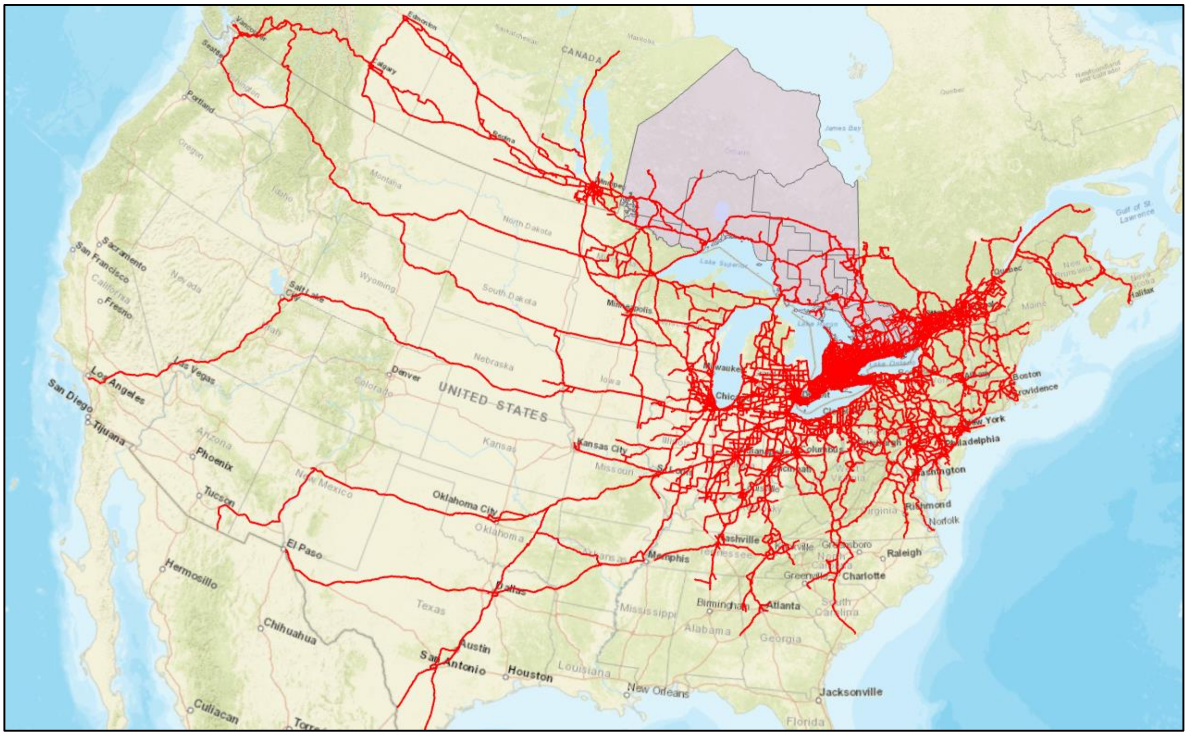


Figure 3-2: Truck Trip Data with Ontario Shown in Purple

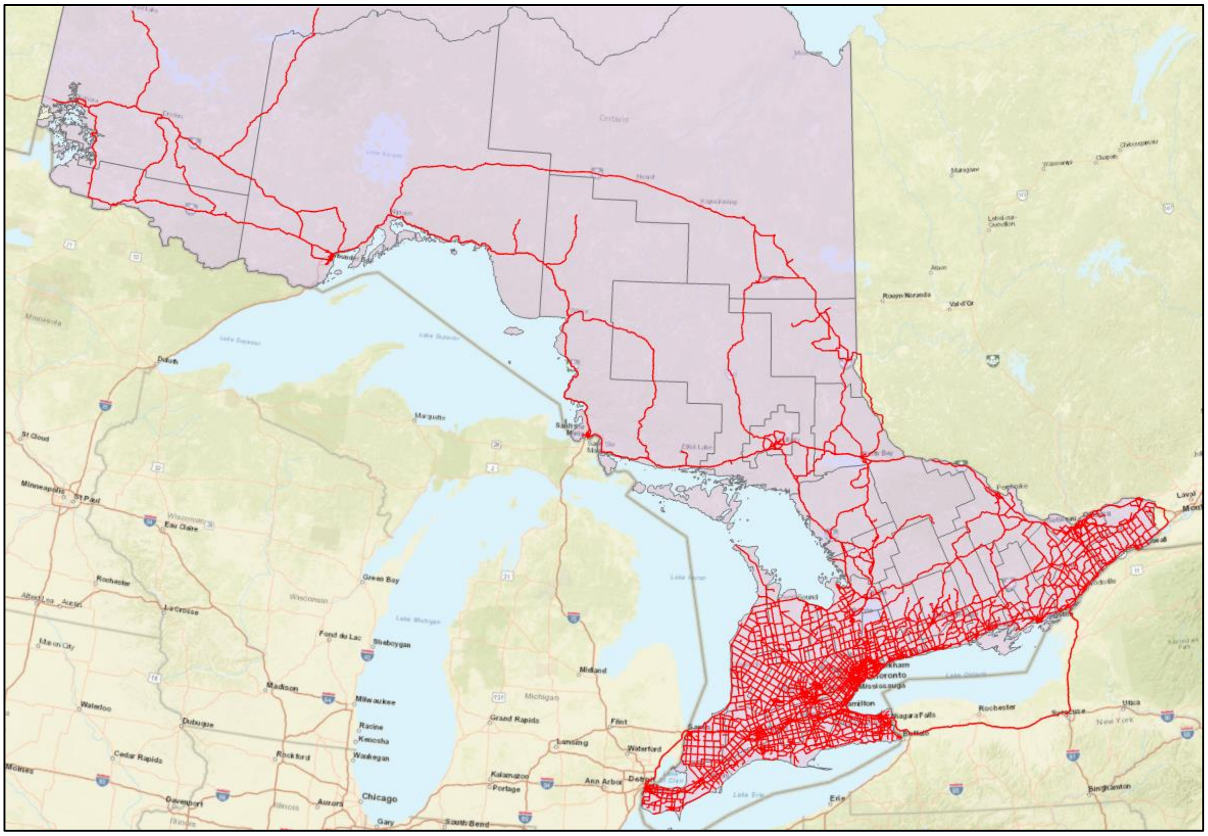


Figure 3-3: Ontario Portion of Truck Trips

To ensure a meaningful number of observations for each OD-pair, those with ten or fewer trips were removed, reducing the dataset to 840 OD-pairs. In the reduced dataset, there are a total of 50,431 recorded trips with 34,081 trips exclusive to Ontario.

3.2.2 Ontario Road Network Data

The road network data for Ontario selected in this study is the ORN Road Net Element file obtained from Ontario's Open Data Catalogue (Ontario Data, 2015). It is available as an ArcGIS GeoDatabase consisting of a shapefile with more than 250,000 kilometers of municipal roads and provincial highways. Roadway segments include information on: official and alternate street names; address information; road classification; number of lanes; road surface; speed limit; structures and underpasses; toll points; blocked passages; jurisdiction; and intersections.

As will be seen in Chapter 5, the road dataset and its tables will be used to determine physical characteristics. For example, proportions of freeways were calculated using the road names and road classification variables.

3.2.3 Additional GIS Data

Scholars Geoportal is an online web-application that provides access to large geospatial databases for researchers in Ontario (Scholars Portal, 2012). Data from DMTI Spatial Inc. was downloaded from this source for use in GIS software.

The following shapefiles have been downloaded in order to analyze the facilities along truck routes:

- Gas Stations: This dataset included information on 3,374 Gas Stations in Ontario, which included 290 stations with diesel pumps. Gas stations providing diesel are more relevant to this thesis as long haul trucks primarily use diesel fuel stations for refuelling and rest.
- Weigh Stations: 43 out of the 229 weigh stations in Canada are located in Ontario. All of these weigh stations are located along major highways and freeways in the province, or located near border crossings.
- HERE data: ArcGIS's built-in network analyst tools are internet based and rely on historical traffic data provided by HERE.com. This is the same GPS probe based data provider that Google Maps utilizes for its online real-time trip planner. This data source will be used to simulate trucking times along routes to estimate route performance metrics.

3.3 Trip Statistics by Zone

Truck trip data for Ontario-based trips has been provided for 323 different truck carriers collected over the course of a week in March 2016.

The top ten origins and destinations in terms of observed trips for the given week are presented in Table 3-2 along with some statistics to provide contextual insight on existing long haul truck trips in Ontario.

Table 3-2: Top 10 OD-Pairs in Ontario

Rank	OD-Census Divisions	OD Zone Descriptions	Observed Trips
1	CD3521_CD3530	Peel Region & Waterloo	977
2	CD2466_CD3501	Stormont Dundas and Glengarry & Montreal	697
3	CD3521_CD3526	Peel Region & Niagara	622
4	CD3512_CD3521	Peel Region & Hastings	605
5	CD3521_CD3539	Peel Region & Middlesex County	593
6	CD3518_CD3521	Peel Region & Durham Region	553
7	CD3521_CD3523	Peel Region & Wellington County	513
8	CD3558_CD3560	Thunder Bay & Kenora	505
9	CD3521_CD3524	Peel Region & Halton Region	476
10	CD3520_CD3521	Peel Region & Toronto	468

In addition to trips within Ontario, Table 3-3 summarizes long haul trips between Ontario and the US. As discussed in Section 1.1, the observed increase in truck traffic applies not only to trips occurring within in Ontario, but also those occurring between American zones.

Table 3-3: Top 10 US-Ontario OD-Pairs

Rank	OD-Census Divisions/ Metropolitan Statistical Areas	OD Zone Description	Observed Trips
1	CD3521_MSA19820	Peel Region & Detroit	376
2	CD3539_MSA19820	Middlesex County & Detroit	318
3	CD3523_MSA19820	Wellington County & Detroit	275
4	CD3521_MSA15380	Peel Region & Buffalo	268
5	CD3532_MSA19820	Oxford County & Detroit	215
6	CD3537_MSA19820	Essex County & Detroit	182
7	CD3530_MSA19820	Waterloo & Detroit	158
8	CD3521_MSA16980	Peel Region & Chicago	150
9	CD3526_MSA40380	Niagara & Rochester	103
10	CD3539_MSA12980	Middlesex County & Battlecreek	103

Many origin and destination zone include multiple route alternatives through which these trips between them have been completed. However, observed routes often exhibit overlap and might not be considered unique. This property will be discussed in the next Chapter since it has influence on the choice set created for a given OD-pair.

Chapter 4: Choice Set Generation

When studying the selection of routes between a given origin and destination, it is necessary to identify all feasible route options available. This list of feasible routes is referred to as the route choice set between an origin and destination. The routes in the choice set should be feasible, exhaustive, and independent from each other.

Using the map-matched truck trip data obtained for each pair of origins and destinations throughout Ontario, the choice set can be identified in GIS. However, defining the Choice Set is challenging since there is some degree of overlap with each potential path. In other words, how do we classify a route to be unique if there exists an alternative route sharing some segments? The level of uniqueness is explored in the next section using the commonality factor measure introduced in Section 2.2.2.

4.1 Choice Set Generation Algorithm

The GPS data includes individually observed truck **trips**. Trips that follow the same path are identified as a unique **route** occurring between a given origin and destination zone pair (**OD-Pair**).The set of unique routes between a given OD-Pair represents the route choice set used in later modelling.

An algorithm was created using the ArcGIS Pro Model-Builder tool to calculate the commonality factor between observed trips for a given origin and destination zone, reduce all trips to a set of relatively unique routes forming the choice set, and allocate each observed trip to the most similar route. This model is then iterated to repeat the process for each origin-destination pair.

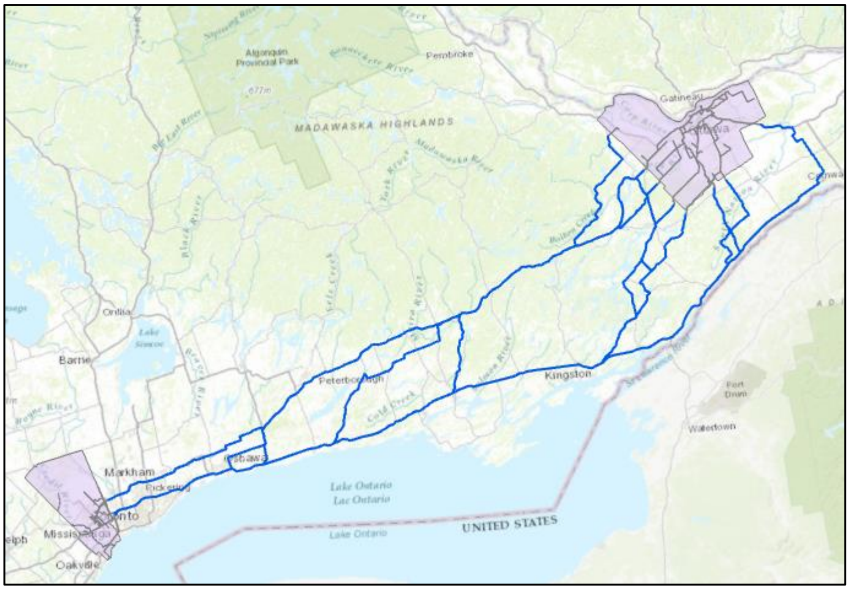
The algorithm is comprised of numerous built-in ArcGIS data analysis tools along with customization using basic SQL and Python scripting. The inputs necessary for the algorithm include (i) a shapefile containing map-matched truck trip data, and (ii) a user defined threshold for the commonality factor. The latter threshold determines the level of uniqueness for the final choice set. From the literature review in Section 2.2.3, authors have used varying thresholds for CFs to define unique routes ranging from 80% to 95%. In this study we initially use a middle-ground CF threshold of 85% as the determinant of uniqueness and later conduct a sensitivity analysis to determine the best CF value.

Each step of the algorithm is described below with accompanying figures to illustrate the processes executed in each step.

Step 1: Trimming Routes to Census Boundaries

Each observed truck trip has a specific location where it begins or ends within the boundaries of an origin or destination zone. As previously stated, the zones in Ontario used for this thesis are census divisions such as the Region of Peel. For the purposes of long haul route analysis, the polylines representing the trips are “trimmed” to the boundaries of their respective origins and destination zones to avoid analyzing intra-city, urban routing. Moreover, this trimming allows for some level of aggregation to create a choice set pertaining to routes identified for a given origin and destination zone.

To account for the portions of trips along the boundary of census divisions, each census boundary shape is buffered by 2.5 kilometers to capture and trim such trajectories to ensure only the inter-zonal portions of trips are retained. Figure 4-1 shows the portions of trips trimmed for a sample dataset.

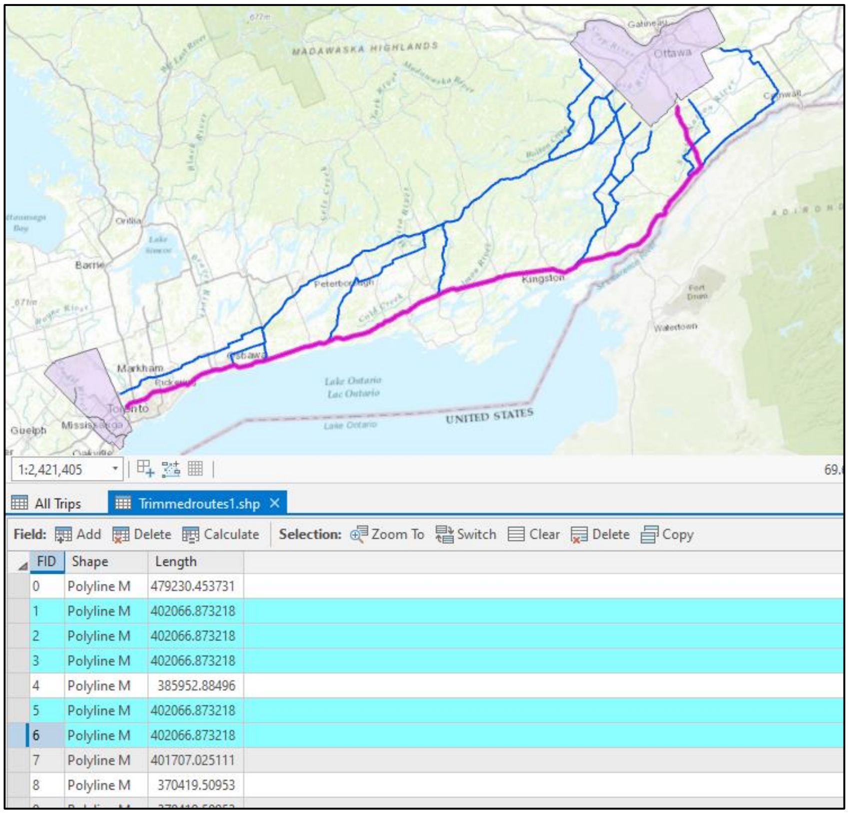


--- Trimmed (removed) trip segments --- Remaining trip segments between OD boundaries

Figure 4-1: Trips Trimmed for Processing

Step 2: Remove Identical Trips

For a given OD-Pair, there may exist several trips that have been completed using the same route as identified be the same calculated length. While this method may theoretically result in a false positive, the value is measured in meters with 6 decimal places as shown in Figure 4-2, and therefore unlikely to be a coincidence. All duplicate routes are deleted to leave behind one instance of this route. This step significantly reduces the size of the dataset to process in the interest of computing time.



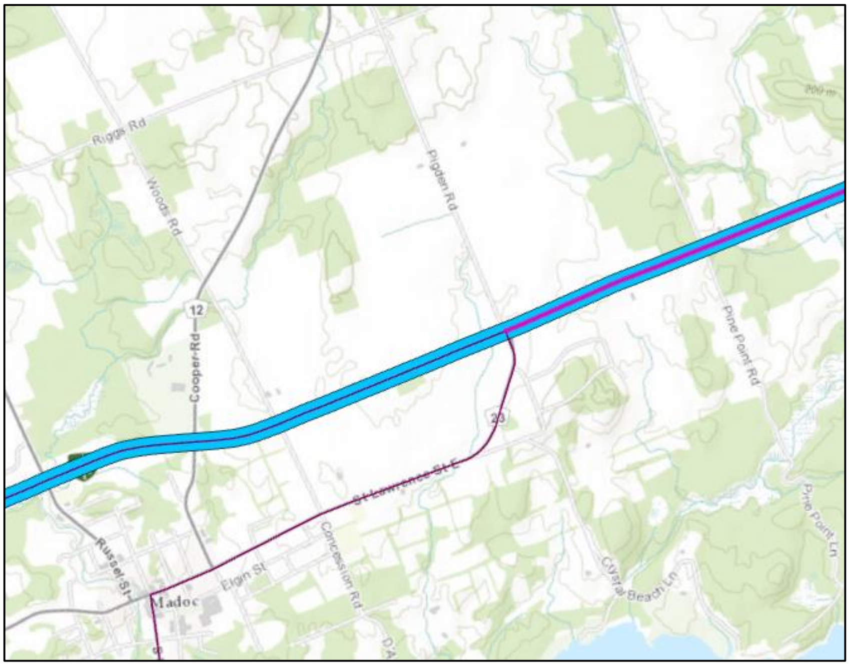
FID	Shape	Length
0	Polyline M	479230.453731
1	Polyline M	402066.873218
2	Polyline M	402066.873218
3	Polyline M	402066.873218
4	Polyline M	385952.88496
5	Polyline M	402066.873218
6	Polyline M	402066.873218
7	Polyline M	401707.025111
8	Polyline M	370419.50953

Figure 4-2: Identical trips

Step 3: Measure l_{ij}

The remaining routes from Step 2 are analyzed in pairs, denoted here as route i and route j, to determine the shared length l_{ij} between them. To calculate this length, a buffer envelope around each route in the dataset is created with a distance of 50 meters. The portions of route j that fall within the 50-meter buffer of route i are retained using the Intersect tool in ArcGIS.

The lengths of the overlapping or “intersected” routes are then measured using the calculate geometry expression: !shape.length@meters!. This represents l_{ij}, the shared portion of routes i and j. The output of this step includes the trip IDs of each overlapping route, their respective lengths L_i and L_j, and the length of the overlapping portion between them l_{ij}.

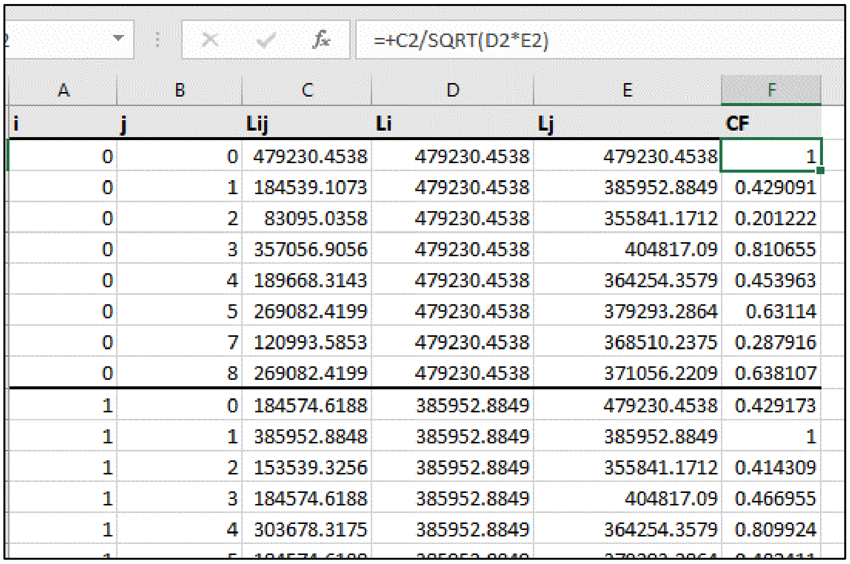


🟦 Buffer of route i --- route j --- Portion of route j falling within buffer of route i

Figure 4-3: Intersected Route

Step 4: Calculate Commonality Factor

Using Equation 2-5, the Commonality Factor is calculated for each pair of intersecting (overlapping) routes. A sample output extracted from GIS is provided in Figure 4-4 to illustrate this step. Note how CF equals 1 when i=j, indicating a 100% overlap of a route with itself.



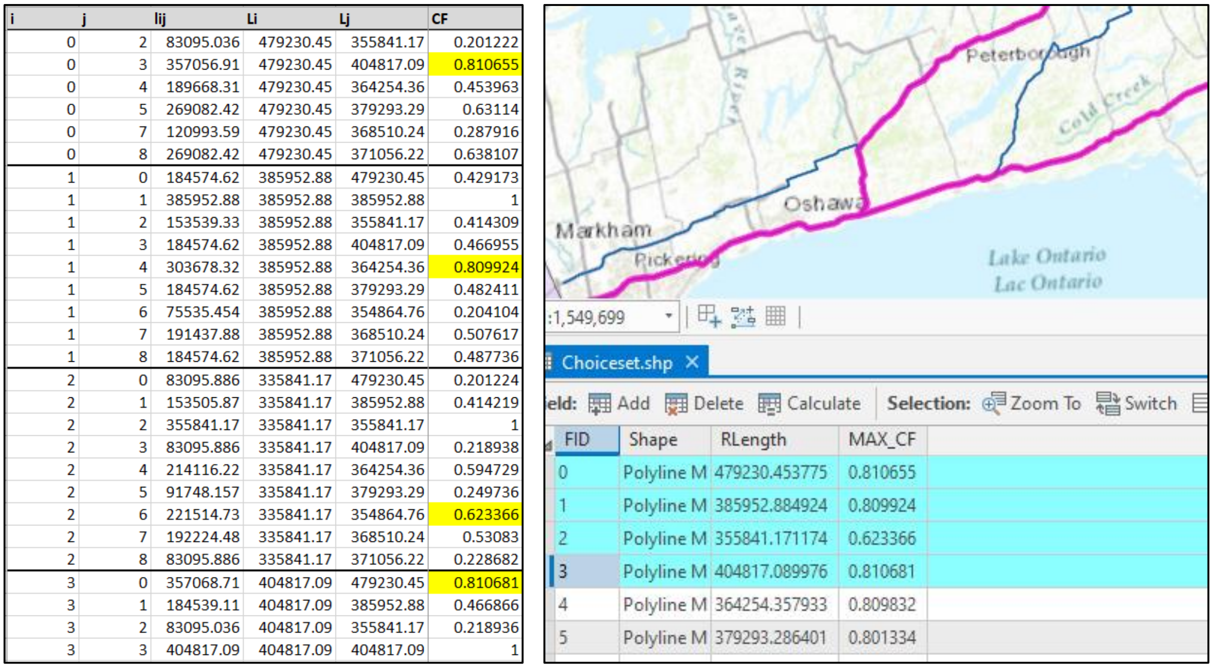
=+C2/SQRT(D2*E2)

i	j	Lij	Li	Lj	CF
0	0	479230.4538	479230.4538	479230.4538	1
0	1	184539.1073	479230.4538	385952.8849	0.429091
0	2	83095.0358	479230.4538	355841.1712	0.201222
0	3	357056.9056	479230.4538	404817.09	0.810655
0	4	189668.3143	479230.4538	364254.3579	0.453963
0	5	269082.4199	479230.4538	379293.2864	0.63114
0	7	120993.5853	479230.4538	368510.2375	0.287916
0	8	269082.4199	479230.4538	371056.2209	0.638107
1	0	184574.6188	385952.8849	479230.4538	0.429173
1	1	385952.8848	385952.8849	385952.8849	1
1	2	153539.3256	385952.8849	355841.1712	0.414309
1	3	184574.6188	385952.8849	404817.09	0.466955
1	4	303678.3175	385952.8849	364254.3579	0.809924

Figure 4-4: Commonality Factor Calculation

Step 5: Find Maximum CF

At this stage, each pair of overlapping routes for the given OD-Pair has a calculated CF. Using “Summary Statistics” function in ArcGIS, the maximum CF pertaining to each route is extracted. For instance, if the CF between route i and j is 90% and route i and k is 70%, the maximum CF for route i is 90%. A table is created to hold the CFs for all pairs of compared routes and identify the maximum CF for each route in the OD-Pair. This table is shown on the left side of Figure 4-5, where the maximum CFs for each route are highlighted in yellow, and displayed for their corresponding routes on the right.



i	j	lij	Li	Lj	CF
0	2	83095.036	479230.45	355841.17	0.201222
0	3	357056.91	479230.45	404817.09	0.810655
0	4	189668.31	479230.45	364254.36	0.453963
0	5	269082.42	479230.45	379293.29	0.63114
0	7	120993.59	479230.45	368510.24	0.287916
0	8	269082.42	479230.45	371056.22	0.638107
1	0	184574.62	385952.88	479230.45	0.429173
1	1	385952.88	385952.88	385952.88	1
1	2	153539.33	385952.88	355841.17	0.414309
1	3	184574.62	385952.88	404817.09	0.466955
1	4	303678.32	385952.88	364254.36	0.809924
1	5	184574.62	385952.88	379293.29	0.482411
1	6	75535.454	385952.88	354864.76	0.204104
1	7	191437.88	385952.88	368510.24	0.507617
1	8	184574.62	385952.88	371056.22	0.487736
2	0	83095.886	335841.17	479230.45	0.201224
2	1	153505.87	335841.17	385952.88	0.414219
2	2	355841.17	335841.17	355841.17	1
2	3	83095.886	335841.17	404817.09	0.218938
2	4	214116.22	335841.17	364254.36	0.594729
2	5	91748.157	335841.17	379293.29	0.249736
2	6	221514.73	335841.17	354864.76	0.623366
2	7	192224.48	335841.17	368510.24	0.53083
2	8	83095.886	335841.17	371056.22	0.228682
3	0	357068.71	404817.09	479230.45	0.810681
3	1	184539.11	404817.09	385952.88	0.466866
3	2	83095.036	404817.09	355841.17	0.218936
3	3	404817.09	404817.09	404817.09	1

FID	Shape	RLength	MAX_CF
0	Polyline M	479230.453775	0.810655
1	Polyline M	385952.884924	0.809924
2	Polyline M	355841.171174	0.623366
3	Polyline M	404817.089976	0.810681
4	Polyline M	364254.357933	0.809832
5	Polyline M	379293.286401	0.801334

Figure 4-5: Overall Commonality Factor of routes

Step 6: Remove Non-Unique Routes

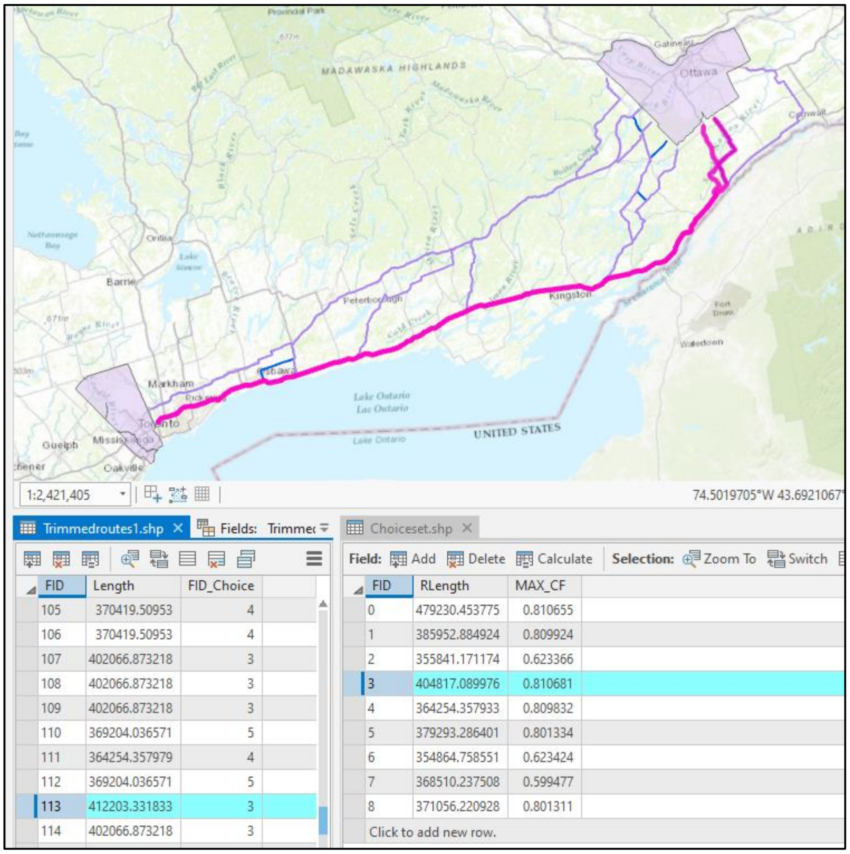
The table from Step 5 is sorted to extract the routes exhibiting a CF exceeding the threshold value of 85%. For each route i having a CF exceeding the threshold, there exists a corresponding route j with same CF value. One such route is selected at random and deleted from the choice set while the other remains.

Step 7: Iterate While Maximum CF Exceeds Threshold

The process is repeated from Step 2 to Step 6 to recalculate the maximum CF for the routes in the given OD-Pair until there remain no routes with a CF exceeding the threshold value. The remaining set of routes then becomes the route choice set between the origin and destination consisting of feasible, unique and independent route alternatives.

Step 8: Allocate Routes to Individual Trips

In this step, the trip trajectories are compared with the set of unique routes obtained from Step 7 to assign the route exhibiting the greatest overlap with a given trip. This utilizes the buffer and intersection technique discussed previously in Step 3. The route exhibiting the highest CF with a given trip is assigned to that trip. Figure 4-6 shows how trip No.113 (FID shown in the left table) chose Route No.3 (FID shown in right table) due to their similarity with each other.



FID	Length	FID_Choice
105	370419.50953	4
106	370419.50953	4
107	402066.873218	3
108	402066.873218	3
109	402066.873218	3
110	369204.036571	5
111	364254.357979	4
112	369204.036571	5
113	412203.331833	3
114	402066.873218	3

FID	RLength	MAX_CF
0	479230.453775	0.810655
1	385952.884924	0.809924
2	355841.171174	0.623366
3	404817.089976	0.810681
4	364254.357933	0.809832
5	379293.286401	0.801334
6	354864.758551	0.623424
7	368510.237508	0.599477
8	371056.220928	0.801311

Click to add new row.

Figure 4-6: Trips Assignment to Routes

Steps 1-8 of this choice set generation process are summarized by the flow-diagram illustrated in Figure 4-7. Detailed algorithms showing every step within each process are provided in Appendix C.

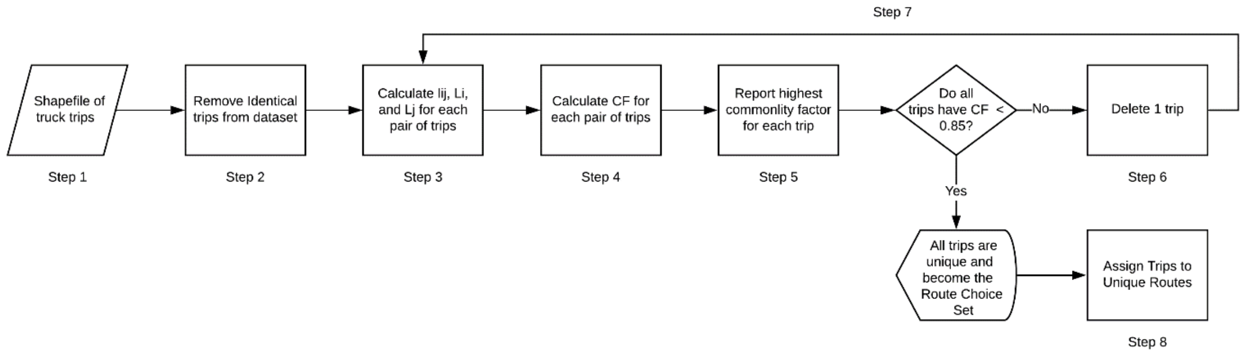


Figure 4-7: Choice Set Generation Flow Diagram

4.2 Iterate algorithm for each OD Pair

The choice set generation model discussed above in Section 4.1 is next replicated for each OD-Pair using a second algorithm developed in model builder. The “Iterate Feature Classes” iterator in ArcGIS model builder was used to implement the recursive loop in this algorithm.

A labelling system is assigned at this stage to avoid duplicate IDs for difference OD-Pairs. The origin and destination zone IDs are included in the final route label. For instance, a given route between census division 3530 (CD3530) and census division 3543 (CD3543) would be denoted as CD3530_CD3543_1. This ensures that each route belonging to a given OD-Pair is treated as a distinct option from all other routes belonging to other OD-Pairs and simplifies data input for the route choice model.

In ArcGIS, this is achieved as an outer algorithm with the previous choice set generator algorithm included as an internal geo-processing tool. Figure 4-8 illustrates this process with additional preprocessing included.

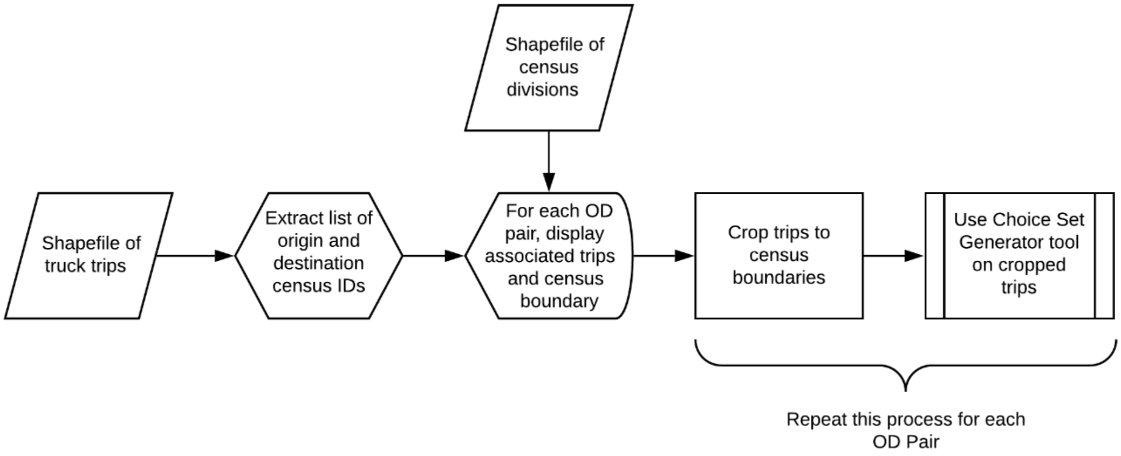


Figure 4-8: Use of Choice Set Generator for Multiple Origins and Destinations in Ontario

4.3 Computation and Model Run-time

The commonality factor algorithm developed in ArcGIS Pro was run for the 50,431 trips belonging to 840 OD-Pairs. This is completed on a computer with the following technical specifications:

- Processor: Inter® Core™ i7-7700 CPU @ 3.60GHz 3.60GHz
- Installed Memory (RAM): 8.00 GB
- System Type: 64-bit Operating System, x64-based processor
- Operating System: Windows 10 Pro

Given the size of the input database, it was expected that the processing time would be substantial. The execution of the model was carefully monitored to identify which processes or OD-Pairs are more time consuming in order to make appropriate adjustments. While OD-Pairs with more diverse route choice sets (requiring more iterations to achieve CF<0.85) were generally more time consuming, the overall run-time time was largely dependent on the computer processor, disk-space, and RAM. Each process within the algorithm originally produced a set of files that were written to the computer's hard drive. Each iteration for a given OD-pair produced a set of files, resulting in heavy consumption and eventual exhaustion of the computer’s RAM and disk-space. When the model was initially run, it was estimated that it would take more than 300 hours to complete.

To reduce the computation time, several outputs of intermediate process such as buffer, intersect and summary statistics were adjusted to be written to ArcGIS’s in_memory workspace. This method was significantly faster than writing files to the hard drive and saves disk-space on the computer. Since the computer’s RAM is being consumed through this method, the algorithm was modified to include commands to clear the in_memory workspace at the completion of each iteration to avoid leaving the computer without any available RAM. With these modifications to the algorithm, the choice set generator for all the OD-Pairs completed within a run-time of 24 hours. This is a suitable processing time for this thesis and fine-tuning was therefore concluded at this stage.

4.4 Commonality Factor Statistics on Processed Routes

The 50,431 analyzed trips for 840 OD-Pairs included 2483 unique routes. The number of routes for each OD-Pair ranged from 1 to 16. To analyze route choice diversity as a function of trip distance, a plot is provided in Figure 4-9 visualizing the number of route choice alternatives by distance. The figure indicates a decreasing trend in feasible alternatives as the trip distance increases. This is consistent with the findings of Luong et al., 2018

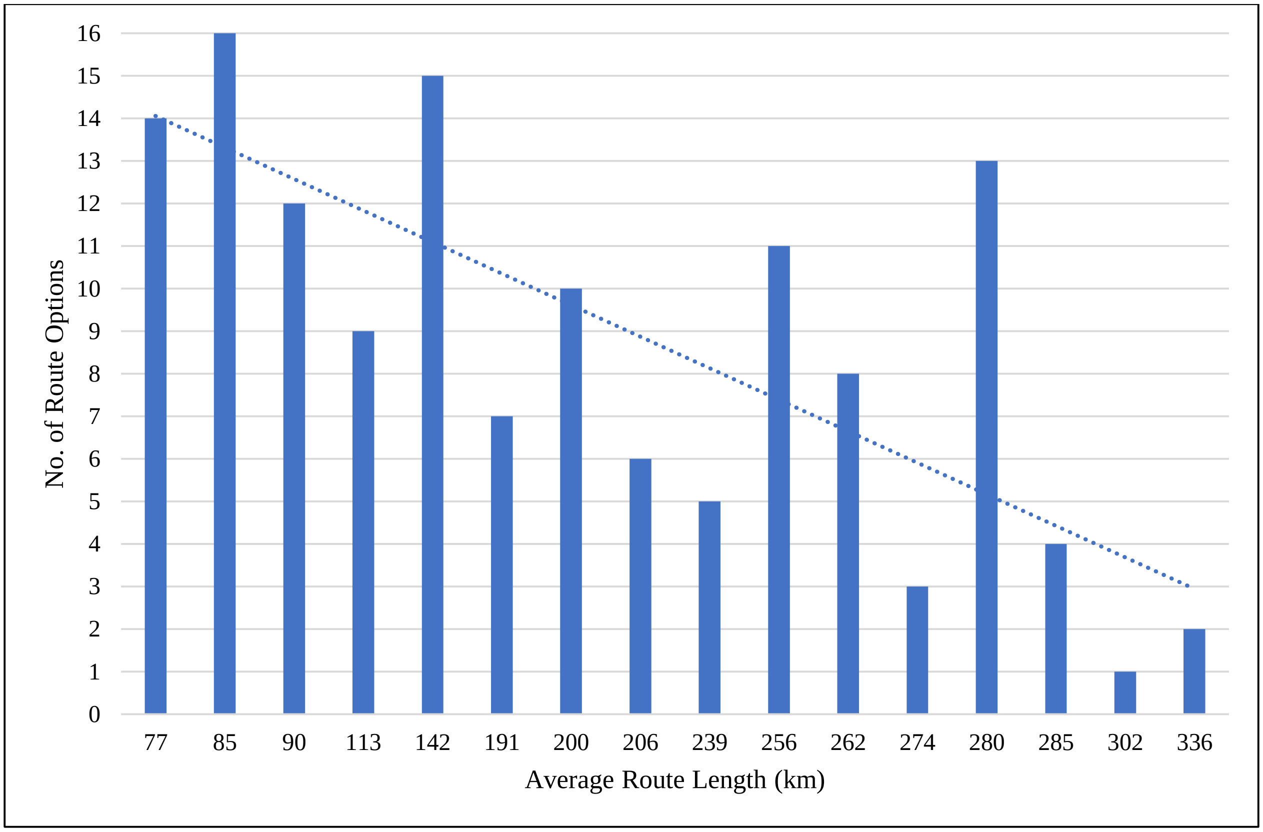


Figure 4-9: Route Choice Set Size vs. Average Trip Distance

Table 4-1 shows the choice set size and number of trips observed in each OD-Pair. There are 263 OD-Pairs in the dataset with only one route option between them. These are OD-Pairs that generated only one route option after completion of the choice set generator algorithm.

OD-Pairs with only one route account for 13,320 trips in the dataset. This effectively means that drivers operating in these ODs have route no route decisions to make since only one option is available. Therefore, these OD-Pairs can be removed for the estimation of the route choice model as they would not add any value to the results.

After removal of the aforementioned OD-Pairs, the resulting analysis dataset now comprises of 577 OD-Pairs that have two or more route alternatives between them. These OD-Pairs also account for 37,111 trips, which is considerable large dataset upon which a reliable route choice model can be estimated.

Table 4-1: Choice Sets by OD-Pair

No. of OD-Pairs	Choice Set Size	No. of Trips
263	1	13,320
216	2	8,562
131	3	7,140
78	4	5,692
47	5	3,278
34	6	3,192
21	7	3,398
18	8	2,671
8	9	394
9	10	1,254
3	11	147
5	12	420
3	13	572
2	14	221
1	15	111
1	16	59

Chapter 5: Analysis of Route Characteristics

With the alternative paths identified for each zone pair in the previous chapter, each route can be evaluated in terms of usage and performance to determine which routes are popularly chosen along with their performance measures. The observations here are used to inform the variables included in the route choice model discussed in later chapters.

The Summary Statistics tool in ArcGIS Pro is used to perform this task. All routes were measured for frequency, average travel time and average trip distance. These metrics were then used as a basis for determining other measures of route performance including average speed, travel time index, planning time index, and buffer time index.

5.1 Average Distance, Travel Time, and Speed

The shortest trip recorded along any given route is 1.18 km completed in a travel time of 0.02 hours. This was a US-Canada trip at the Niagara border. The longest trip in the dataset was 1516 km, with a duration of 65.9 hours. This trip occurred between Region of Peel and Winnipeg, Manitoba. Longer trips are included in the original dataset but have been filtered out by pre-processing steps due to the lack of sufficient sample size for a given origin-destination pair zone pair. Other statistics about these routes in the modelling dataset are provided in Table 5-1.

Table 5-1: Final Route Choice Dataset Statistics

Statistic	**Value**
No. of OD-Pairs	*577*
No. of Unique Routes	*2220*
No. of Trips	*37,111*
Minimum Number of Alternatives for OD Zone Pairs	*2*
Maximum Number of Alternatives for OD Zone Pairs	*16*
Minimum Trip Distance (km)	*1.182*
Maximum Trip Distance (km)	*1516*
Minimum Travel Time (Hrs)	*0.02*
Maximum Travel Time (Hrs)	*65.9*
Minimum Average Speed (km/h)	*22.9*
Maximum Average Speed (km/h)	*69.3*

The variance of average speeds can be used to explain the reliability of the route alternatives. Routes showing high variance or range of speeds are less reliable than routes exhibiting with lower variance of average speeds. The lowest recorded average speed of 22.9 km/h was of along the route connecting Peel Region to Winnipeg. This may be due to the variety of road facilities along this route and the time spent on driver rest and recovery. The highest recorded speed for a route was 69.3 km/h between Stormont, Dundas and Glengarry Counties and Middlesex County, which is mostly connected by freeways. Knorring et al. (2005) stated that long haul truck drivers are first- and foremost-time minimizers. Therefore, it is valuable to measure the variance of trip duration for each route. Like average speed, routes exhibiting high ranges of travel times can be considered less reliable than routes with lower ranges of travel time. For instance, a box and whisker plot for OD-Pair CD3519_CD3524 (York Region and Halton Region) provided in Figure 5-1 shows the ranges of trip travel times for each route. Routes 4, 5, and 6, do not have sufficient observation counts to develop a statistically relevant range of travel times. Route 1 is observed with a higher travel time compared to route 0, but is more reliable due to a smaller range of observed travel times.

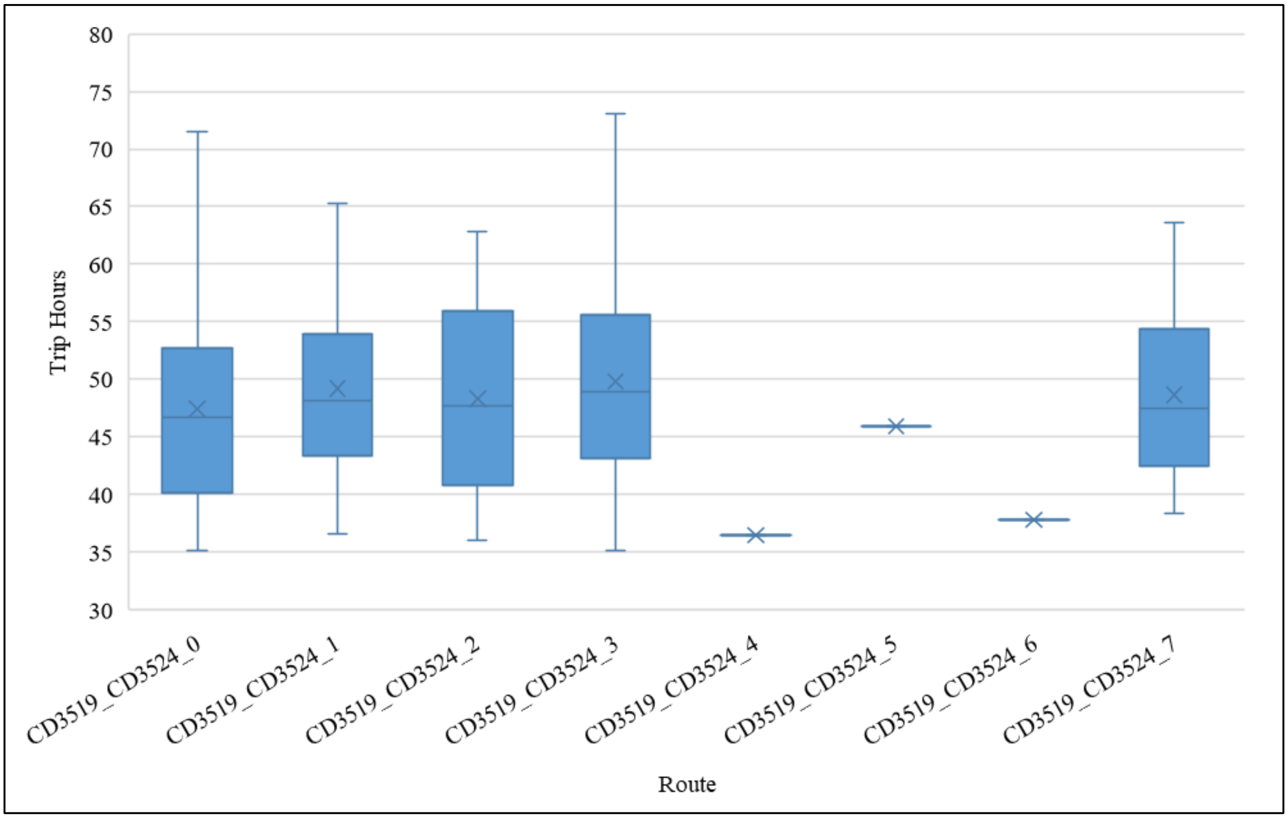


Figure 5-1: Sample Box-Whisker Plot for Travel Time Variability by Route

The total travel time given in Figure 5-1 represents time spent along major highways in free-flow conditions as well as time spent in the urban road network in stop-and-go conditions even though intra-zonal portions of the trip in the origin and destination zones have been removed. It is possible that two routes with similar trip distances can have significantly varying trip durations due to the locations of these start and stop points This issue is most problematic for zone pairs that are located near each other, but becomes less influential as the zone pairs are located further apart. For this reason, variances in travel time will be tested with this known limitation in mind while building the route choice model.

Another issue with travel times is a potential bias due to the sample size derived for each OD-Pair from the GPS data source. As an alternative, the world traffic map in ArcGIS Pro and the built-in network analyst tools were used to simulate route travel times based on traffic conditions at different times during a typical Thursday during the study period in March 2016. This network analyst relies on historical traffic data provided by HERE.com.

The 2,220 routes in the dataset were sequentially input to the network analyst and the truck travel times for each route were simulated for 6 different times of day. These times included 8:00 am, 12:00 pm, 4:00 pm, 8:00 pm, 12:00 am, and 4:00 am. These time periods were selected such that every truck route would experience at least one peak period throughout a truck journey. The simulated truck trips for each route were compared with the actual GPS recorded truck travel times along those routes. The lowest simulated travel times obtained were similar to the GPS recorded minimum travel times for most routes in the dataset, while the highest simulated travel times were substantially lower than the GPS recorded maximum travel times.

ArcGIS's network analyst is used here to determine the truck travel time between the origin and destination points for each route based on road network properties and trip departure time. To ensure that the path generated by the software matches the observed routes, additional waypoints are added at regular intervals as seen in Figure 5-2. This process was repeated for different times of day and for each route using an iterative algorithm in ArcGIS model builder.

Original Route from GPS Dataset	Route Divided into Points	Solution Route in Network Analyst

Figure 5-2: Modeling GPS Routes in Network Analyst

5.2 Route Reliability

The travel times for the routes obtained from the observed GPS data can be used to calculate reliability indexes. The three commonly used indexes are the Travel Time Index, Buffer Time Index, and Planning Time Index (Russell, 2014). The travel time index represents the ratio of average travel time caused congestion to the travel time under free flow conditions. This accounts for the delay expected to be faced along a given route.

$$Travel\ Time\ Index = \frac{Average\ travel\ time}{Freeflow\ travel\ time} \tag{5-1}$$

The Buffer Time Index represents the additional time to add to a trip to account for unexpected delay along the route compared to the average delay. The former is based on the 95% percentile (or near-worst-case) travel time but can be adjusted with other percentile values.

$$Buffer\ Time\ Index = \frac{95th\ percentile\ travel\ time\ - average\ travel\ time}{Average\ travel\ time} \tag{5-2}$$

The Planning Time Index is an additional measure of reliability and represents the total travel time to plan for a trip included both expected and unexpected delays when compared to the free flow travel time (Russell, 2014).

$$Planning\ Time\ Index = \frac{95th\ percentile\ travel\ time}{Freeflow\ travel\ time} \tag{5-3}$$

The Travel Time Index, Buffer Time Index, and Planning Time Index were calculated using the observed GPS trip data. Routes which experienced only one trip were assigned the averaged Travel Time, Buffer Time, and Planning Time indexes of the other routes in their respective choice sets. These statistics for each route are then included in the input data spreadsheet and will be tested for significance in the route choice model.

5.3 Usage of Freeways

The road network of Ontario consists of several classes of roadways. From the Ontario Road Network (ORN) Net Element Dataset, these classes include Alleyways, Local roads, Collectors, Arterials, Expressways / Highways, and Freeways. The observed GPS trips predominantly utilized arterials, highways, and freeways.

A driver’s willingness to choose a route is expected to be influenced by it composition of freeways, as this type of roadway provides sufficient geometric conditions and high speeds while also often offering freight related amenities nearby such as rest areas. In Ontario, freeways are defined as controlled access expressways with posted speed limits of 90-100 km/hr. Ontario’s freeway network includes the 400-series freeways including Highways 400, 401, 402, 403 and the Queen Elizabeth Way (451) and municipal expressways such as the Don Valley Parkway, and Red Hill Valley Parkway. The existing freeway network of Ontario provided by the ORN Net Element Dataset is shown in Figure 5-3. Note that this includes two tolled freeways, Highway 407 and Highway 412, but excludes the Trans-Canada highway as it is not a controlled access roadway which is a defining feature of freeways. A shapefile of US Interstate Freeways was also obtained from the United States Census Bureau Open Data Catalog (United States Census Bureau, 2016) and merged with the Ontario freeway network to appropriately measure the usage of freeways for US portions of each trip.

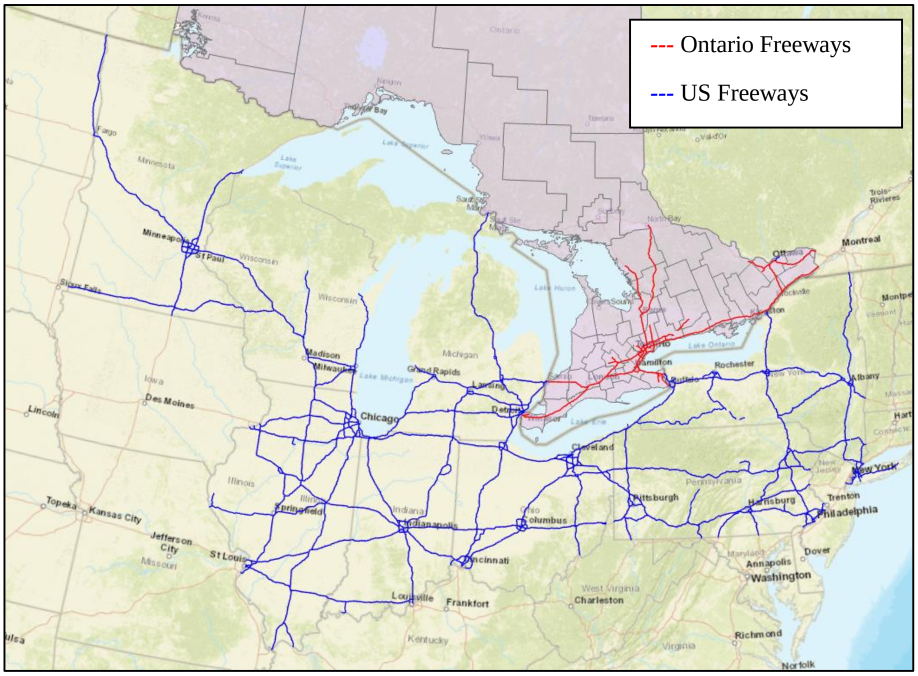


Figure 5-3: Analysis Freeway Network

ArcGIS was used to measure the proportion of freeway segments in each route. This was achieved by creating a buffer around each freeway segment and measuring the proportion of each route passing through the buffer using an intersect tool.

As an example, Figure 5-4 shows the full route CD2454_CD3521_0 travelling between Les Maskoutains and Peel Region. The bottom image shows the section of this route which overlaps with the freeway segments. The lengths of highlighted freeway portions are then used to calculate a freeway proportion (FWP) for this route which in this case was 41%, indicating that less than half of the route followed a freeway. This procedure is applied to all 2,220 routes in the choice set.

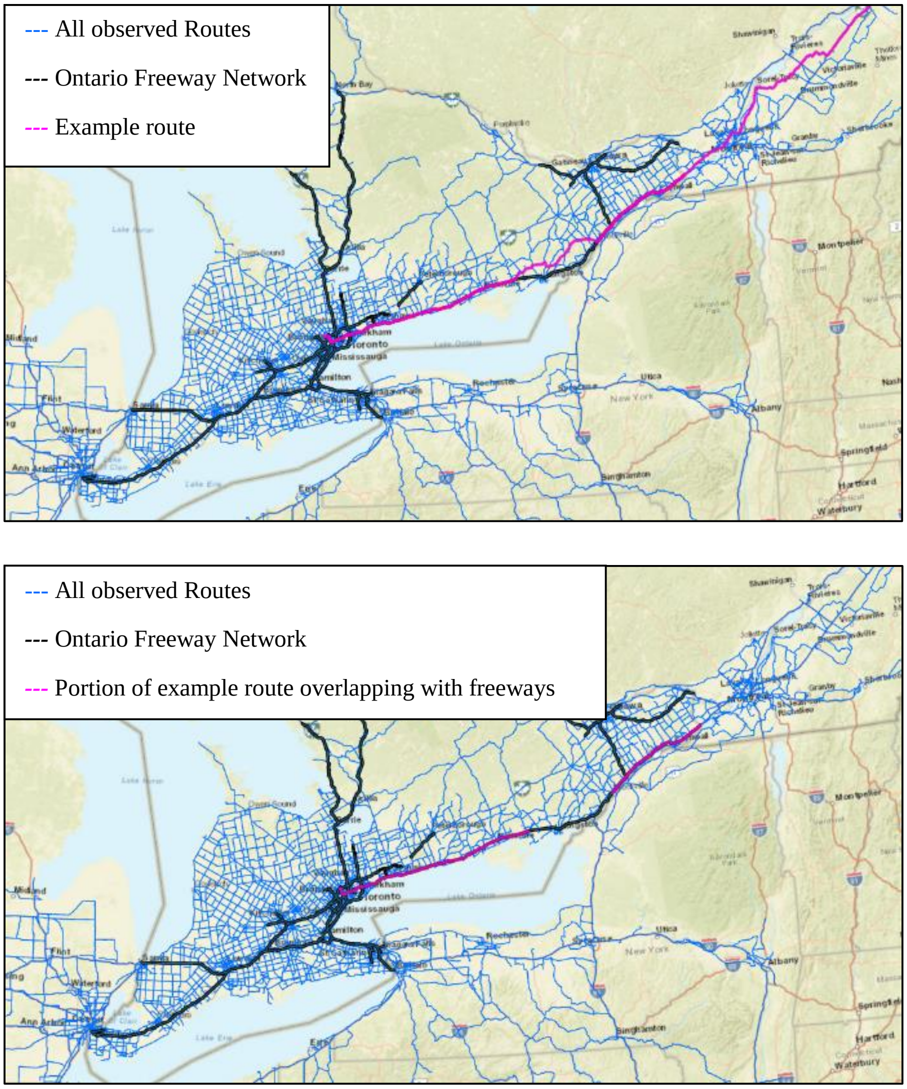


Figure 5-4: Example route between Les Maskoutains and Peel Region (top) and portion of route overlap with freeways (bottom)

In addition to the freeway proportion comprising a given trip, additional variables have also been measured to understand the usage of specific freeways in Ontario. A variable FWP401 has been created to measure the proportion of Highway 401 given that this is the busiest roadway in Canada. In the same manner, a variable FWP4XX has been created to account for the usage of other 400 series highways.

5.4 Number of Intersections

Another factor investigated here for routing decisions is the number of intersections encountered along a given route. As mentioned in the literature review, previous studies on route choice in metropolitan areas have considered the number of left and right turns needed to be made by truck drivers at intersections. As this thesis includes a higher-level study covering predominantly non-urban travel, such a microscopic analysis would be cumbersome. However, it may be useful to determine if the number of intersections encountered along a route may affect the attractiveness of a given route.

To this end, a shapefile of all intersections in Ontario shown in Figure 5-5 was extracted from the ORN Net Element Database and a count of these features was calculated for each route.

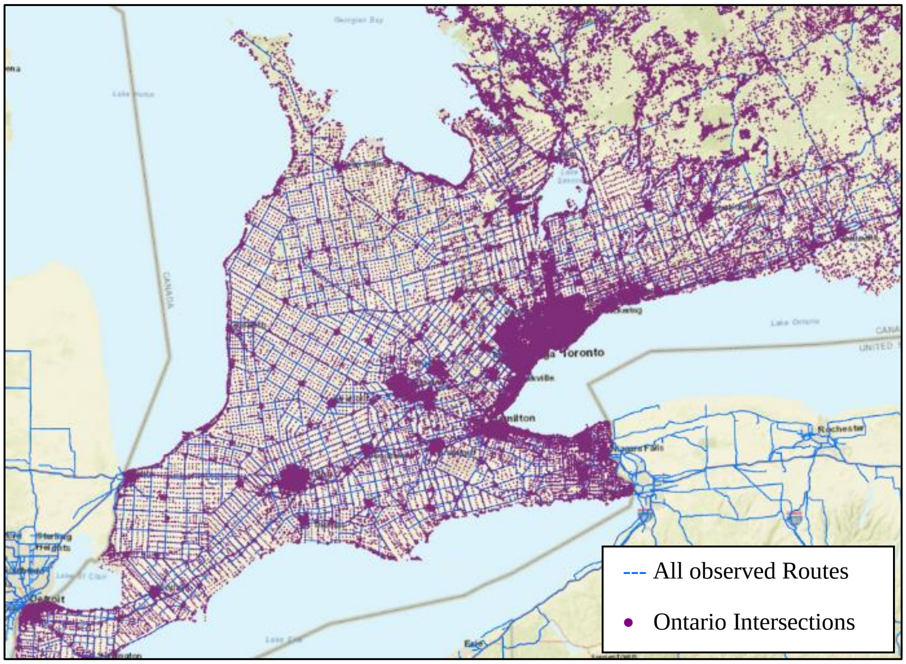


Figure 5-5: All intersections in Ontario

It is apparent that some intersections in the dataset include underpasses, highway ramps, and two-way stop controlled intersections along expressways. Such intersections are unlikely to incur delays to truck trips as they do not necessarily require any stoppage. These intersections have been identified and removed using the "select by location tool" in ArcGIS where presence along a freeway or expressway was the criterion for selection.

5.5 Usage of Tolled Roads

While there is only one tolled roadway in the Province of Ontario, Highway 407, it passes through a central region of Ontario where substantial congestion and freight activity occurs. Highway 407 spans west to east from Burlington to Clarington respectively, passing through Oakville, Mississauga, Brampton, Vaughan, Markham, Pickering, Whitby, and Oshawa. As of June 2016 and December 2019, two new north-south tolled routes namely Highway 412 and Highway 418 have opened in Durham Region. These routes however will not be analyzed in this study as the available road network data and truck GPS data is current to the year 2016. Figure 5-6 shows the tolled roadway in context of the truck routes being analyzed.

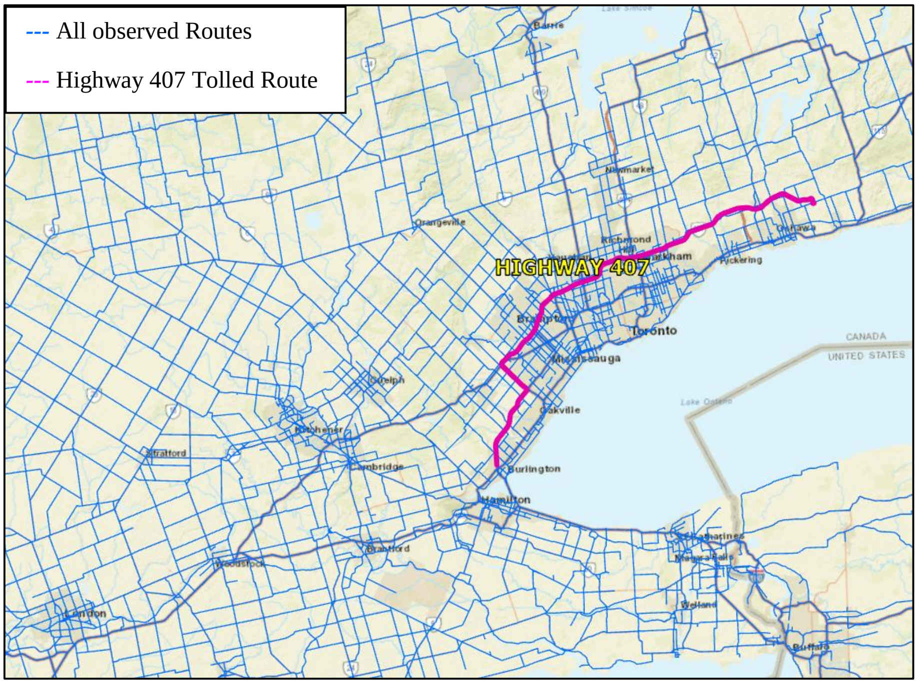


Figure 5-6: Highway 407 Express Toll Route

A variable, FWP407, is created to measure tolled route usage in terms of physical distance similarly to other FWP variables described in Section 5.3. A cost-based variable for the tolled route is not included due to the limited number observations using this route and missing information such as entry and exit times.

5.6 Gas Stations

Gas stations are an integral component to the freight industry to provide truck refueling as well as rest and recovery opportunities for drivers. Therefore, it is expected that their presence along important truck routes would become a major determinant of truck route choice. Specifically, gas stations that provide diesel fuel would have the most relevance to truck operators.

To study the influence of diesel gas stations in route choice, a shapefile containing all gas stations in Ontario created by DMTI Spatial Inc. was queried to retrieve stations that provide diesel fuel as shown in Figure 5-7. A count of these stations within a buffer of 100 m around each route is calculated. This frequency is added to the route choice model as a variable to investigate its influence on route selection.

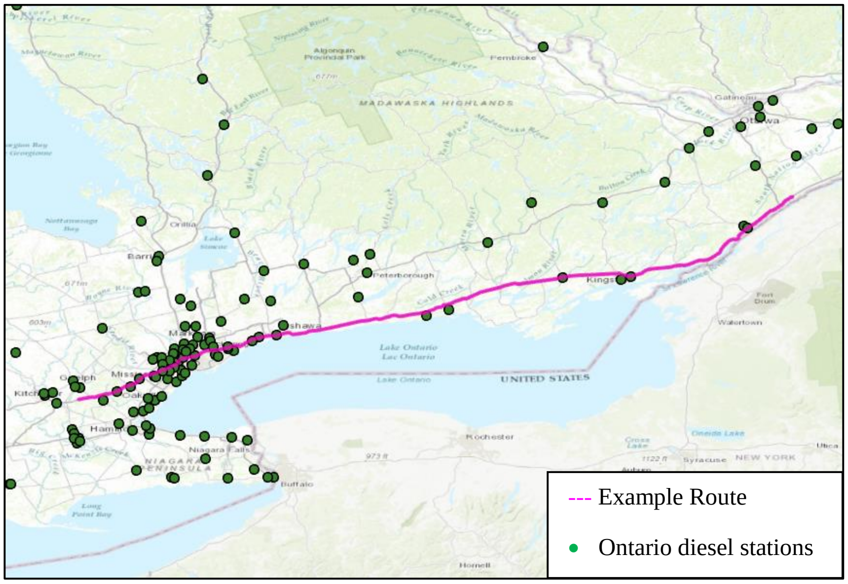


Figure 5-7: Diesel Fuel stations located near a sample truck route

5.7 Weigh Stations

Like gas stations, weigh stations are also important facilities for the freight industry as they have been constructed to inspect the weights of trucks along major truck routes. Their role along highway networks is to ensure that roads and structures are not compromised by the weights of the vehicles due to their payloads. A shapefile created by DMTI Spatial Inc. shown below in Figure 5-8 was used to map out the locations of all the weigh stations in Ontario. The majority of the weigh stations are located directly along major expressways and freeways.

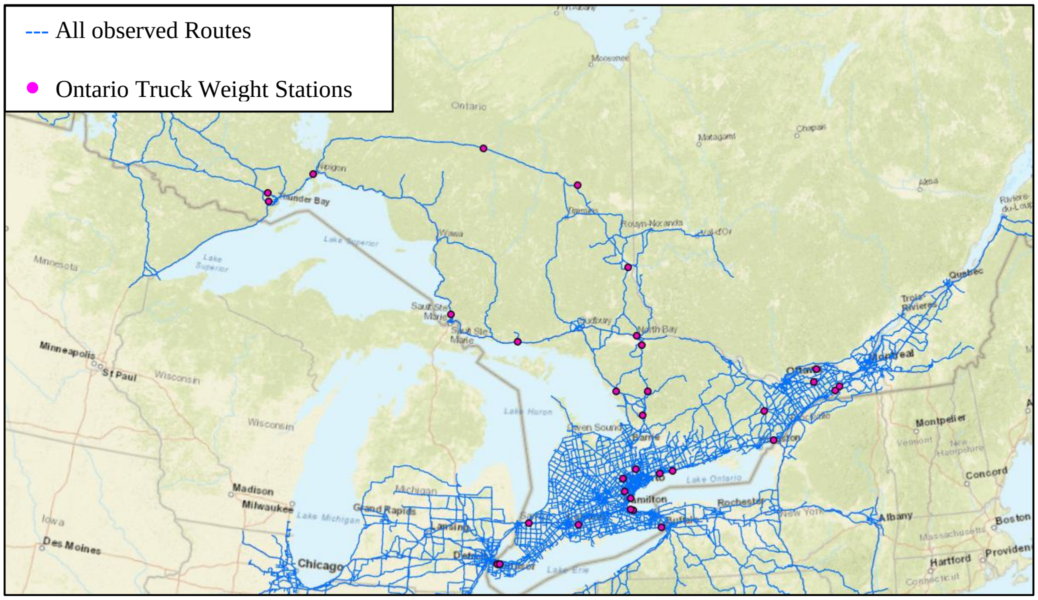


Figure 5-8: Truck Weigh Stations in Ontario

While these facilities have been built to serve transportation authorities, they may play a role in long haul route choice. While trucks would be required to stop when prompted by flashing beacons at the weigh stations, there may be some drivers who might avoid frequent inspections.

Chapter 6: Route Choice Model Estimation

In this chapter, the variables defined in Chapter 5 are used to estimate a route choice model to explain route decision making for long-haul trucks in Ontario. This step corresponds to the third objective of this thesis; using NLogit Software to estimate the route choice model.

6.1 NLogit Software

As shown by the literature review, logit models are a popular approach for route choice selection. Based on the literature review, the multinomial logit (MNL) model is used to explain route choice. The general form of multinomial logit model was presented in Section 2.2.1 in Equation 2-1. Since the commonality factor was used in defining the route choice set for each OD-Pair, it is appropriate to use the C-Logit variant of the MNL model as described by Equation 2-2. This allows the model to take into account the level of overlap between routes.

The output of the model is the probability of a route being selected based on its relative utility. A probability means that a driver is more likely to select a given route in order to maximize their utility. This is useful for understanding the current choices made by truck drivers. The factors and measures included in the model are referred to here as explanatory variables, which may have positive or negative coefficients depending on whether they promote or discourage an outcome. The significance of the coefficients will reveal which variables tend to be influential in the decision making process. An analysis of the elasticities of these variables will also be conducted to determine the responsiveness of the model.

6.2 Formatting the Dataset

NLogit software is able to interpret data in three formats; individual, aggregate, and proportional. If this excel data is organized by Trip ID, we would use a discrete format to organize the data set. Due to large number of observations in this model, the aggregate method is selected for this thesis.

This means that the input data, as shown in Table 6-1, includes a variable (Choice) containing the number of observations for a given route used for a particular OD-Pair. For example, the table indicates that the three routes for the OD-Pair CD3531_CD3542 in this sample had five, three, and seven observations. The choice variable therefore denotes the dependent variable for this model, along with the total number of alternatives available in the given choice set (Cset). This is then followed by a number of potential explanatory variables such as the Travel Time (TTime) or Buffer Time Index (BTIndx). These variables were introduced in Chapter 5 and will be further discussed below to determine a priori expectations when included in the model.

Table 6-1: Sample format for Aggregated Choice Data

Route Name	AltID	Cset	Choice	TTime	Dist	Avgspd	BTIndx	PTIndx	TTMIN
CD3531_CD3542_0	1	3	5	1.90	90.651	48.174	30.441	1.519	1.634
CD3531_CD3542_1	2	3	3	2.011	95.693	47.570	0.267	1.004	2.008
CD3531_CD3542_2	3	3	7	1.759	80.180	46.3644	24.281	2.504	0.872

Any potential explanatory variables were added to an Excel spreadsheet using the Table 6-1 format and exported into NLogit software for model estimation.

6.3 Expectations of Independent Variables

As previously discussed, the measured performance metrics shall be examined for significance on route choice. Table 6-2 provides a summary of all the variables included in the model estimation and their expected influence on route choice. A positive expectation indicates that an increase in the variable leads to greater desirability of a given route. A negative expectation indicates that a decrease in the variable leads to greater desirability of a given route.

Table 6-2: Expectations of Independent Variables on Route Choice

Variable	Description	Expectation
AVGSPD (km/h)	Average speed in km/hr for a route based on observed GPS trips	+
TTMIN (Hrs)	Minimum travel time for a route based on observed GPS trips	-
TTIME (Hrs)	Average travel time in hours for a route based on observed GPS trips	-
TTMAX (Hrs)	Maximum travel time for a route based on observed GPS trips	-
SIMSPD (km/h)	Average simulated speed for a route based on the ArcGIS network analyst using HERE data	+
SIMTTMIN (Hrs)	Minimum simulated travel time for a route based on the ArcGIS network analyst using HERE data	-
SIMTTIME (Hrs)	Average simulated travel time for a route based on the ArcGIS network analyst using HERE data	-
SIMTTMAX (Hrs)	Maximum simulated travel time for a route based on the ArcGIS network analyst using HERE data	-
TTINDX	Travel Time Index for a route based on observed GPS trips	-
BTINDX	Buffer Time Index for a route based on observed GPS trips	-
PTINDX	Planning Time Index for a route based on observed GPS trips	-
FWP	Observed proportion of freeways comprising a given route	+
FWP401	Observed proportion of Highway 401 comprising a given route	+
FWP4XX	Observed proportion of freeways other than Highway 401 comprising a given route	+
FWP407	Observed proportion of Highway 407 comprising a given route	-
INTRSCT	Number of intersections along a given route	-
DIESEL	Number of diesel gas stations along a given route	+
WEIGH	Number of weigh stations along a given route	+
CF	Commonality Factor representing level of overlap for a given route and all other alternatives	+

The expected effects of each variable are explained as follows:

AVGSPD and SIMSPD: Average speed, is function of distance and time. However, a greater speed implies better route performance, and would therefore be preferred to drivers. Hence, it would have a positive effect on route choice.

TTIME: Intuitively, as truck drivers are time minimizers, a longer travel time for a route should cause it to be less desirable.

TTMIN and TTMAX: Observed minimum and maximum travel times should have negative influences on route selection as they imply greater expected travel times for a driver, to which a driver is averse to.

SIMTTIME/SIMTTMIN/SIMTTMAX: Following the same reasoning as with observed minimum, average, and maximum travel times along a given route, the simulate travel times are expected to have the same negative affect on the utility of a route as drivers are time minimizers.

TTINDX, BTINDX, and PTINDX: These reliability indexes are functions of the observed travel time, free-flow travel time, and the maximum travel time. In general, the lower these values are for a given route, the more reliable it is perceived to be. Hence, all three should have negative influence on route choice. These variables are not included in the final model due to difficulties obtaining detailed information on the 95th percentile travel times.

FWP/FWP401/FWP4XX: It is expected that Freeway Proportion should have a positive influence on a given route being selected as truck drivers may prefer to use high speed uninterrupted transportation facilities to carry out deliveries. The more amount of freeway a given route has, the more likely it is to be selected compared to a route that may be comprised mostly of expressways and arterials. Highway 401 provides connection between east and west Ontario as well as to several other freeways and the American borders.

FWP407: The usage of toll facilities is expected to have a negative impact on route choice for a given route due to monetary costs. Highway 407 is the only major toll facility in Ontario passing through Toronto between Burlington and Oshawa. With regards to the Ontario OD-Pairs, trips utilizing Highway 407 account for a small fraction of the trip data set.

INTRSCT: The number of intersections a long a given route is expected to have a negative influence on route choice due to the assumption that drivers are time minimizers and intersections tend to be delay incurring features due to traffic lights, traffic exposure, and turning maneuvers.

DIESEL: The availability of refueling stations with diesel pumps are expected to have a positive influence on routing decisions as they tend to be the go-to opportunity of rest and recuperation along with fuelling opportunities. Drivers may tend to choose routes with more diesel stations in order to reduce uncertainties for long haul trips.

WEIGH: Weigh stations are present along major freeways and expressways in Ontario. Therefore, it is expected that this variable will appear positive in the model due to correlation with heavily utilized truck routes. However, this variable should be interpreted with caution since station locations are selected from a governmental perspective to ensure compliances by may not be desirable from a driver's perspective. For routes concerning rural areas where weigh stations are scarce, this variable may not have any significance at all.

CF: The commonality factor is a definitive metric in the choice sets for each OD-pair. The CF for a given route defines how unique it is compared to other routes within the choice set. A higher CF indicates increasing overlap with other routes in the choice set. Therefore, routes with higher overlaps tend to exhibit a higher number of truck trips as well. This is utilized in the MNL model (also called a C-Logit when CF is included) as a correction variable to account for overlap bias. This factor is expected to be positive and significant for each model estimation.

6.4 Descriptive Statistics

Some contextual information about the variables imported into NLogit was obtained by analyzing their descriptive statistics. The results are shown in Table 6-3.

Table 6-3: Descriptive Statistics of Variables

Variable	Mean	Standard Deviation	Minimum	Maximum	Observations
Choice Set	5.35	3.24	2	16	37,111
TTIME (hr)	4.35	3.16	0.49	31.04	37,111
TTMIN (hr)	3.91	3.03	0.06	27.10	37,111
TTMAX (hr)	5.09	3.87	0.49	65.91	37,111
AVGSPD (km/hr)	56.81	11.21	35.03	90.95	37,111
SIMTTMIN (hr)	2.78	2.50	0.01	23.91	37,111
SIMTTIME (hr)	3.39	2.72	0.01	23.98	37,111
SIMTTMAX (hr)	4.03	3.05	0.01	24.12	37,111
SIMSPD (km/hr)	83.70	9.96	32.08	104.45	37,111
FWP	0.29	0.34	0.00	1.00	37,111
FWP401	0.24	0.32	0.00	1.00	37,111
FWP4XX	0.17	0.25	0.00	1.00	37,111
FWP407	0.02	0.06	0.00	0.81	37,111
DIESEL	0.93	1.64	0.00	21	37,111
WEIGH	1.84	1.94	0.00	11	37,111
INTRSCT	35.33	50.18	0.00	434	37,111
CF	0.57	0.25	0.00	0.85	37,111

6.5 Preliminary C-Logit Model Results

The variables to be tested for significance and effects have been categorized into time-based variables and infrastructure-based variables. Variables under each category are tested individually and in combination with others by estimating preliminary C-logit models in NLogit. The model results indicate the statistical significance of each variable and the sign of its coefficient.

Variables exhibiting low statistical significance or non-intuitive signage are dropped from the model, while statistically significant variables with expected signs are retained and further inspected for linear correlations. One of two variables in a pair of highly correlated variables shall be retained while the other is dropped. The quality of the model is determined by it's ρ^2 value, which indicates how well the model explains variability in the dataset. Models with higher ρ^2 values are considered to perform better.

The best variables from each model are combined to create a final C-Logit model that best explains route choice for long haul trucks. Further testing is discussed later to analyze the performance of the final model.

6.5.1 Time-based variable models

1. Models based on GPS Based Travel times

The GPS based travel time variables included TTMIN, TTIME, and TTMAX. These are based on the actual recorded travel times for each truck trip from its point of origin to destination, including the portion travelled within their origin and destination zones. Each variable was tested independently and in combination with the other variables in this category to determine their effects on route choice. Initial models with only one variable alongside the CF are discussed below and their results provided in Appendix E. Table 6-4 shows a combined model with all variables of this included.

TTMIN is a dominant and intuitively negative variable when tested as a standalone variable to explain the dataset. This explains that routes with lower experienced minimum travel times tend to be more popular amongst drivers. TTIME, representing the average observed travel time, is also a dominant variable but does not explain route choice as well as TTMIN does. The ρ^2 value for the model based on TTIME is very low. The variables sign is intuitively negative when tested as a standalone variable to explain the dataset.

TTMAX is a very significant variable but shows a positive influence on route choice. This is non-intuitive. The positive influence of maximum travel time is likely caused by the greater likelihood of an extended time when there is are larger set of observations. This manifests in the model as a positive parameter but is not likely indicative of route choice preference.

Table 6-4: Model results for all observed travel time variables

Variable	Coefficient	T-Statistic
TTMIN	-2.24***	-54.88
TTIME	.481***	8.7
TTMAX	.805***	45.15
CF	.990***	18.84
LL(0)	-52494.89	
LL(β)	-26013.05	
ρ^2	0.504	
No. of Observations	37,111	

***** *indicates the parameter is statistically significant with 99% confidence***

The most significant variables when including all observed travel time factors are the TTMIN and TTMAX while TTIME does not have much significance. A Pearson correlation is performed to determine the linear correlation between these pairs of variables as shown in Table 6-5.

Table 6-5: Pearson Correlation for observed travel time variables

	TTIME	*TTMIN*	*TTMAX*
TTIME	1		
TTMIN	0.983	1	
TTMAX	0.934	0.867	1

The Pearson correlation shows that all GPS Travel time based variables are highly correlated with each other. As a result, two of the three variables will not remain in further models. The TTIME variable is removed due to a lack of sign consistency. TTMAX is removed due to a non-intuitive sign in all estimated models. TTMIN remains as the best variable to explain route choice based on the GPS based travel time.

In terms of route performance, the observed average speed of a given route (AVGSPD) is explored for its influence on route choice, both independently and in combination with the observed minimum travel time variable TTMIN. These two model configurations are shown in Appendix E and Table 6-6 respectively.

Table 6-6: Model results of observed average speed and minimum travel time

Variable	Coefficient	T-Statistic
TTMIN	-2.75***	-119.47
AVGSPD	.0234***	15.01
CF	1.145***	22.29
LL(0)	-52494.89	
LL(β)	-30133.96	
ρ^2	0.426	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Both models show that AVGSPD is a significant and positive variable as expected, indicating driver preference to select routes with higher average speeds. When paired with TTMIN, the model performs better based on its ρ^2 value. A Pearson correlation in Table 6-7 shows that these two variables are uncorrelated and can both be included in the final model.

Table 6-7: Pearson correlation for observed route performance variables

	TTMIN	AVGSPD
TTMIN	1	
AVGSPD	0.203	1

As discussed already for TTMAX, there is an inherent potential bias when using the observed GPS data for observations and independent variables. For example, routes with more observations become more likely to obtain an outlier travel time due to a larger sample size. This would therefore reflect the data structure instead of the route choice behaviour we are modelling. As an alternative, historical travel time information is discussed in the next section to provide externally sourced information without implicit bias.

The TTINDX, BTINDX, and PTINDX reliability variables exhibited similar behaviour as the TTMAX variable, being positive in each estimated model. This was expected based on the discussion in Section 6.3 and the same inherent bias as for TTMAX. Consequently, these reliability indexes have also been dropped from the model.

2. HERE based models

The process used for the GPS based travel time variables is replicated here with simulated travel time variables obtained from ArcGIS Pro and built-in HERE data travel times. The process for deriving these times has been discussed previously in section 5.1. The C-logit model estimations for the HERE based minimum, average, and maximum travel times as standalone variables are shown in Appendix E.

In all three models, the simulated travel time variables are significant and negative, indicating a preference for routes that have lower perceived minimum, average, and maximum travel times. This is an intuitive result from the perspective of drivers, and also provides confirmation that the previous counter-intuitive relationship for TTMAX may be been due to an inherent bias.

A combined model comprising all of the simulated travel time variables SIMTTMIN, SIMTTIME, SIMTTMAX is estimated and shown in Table 6-8.

Table 6-8: Model results for simulated travel time variables

Variable	**Coefficient**	**T-Statistic**
SIMTTMIN	-.606***	-6.42
SIMTTIME	-.469***	-2.6
SIMTTMAX	.398**	4.52
CF	2.64***	57.22
LL(0)	-52494.89	
LL(β)	-48239.98	
ρ^2	0.081	
No. of Observations	37,111	

***** indicates the parameter is statistically significant with 99% confidence*

In all model versions tested, the SIMTTMIN variable retains its negative sign and remains the most dominant variable in all the models. The Pearson correlation result shown in Table 6-9 suggests that all the simulated travel time variables are highly correlated with each another. Consequently, SIMTTMIN will be retained as the most explanatory variable of route choice due to its statistical significance and intuitive sign.

Table 6-9: Pearson correlation for simulated travel time variables

	SIMTTIME	***SIMTTMIN***	***SIMMAXHR***
SIMTTIME	1		
SIMTTMIN	0.983	1	
SIMMAXHR	0.987	0.943	1

In terms of route performance, the simulated average speed of a give route (SIMSPD) is explored for its influence on route choice. A model configuration based on SIMSPD is shown in Appendix E.

SIMSPD is a significant and positive variable, indicating drivers’ preference to select routes with higher average speeds. This is an expected and intuitive result that explains route choice from the perspective of perceived route performance.

Using SIMTTMIN as the best time-based variable, a combined model using SIMTTMIN and SIMSPD is estimated in Table 6-10.

Table 6-10: Model results for simulated minimum travel time and average speed

Variable	Coefficient	T-Statistic
SIMTTMIN	-.678***	-48.45
SIMSPD	.107***	83.84
CF	2.168***	45.39
LL(0)	-52494.89	
LL(β)	-44381.62	
ρ^2	0.155	
No. of Observations	37,111	

******* *indicates the parameter is statistically significant with 99% confidence*

The model shows that when SIMTTMIN is paired with SIMSPD, the model performs better based on its ρ^2 value. A Pearson correlation reveals that these two variables are partially correlated.

Table 6-11: Pearson correlation for HERE based route performance variables

	SIMSPD	***SIMMINHR***
SSIMSPD	1	
SIMTTMIN	0.536	1

When comparing the GPS time-based model to the HERE-based models, there is consistency in the behavior of the variables representing travel time and average speed. The variables representing minimum travel time, TTMIN and SIMTTMIN, are both dominant and have intuitive negative signage explaining route choice well. Similarly, the speed variables AVGSPD and SIMSPD are also significant and have intuitive positive signs in the model configurations shown in Table 6-6 and Table 6-10.

The GPS time-based models substantially outperform the models estimated using HERE data. The significance and ρ^2 values of the GPS-based models warrant the usage of the TTMIN and AVGSPD variables. Furthermore, GPS-based models are much easier to replicate since the TTMIN variables can be extracted from the trip dataset rather than running several simulations to attain the SIMTTMIN values in the HERE-based models.

6.5.2 Infrastructure-based variable models

An initial configuration of the model containing all infrastructure based variables is shown in Table 6-12, along with a Pearson correlation matrix for the infrastructure based variables in Table 6-13.

Table 6-12: Model results for infrastructure variables

Variable	Coefficient	T-Statistic
FWP	1.512***	21.57
FWP401	2.072***	43.48
FWP4XX	-.626***	-8.74
FWP407	0.259	1.34
INTRSCT	-.0138***	-40.8
DIESEL	.382***	83.94
WEIGH	.301***	35.29
CF	1.53***	27.49
LL(0)	-52494.89	
LL(β)	-30255.60	
ρ^2	0.424	
No. of Observations	37,111	

****** indicates the parameter is statistically significant with 99% confidence*

Table 6-13: Pearson correlation for infrastructure variables

	FWP	*FWP401*	*FWP4XX*	*FWP407*	*INTRSCT*	*DIESEL*	*WEIGH*	*CF*
FWP	1							
FWP401	-0.286	1						
FWP4XX	0.669	-0.228	1					
FWP407	0.256	-0.004	0.091	1				
INTRSCT	-0.193	-0.046	-0.111	-0.072	1			
DIESEL	0.002	0.142	0.104	0.064	0.023	1		
WEIGH	0.202	0.434	0.176	-0.006	-0.064	0.171	1	
CF	0.143	0.219	0.058	0.057	-0.029	-0.028	0.209	1

All variables are statistically significant and have intuitive influences in terms of signage based on previous a priori expectations. The freeways proportion (FWP) and Highway 401 freeway proportion (FWP401) are both statistically significant and positive variables, indicating the preference of drivers to use uninterrupted high speed facilities. Another intuitive yet less significant infrastructure variable is the proportion of 400 series freeways other than Highway 401 (FWP4XX). This indicates that long haul drivers are less attracted to using freeways other than Highway 401 such as Highway 400, 403, and the QEW but still find these roads more attractive compared to non-freeways. This is an intuitive result given the lesser extent of connectivity these other 400-series highways provide to freight hubs compared to Highway 401. The FWP4XX parameter is not as statistically significant based on its t-statistic. For this reason, this variable will be dropped from the final model.

A similar argument can be made for the usage of Highway 407 (FWP407). Although significant with a relatively lower t-statistic, the positive sign of this variable implies than drivers are willing to use routes that include tolled facilities, maybe in the interest of time saving or access to specific destinations. However, given its low significance, this variable is subsequently dropped from the model.

As expected, the availability of diesel gas stations (DIESEL) and the truck weigh stations (WEIGH) are both significant and positive variables in the initial configuration of the model. They imply that routes with more rest and refueling opportunities are more attractive to drivers compared to routes with possibly inadequate diesel gas stations. The number of intersections (INTRSCT) is a significant and negative variable, confirming that drivers are averse to routes with many intersections and interruptions.

As discussed in Section 5.7, although WEIGH is a positive and dominant variable for route choice, the availability of weigh stations is dependent on existing truck route patterns as established by transportation authorities. Since it is likely not a cause for route choice from driver’s perspective, it is dropped from the final model.

Based on the above analysis, the remaining infrastructure based variables include FWP, FWP401, INTRSCT, and DIESEL. From the correlation matrix, it is evident that none of these variables are sufficiently correlated with each other to cause problematic interactions in the model. A C-logit model estimated using the aforementioned preferred infrastructure variables is provided below in Table 6-14.

Table 6-14: Model results for best infrastructure variables

Variable	**Coefficient**	**T-Statistic**
FWP	1.452***	28
FWP401	2.393***	53.48
INTRSCT	-.0163***	-48.65
DIESEL	.403***	91.09
CF	1.627***	29.42
LL(0)	-52494.89	
LL(β)	-30927.91	
ρ^2	0.410	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

A comparison of the models in Table 6-12 and Table 6-14 shows that removal of the variables FWP4XX, FWP407, and WEIGH does not substantially impact the performance of the model. When comparing the performance of the infrastructure based models to the time-based models, the time based variables outperform the former, but have very similar ρ^2 values. While there may be some indirect correlation between time-based variables and infrastructure such as TTMIN and FWP, it seems that route choice may be more dependant on the presence of physical infrastructure along routes than the perceived route performance. A combined model using both time-based and infrastructure-based variables is presented in the next section.

6.5.3 Combined model

The best time-based and infrastructure-based variables from sections 6.5.1 and 6.5.2 were tested in combination using NLogit. This model is shown in Appendix E.

In the model using GPS time based variables in combination with the infrastructure variables, AVGSPD has a counterintuitive negative sign and is statistically insignificant. As a result, it can be removed from the model. Table 6-15 shows the final combined model which incurs minimal change after removal of AVGSPD.

Table 6-15: Final model results

Variable	Coefficient	T-Statistic
TTMIN	-1.666***	-77.95
FWP	1.18***	22.56
FWP401	1.975***	42.67
DIESEL	.269***	57.18
INTRSCT	-.006***	-17.13
CF	.687***	11.62
LL(0)	-52494.89	
LL(β)	-25021.69	
ρ^2	0.523	
No. of Observations	37,111	

***** *indicates the parameter is statistically significant with 99% confidence***

A Pearson Correlation analysis for the above model is provided in Table 6-16 and shows that all the infrastructure and time based variables are not correlated with each other.

Table 6-16: Pearson correlation for route performance and infrastructure variables

	TTMIN	*FWP*	*FWP401*	*INTRSCT*	*DIESEL*	*CF*
TTMIN	1					
FWP	0.108	1				
FWP401	0.005	-0.286	1			
INTRSCT	0.115	-0.193	-0.046	1		
DIESEL	-0.031	0.002	0.142	0.023	1	
CF	0.163	0.143	0.219	-0.029	-0.028	1

The variables from Table 6-15 are retained for the remainder of this thesis for conducting a sensitivity analysis of the commonality factor, evaluating model accuracy in predicting route choice, and testing future scenarios.

Chapter 7: Model Performance

The predictive power of the estimated route choice model and its variables is assessed in this chapter through four analyses. First, a sensitivity analysis of the commonality factor threshold value is conducted in Section 7.1 to explore its effect on the quality of the model. Recall that the CF threshold was a voluntarily defined value of 0.85.

In Section 7.2, the variable coefficients obtained from the models in Section 7.1 estimation are then used to predict the trip frequencies of routes based on their utilities. The predicted frequencies of routes are plotted against the observed frequencies to investigate the accuracy of each model. The best CF threshold value and corresponding model is then determined based on these plots and the ρ^2 measure. The latter measure identifies the explanatory power of a given model compared to a model lacking any explanatory variables.

In Section 7.3, the final model is further examined and improved with consideration to any visible outliers in its predicted frequency vs. observed frequency plots. Lastly, the elasticity of each variable in the model is discussed in Section 7.4.

7.1 Commonality Factor Sensitivity Analysis

The commonality factor is the primary determinant of the number of unique routes that make up the route choice set. In theory, increasing or decreasing the threshold for uniqueness would directly affect the number of alternatives in the choice set and could subsequently influence the results. The impact of this change is not well documented in the existing literature.

To analyze the impact of CF on the route choice model, the Choice Set Generation algorithm from Chapter 4 was rerun several times with the CF Threshold adjusted to 0.65, 0.70, 0.75, 0.80, 0.90, and 0.95. OD-Pairs found to have only one route between them were removed from the analysis as was conducted in earlier chapters. The results of each run of the algorithm is shown in Table 7-1.

The initial CF threshold of 0.85 is highlighted as a base reference. It is evident that a lower CF threshold results in small choice sets of more numerous but less unique routes.

Table 7-1: Commonality factor statistics

CF Threshold	Number of OD-Pairs	Number of Routes	Number of Trips
0.65	471	1,509	33,342
0.70	492	1,652	33,946
0.75	522	1,820	34,962
0.80	551	2,007	36,179
0.85 (original)	*577*	*2,220*	*37,111*
0.90	605	2,512	38,012
0.95	650	2,890	39,680

For each choice set configuration derived from the scenarios in Table 7-1, the route choice model was reproduced following the model specification from Chapter 6. After obtaining the relevant data under each variable, a C-Logit model is estimated in NLogit for each dataset. The results are shown in Table 7-2.

Table 7-2: Sensitivity Analysis Results

CF Threshold	**0.65**		**0.70**		**0.75**		**0.80**		**0.85**		**0.90**		**0.95**	
Variable	β	(t-stat)	β	(t-stat)	β	(t-stat)	β	(t-stat)	β	(t-stat)	β	(t-stat)	β	(t-stat)
TTMIN	-1.81	(-73.03)	-1.83	(-75.41)	-1.94	(-83.31)	-1.86	(-83.54)	-1.66	(-77.95)	-1.57	(-78.22)	-1.72	(-91.33)
FWP	.984	(19.2)	1.204	(25.26)	1.128	(22.01)	1.288	(25.11)	1.179	(22.56)	1.058	(19.52)	1.06	(19.38)
FWP401	1.88	(37.95)	1.993	(43.81)	2.155	(44.4)	2.252	(45.49)	1.974	(42.67)	2.118	(40.99)	2.16	(41.92)
DIESEL	.173	(52.96)	.172	(52.2)	.16	(46.4)	.171	(48.98)	.269	(57.18)	.206	(63.24)	.19	(55.88)
INTRSCT	-.004	(-10.65)	-.138	(-8.52)	-.003	(-10.66)	-.004	(-12.44)	-.006	(-17.13)	-.005	(-15.85)	-.008	(-20.14)
CF	.368	(4.82)	.237	(3.39)	.391	(6.13)	.412	(6.84)	.686	(11.62)	.627	(7.02)	.38	(7.52)
LL(0)	-39879.79		-42270.76		-45613.51		-49537.21		-52494.89		-56739.94		-62068.56	
LL(β)	-18019.38		-19198.23		-21310.22		-23256.99		-25021.69		-27311.57		-32710.01	
ρ^2	0.548		0.546		0.533		0.531		0.523		0.518		0.473	
Observations	33,342		33,946		34,962		36,179		37,111		38,012		39,680	
One trip routes	482		569		660		775		914		1108		1331	

The sensitivity analysis of the commonality factor revealed interesting observations about route choice modelling. These are discussed below with reference to the original Chapter 6 results obtained using the original CF threshold value of 0.85.

Result 1: Increasing CF leads to more routes

Reducing the CF resulted in smaller but more unique choice sets. Consequently, there would be OD-Pairs with only one observed route remaining in their choice set, which are subsequently removed from the model estimation. Conversely, a higher CF value would result in larger choice sets of less unique routes. For example, OD-Pairs that previously had only one route between them when CF=0.85 may now have two or more routes when CF=0.95. Therefore, more OD-Pairs are included in the model scenarios with higher CF values.

Result 2: Smaller choice sets have routes with more trips (fewer one trip routes)

When CF exceeds the original 0.85 value, more routes in a given choice set exhibited a fewer number of trips. When CF is set below the 0.85 value, more trips are observed along routes in a smaller choice set. This is indicated by the decreasing number of one trip routes in Table 7-2. For example, a hypothetical choice set with three routes with trip frequencies of 10,1,1 when CF=0.95; whereas, that choice set would be reduced to two routes having trip frequencies of 10,2 when CF=0.65. While higher CF's can reveal the more popular routes, lower CF can better compare and explain route choice in terms of observations.

Result 3: Consistency of variable effects

Each of the variables in the model showed relatively similar levels of significance and signage as CF was adjusted. However, the highest significance for a given variable occurs in different models with no identifiable trend as the CF is increased or decreased. For example, TTMIN was at its highest when CF=0.95, while DIESEL was at its highest at CF=0.9. This suggests that no single model performs best from this perspective.

Result 4: ρ^2 improves with lower CF threshold values

The ρ^2 value that indicates the explanatory power of the model was lowest at CF=0.95 and highest at CF=0.65. This can be explained by the lower number of routes in a choice set when the CF is lower. The choice set, and subsequently the individual choice of route, becomes simpler.

Based on the ρ^2 value alone, the optimal Commonality Factor threshold for route choice analysis in this thesis is 0.65 (ρ^2= 0.548).

7.2 Observed and Predicted Model Comparisons

The performance of a model is assessed in this section by its ability to predict the observed route frequencies using the parameters from the models in Table 7-2. Using Equation 2-1 introduced in Section 2.2.1, the utility of each route in a choice set is calculated using the independent variables, TTMIN, FWP, FWP401, DIESEL, INTRSCT, and CF.

As an example, the coefficients obtained for the model estimated using a CF threshold value of 0.65 have been adapted into Equation 2-1, as shown below.

$$\begin{aligned} V_k = & -1.816(TTMIN_k) + .985(FWP_k) + 1.88(FWP401_k) \\ & + .173\,(DIESEL_k) - 0.007(INTRSCT_k) + .368(CF_k) \end{aligned} \tag{7-1}$$

Where:

- V_k is the calculated utility of a given route k
- $TTMIN_k$ is the observed minimum GPS based travel time along route k
- FWP_k is the measured proportion of Freeways comprising a route k
- $FWP401$ is the measured proportion of Highway 401 comprising a route k

- $DIESEL_k$ is the number of diesel stations observed along a given route
- $INTRSCT_k$ is the number of intersections observed along a route k
- CF_k is the commonality factor of route k

Using the relevant parameters, the utilities of all routes in a choice set are calculated using each model shown in Table 7-2. The probability of a route is then calculated using Equation 2-2. The predicted frequency is then determined by multiplying the number of observations for a given OD-Pair by the probability of selecting a given route. These predicted values are then compared to the observed choice frequency. An example prediction is shown in Table 7-3 for an OD-Pair with five potential routes in the choice set.

Table 7-3: Predicted frequencies of routes in example choice set

Routes	Obs	TTMIN	FWP	FWP401	Diesel	INTRSCT	CF	Total Obs	$\exp(V_k + \beta_{CF}.CF_k)$	$\sum_{l \in C} \exp(V_l + \beta_{CF}.CF_l)$	P_k	Predicted Obs
CD3506_CD3512_0	29	2.86	0.18	0.80	0	1	0.28	104	0.03	0.09	0.38	39
CD3506_CD3512_1	17	2.35	0.00	0.00	1	42	0.60	104	0.02	0.09	0.20	21
CD3506_CD3512_2	54	2.43	0.00	0.00	4	22	0.60	104	0.03	0.09	0.32	33
CD3506_CD3512_3	1	3.20	0.06	0.00	0	53	0.00	104	0.00	0.09	0.03	3
CD3506_CD3512_4	3	3.27	0.00	0.30	2	46	0.28	104	0.01	0.09	0.07	7

The route choice predictions were conducted for all routes in the seven models representing different CF threshold values. A linear trend line is generated through the scatterplot shown in Figure 7-1 to indicate how similar the predicted frequencies are to the observed frequencies for CF=0.65. The R^2 value of 0.93 for the resulting trend line suggests that the predicted values generated by the model are relatively accurate.

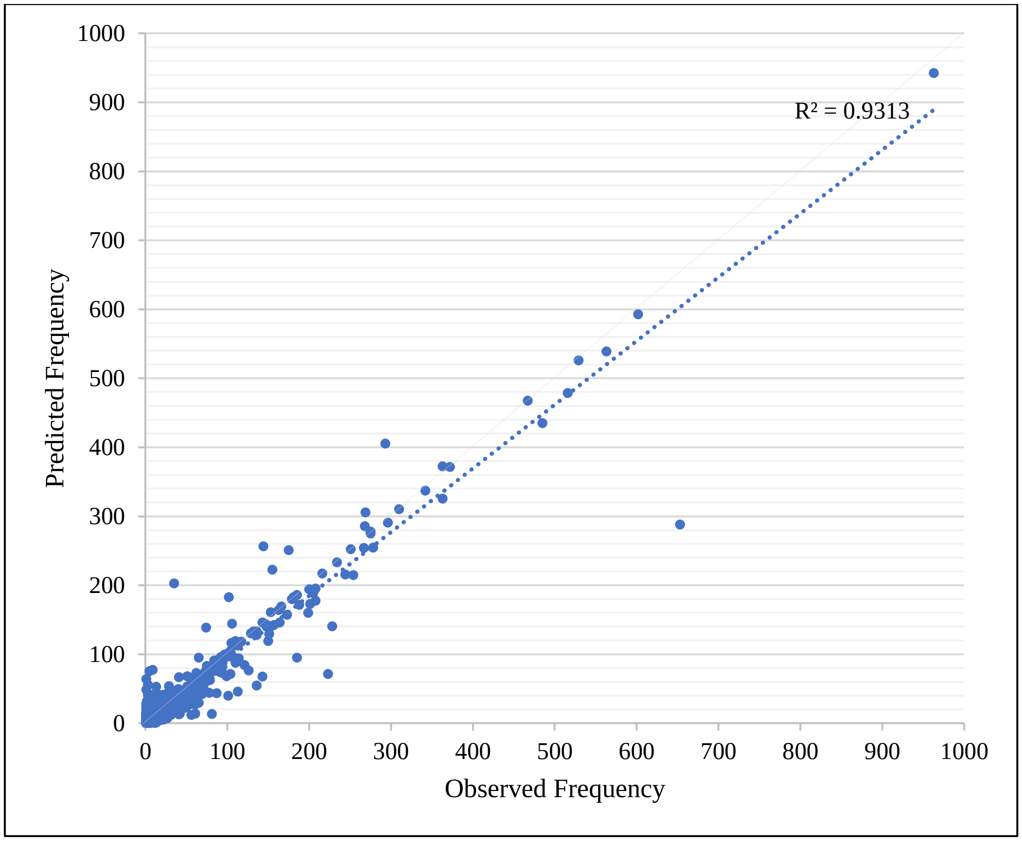


Figure 7-1: Model predicted frequency vs. observed frequency of route choice for CF=0.65

In the same manner, trend lines have been generated for comparison with the models generated from other CF threshold values. These trend lines are combined into a single graph shown in Figure 7-2 for comparison.

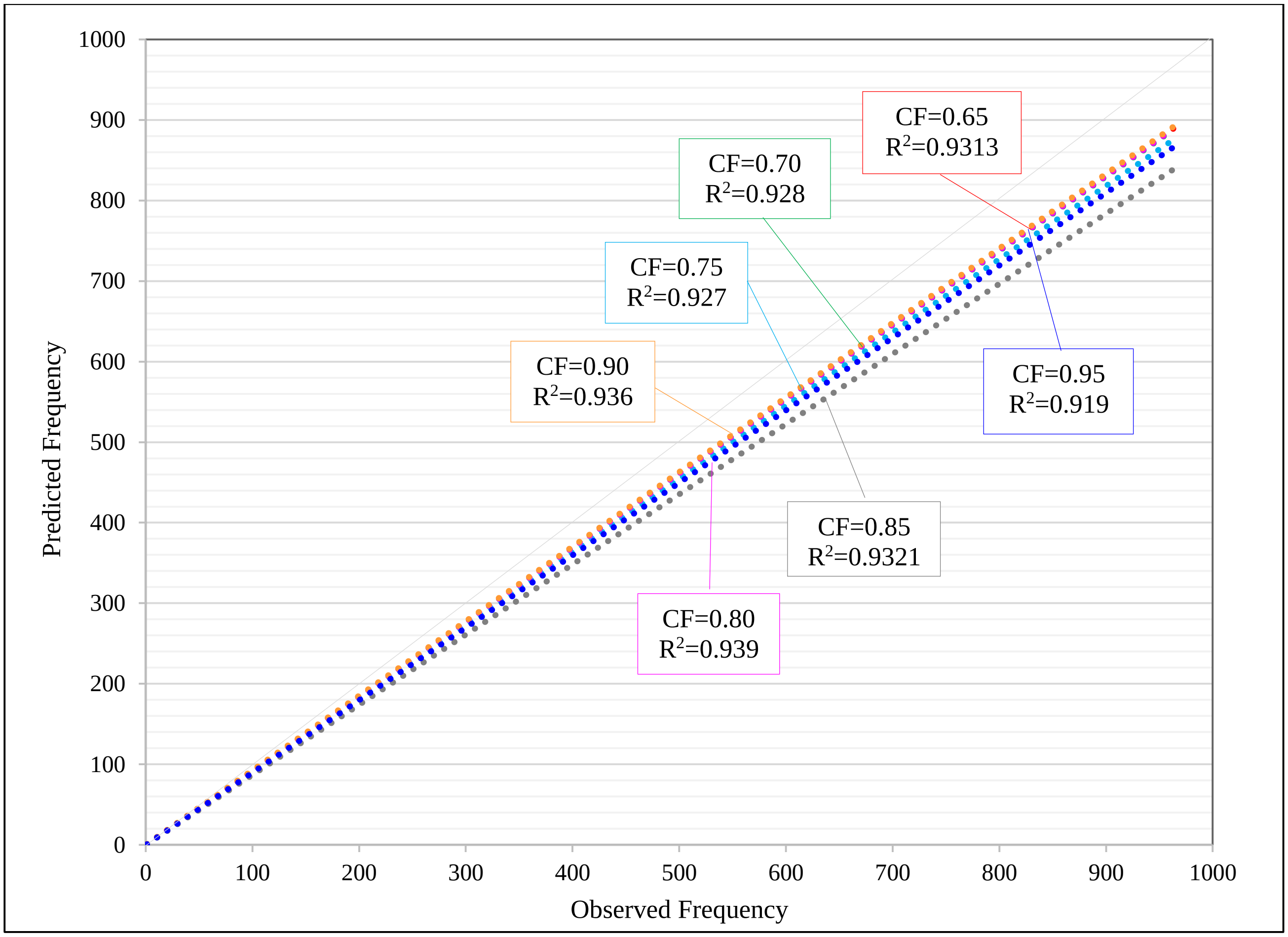


Figure 7-2: Predicted vs. observed frequency plots for various model estimations

From Figure 7-2, it is apparent that all the models behave similarly and perform reasonably well in predicting route frequencies for their respective datasets. This is suggested by their R^2 values which are all above 0.9. In contrast to an increasing ρ^2 value observed with decreasing CF threshold values, there appears to be no relationship between the R^2 values calculated here and the CF thresholds. This suggests that the CF threshold and subsequent choice set do not have a significant impact on the predictive power of the model in this thesis.

Another scatterplot is created to investigate the applicability of model parameters on choice sets based on different CF threshold values. In this example, the model parameters obtained for CF=0.65 are used to predict the route frequencies of a dataset consisting of choice sets generated using CF=0.85. This is shown in Figure 7-3. The scatterplot for the frequencies predicted using the model parameters from CF=0.85 is also provided for comparison.

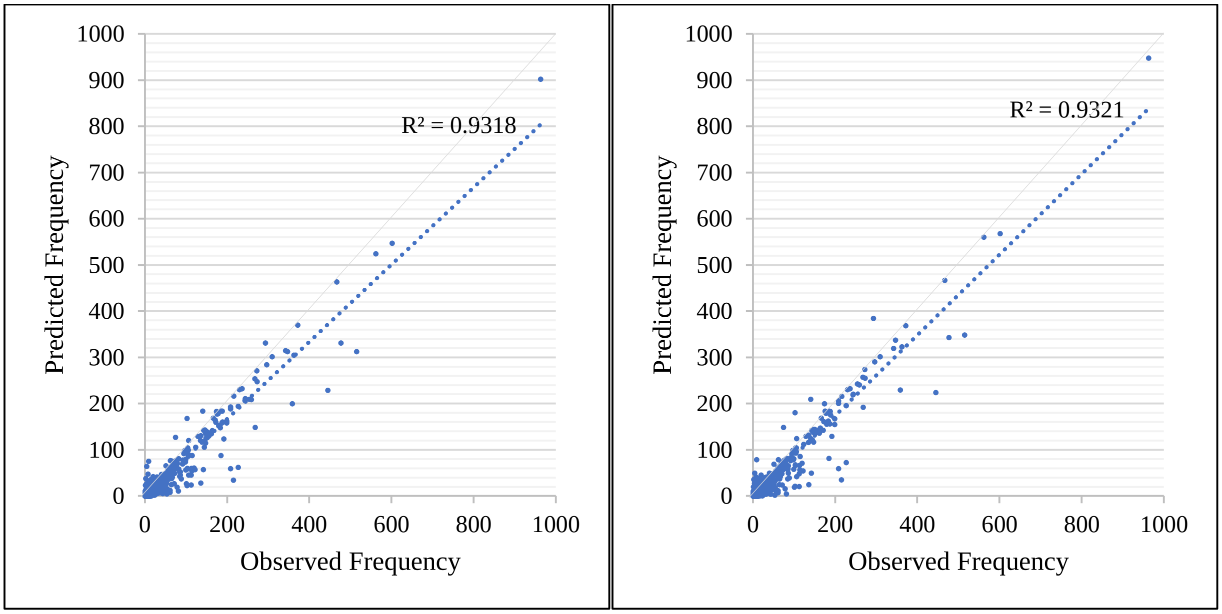


Figure 7-3: Scatterplots of CF=0.65 parameters on CF=0.85 choice set (left) compared to CF=0.85 parameters on CF=0.85 choice set (right)

Figure 7-3 shows that the model parameters for CF=0.65 predict the route frequencies for the 0.85 dataset just as well as the model parameters based on CF=0.85. This may be due to the parameter values being very similar in both model estimations. However, this further confirms that the model appears insensitive to the chosen CF threshold value. From a practical perspective, a modeler applying the results of this model can therefore have confidence that the parameters can be applied to a choice set based on different assumptions for the number of routes, all other things being equal.

While the CF threshold value has ultimately little impact on the overall model, the preferred model to continue in this thesis is selected for a CF=0.65 due to the ρ^2 values discussed in Section 7.1.

7.3 Outlier Analysis and Final Model

The previous scatterplot from Figure 7-1 is shown again here with two outliers identified.

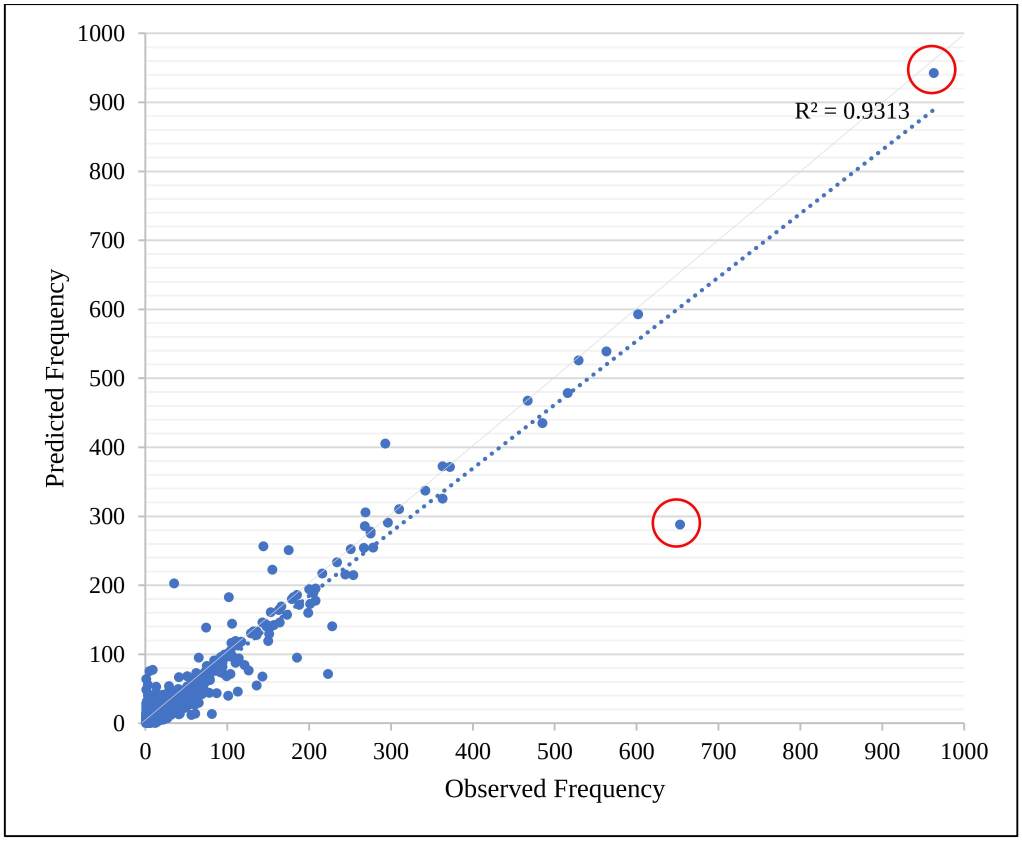


Figure 7-4: Model predicted frequency vs. observed frequency of route choice with outliers identified

Outlier #1 represents a route with an observed frequency of 653 and a predicted frequency of 287, therefore under predicting by more than 50%. This point represents a route between zones Montreal (CD2466) and Stormont, Dundas, and Glengarry Counties (CD3501). A close examination of the trips using this route showed that poor prediction arises from the spatial location of local trip end points within the origin and destination zones. As a reminder, routes in this thesis have only been defined between zonal boundaries and therefore exclude first and last mile routing. Therefore, it is recommended that local origin and destination proximities to established routes should be considered for future studies in route choice between Canadian zones.

Since Outlier #1 may bias the parameter results, the route choice model was re-estimated with the removal of all trips in the OD-Pair containing this outlier route. The re-estimated route choice model with outlier treatment is shown in Table 7-4 along with an updated scatterplot visualizing observed and predicted model results in Figure 7-5.

Table 7-4: Model results with outlier removed

Variable	Coefficient	T-Statistic
TTMIN	-1.65***	-68.11
FWP	1.17***	22.53
FWP401	1.999***	40.21
DIESEL	.180***	54.74
INTRSCT	-.003***	-9.62
CF	.264***	3.42
LL(0)	-38523.49	
LL(β)	-17344.63	
ρ^2	0.550	
No. of Observations	32,645	

**** indicates the parameter is statistically significant with 99% confidence*

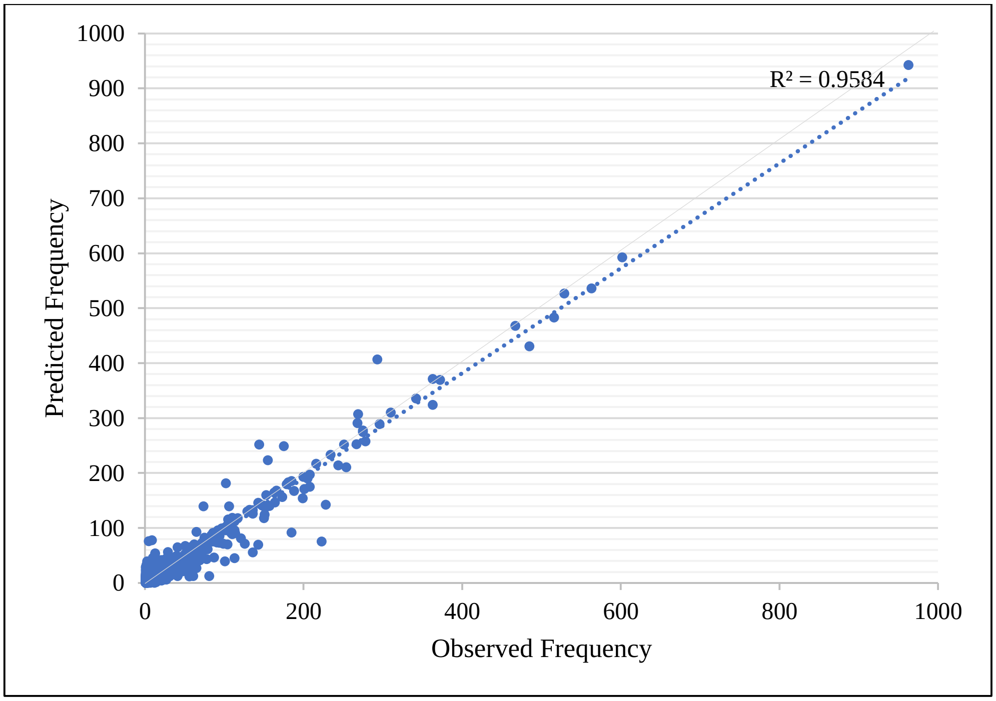


Figure 7-5: Model predicted frequency vs. observed frequency with outlier removed

The ρ^2 of the re-estimated model indicates the model performs slightly better without the OD Pair containing the outlier route. The re-estimated model also produces a better fit to the dataset compared to the previous model based on the scatterplot in Figure 7-4. This confirms that the removal of this outlier subsequently improved the model.

Outlier #2 shown in Figure 7-4 is a route exhibiting a predicted frequency of 963 against an observed frequency of 942. Although this is a good example of how well the model functions outside of the typical range of frequency values, the large amount of data associated with this route represents a larger influence compared to the rest of the dataset. It is noteworthy that this route is between the Region of Peel (CD3521) and the Region of Waterloo (CD3530), which was the largest OD-Pair in terms of observed trips as seen previously in Table 3-1. A closer analysis of the characteristics of the most attractive route showed that the travel time associated with it was substantially lower (almost half) when compared to other route options in the route choice set. This route comprises of Highway 401 and regional roads passing through Milton including Appleby Line, Steeles Avenue, James Snow Parkway, and Derry Road. Hence extreme values for this route seen in the scatterplot is caused by the large amount of traffic between these two zones and lack of route competitiveness.

To understand the influence of the route on model, the C-Logit model was re-estimated with this OD-Pair and associated trips removed from the dataset. The re-estimated route choice model with removal of this outlier had a lower ρ^2 and subsequently lower R^2 value in its predicted vs. observed frequency scatterplot. As removal of this OD-Pair did not improve the model results, it was retained.

The variant of the model shown in Table 7-4 is the optimized route choice model for this thesis. Nonetheless, it is recommended for future studies to further explore such outliers and experiment with other new variables such as local proximity to improve the predictive power of the route choice models.

7.4 Variable Elasticities

The value of a parameter in an estimated model is not indicative of the amount of influence a given variable has on route choice. The influence of each variable is instead tested here using direct elasticities, which measure the change in probability for a given route if a specified variable increased by 1%.

The Effects function was used in NLogit to investigate the elasticities of each route choice model. The outputs from NLogit for the effects of each variable on selecting an alternative are provided in Appendix F. The elasticities are originally given separately for each alternative, but have been merged into one value using a weighted average. The elasticities for each variable are shown below in Table 7-5.

Table 7-5: Weighted average elasticities of route choice variables

Variable	Elasticity
TTMIN	-3.306
FWP	0.212
FWP401	0.166
DIESEL	0.101
INTRSCT	-0.069

Minimum Travel Time (TTMIN): TTMIN has a negative elasticity. This suggests that a positive change (or increase) in the minimum travel time for a given route negatively affects its probability of being selected, and promotes the selection of other routes in the choice set, and vice-versa. The magnitude of the elasticity of TTMIN is larger than 1, indicating TTMIN to be very elastic and have substantial impact on route choice if its value slightly changes.

Proportion of Freeways (FWP) and Proportion of Highway 401(FWP401): FWP and FWP401 have positive elasticities, suggesting that if a given route was to increase the amount of freeways or Highway 401 in its composition, it would become more attractive to a driver over other available routes. The magnitude of FWP and FWP401's elasticity is however less than 1, suggesting they are inelastic variables. This means if a route's composition in terms of freeway facilities were to change, it would not become significantly more or less attractive compared to other routes in the choice set.

Number of Diesel Stations (DIESEL): The elasticity of the frequency of diesel stations along a given route is positive but less than 1. This suggests that increased opportunities for refueling and rest along a route would have low positive impact on its relative utility over other routes in a choice set. Due to the nature of trucking activities, drivers do not need to stop and refuel that often as it would affect their trip duration, which is considered to be their first priority. Furthermore, the presence of diesel stations along observed truck routes does not necessarily indicate their utilization. Therefore, the addition of a new diesel station along a route may not substantially improve a routes utility, hence making DIESEL an inelastic variable.

Number of Intersections (INTRSCT): INTRSCT has negative value for elasticity, suggesting that an increase in the number of intersections along a given route negatively affects its relative utility. This is intuitive as more intersections along a route incur increased stoppage, making other routes more attractive. INTRSCT is however extremely inelastic, with an absolute value far less than 1. Small changes in the frequency of intersections along routes would have low influence on the relative utilities of route alternatives for a given OD-Pair as they may already exhibit similar frequencies of intersections. This was also an expected result as INTRSCT has a lower significance in the model compared to the other variables.

Based on the above discussion, it is inferred that Minimum Travel Time, Proportion of Freeways, and Proportion of Highway 401 are the best determinants of route choice, followed by the frequency of diesel stations, and number of intersections along a route.

Given that the elasticities of the variables and the much greater influence of the TTMIN variable on route choice, a question arises if route choice can be sufficiently predicted based on TTMIN only. If route choice can be predicted accurately using only TTMIN, then all other variables can be removed, reducing the route choice model to a simple time based route assignment model.

The results of this model is shown in Table 7-6 using the CF threshold of 0.65. The corresponding scatterplot visualizing predicted frequency and observed frequency is shown in Figure 7-6.

Table 7-6: Model results based on TTMIN only

Variable	**Coefficient**	**T-Statistic**
TTMIN	-2.852***	-111.32
CF	.51264***	8.68
LL(0)	-38523.490	
LL(β)	-21939.528	
ρ^2	0.430	
No. of Observations	1502/470	

***** indicates the parameter is statistically significant with 99% confidence**

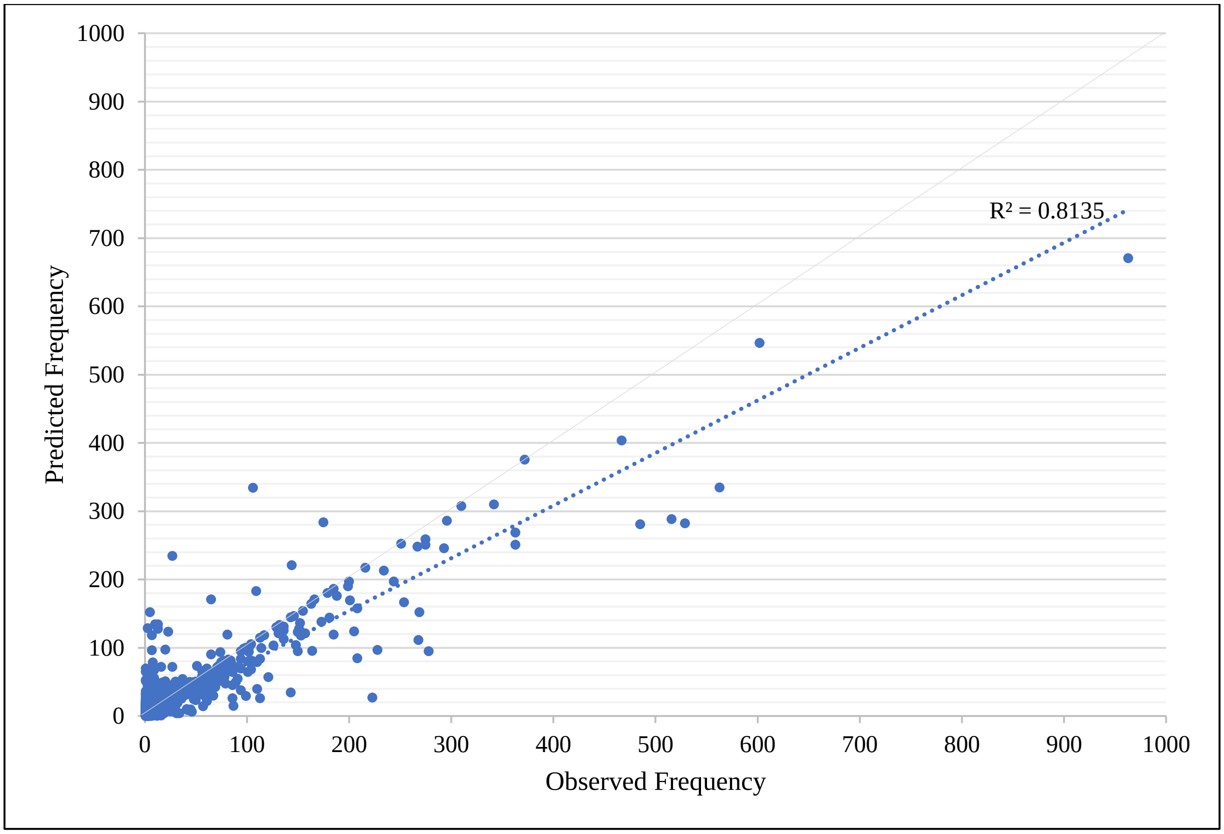


Figure 7-6: Model predicted frequency vs. observed frequency of route choice based on TTMIN only

The results indicate that the model performance deteriorates without other variables. For example, the R^2 value has reduced from 0.96 in Figure 7-5 to 0.81 in Figure 7-6. It is also apparent that the model based on TTMIN only is under-predicting route choice in the dataset as indicated by its trend line skewing below the y=x line. Therefore, the inclusion of the infrastructure based variables is justified despite their relatively lower elasticities. However, it should be noted that the Minimum Travel time does not account for congestion at different times of the day. Therefore, the average travel time variable is explored in Section 7.5.

7.5 Alternative Model 1: Using Average Travel Time

Congestion has been noted as a potential implication of the expected growth of truck volumes in Ontario based on observed trends. However, the route choice model derived in this section is based on the minimum observed travel time, which is similar to travel time without any congestion. The route reliability variables were also dropped from the model due to inherent biases in the dataset previously discussed. Therefore, to account for congestion in route choice, it may be suitable to estimate a model based on the observed average travel time variable, TTIME, in place of the TTMIN variable as it accounts for travel time variations along a route due to congestion or the urban routings each individual trip.

An alternative C-logit model based on observed average travel time is provided in Table 7-7, and its corresponding scatterplot is provided in Figure 7-7.

Table 7-7: Results for alternative model based on Average Travel Time

Variable	**Coefficient**	**T-Statistic**
TTIME	-.89879***	-42.5
FWP	1.69442***	33.93
FWP401	2.52498***	54.47
DIESEL	.26787***	79.8
INTRSCT	-.00551***	-18.57
CF	.76824***	10.36
LL(0)	-38523.49	
LL(β)	-20452.16	
ρ^2	0.47	
No. of Observations	32,645	

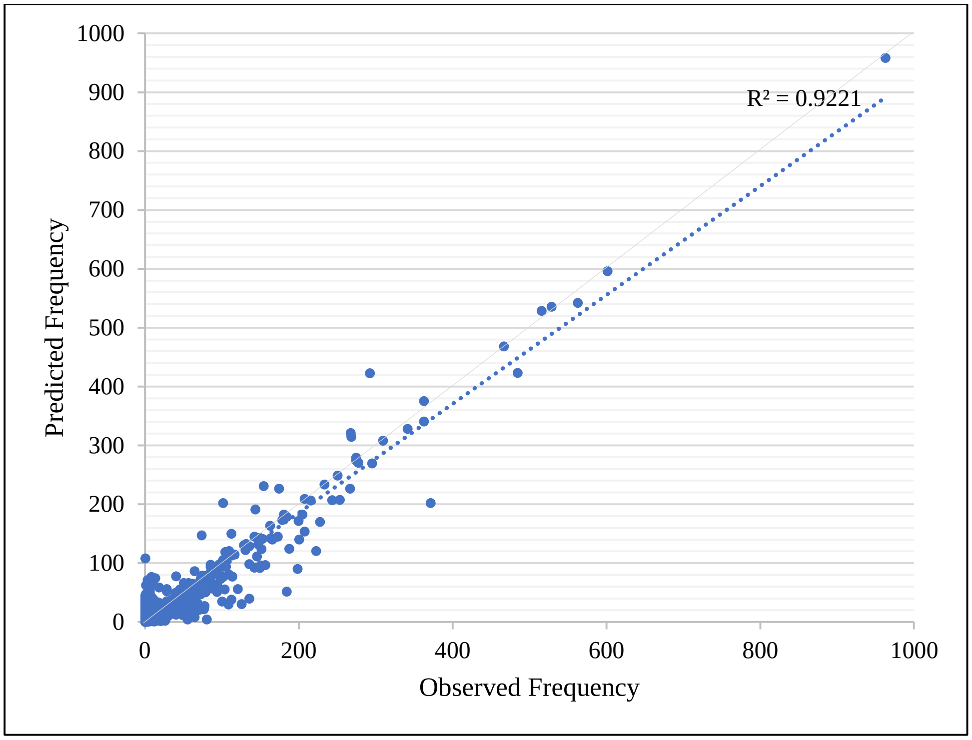


Figure 7-7: Average travel time based model predicted frequency v.s. observed frequency

As shown by Figure 7-7, the route choice model based on observed average travel time does not predict route choice as accurately as the minimum travel time based model. This suggests that drivers tend to select routes based on a best-case scenario travel time of a route without accounting for potential expected or unexpected congestion. However, this model may be more useful for infrastructure planning and policy purposes as it can be used when assigning trucks based on a user equilibrium traffic assignment.

7.6 Alternative Model 2: Adjusting for Under-predictions

As seen from Figure 7-5, the model tends to under-predict the route choice data despite the removal of a visible outlier. While the variables contributing to this general effect are unknown, a dummy variable representing the model tendency to under predict the frequencies of specific routes can be incorporated to the route choice model presented in Table 7-4.

A dummy value of 1 or 0 is assigned to the routes that have been under-predicted by more than 40 trips. Through this method, 15 such routes have been identified that are under-predicted by more than 40 trips.

The resulting variable, 'UNDERPREDICTION', is imported to NLogit along with the other variables from Table 7-4 and a C-logit model is estimated. The results of this model are shown in Table 7-8, and the corresponding scatterplot is provided in Figure 7-8.

Table 7-8: Results for alternative model including under-prediction dummy variable

Variable	Coefficient	T-Statistic
TTMIN	-1.633***	-66.71
FWP	1.29***	23.95
FWP401	1.948***	38.17
DIESEL	.203***	58.54
INTRSCT	-.002***	-7.15
CF	.160**	1.97
UNDERPREDICTION	1.72***	39.21
LL(0)	-38523.49	
LL(β)	-16489.28845	
ρ^2	0.572	
No. of Observations	32,645	

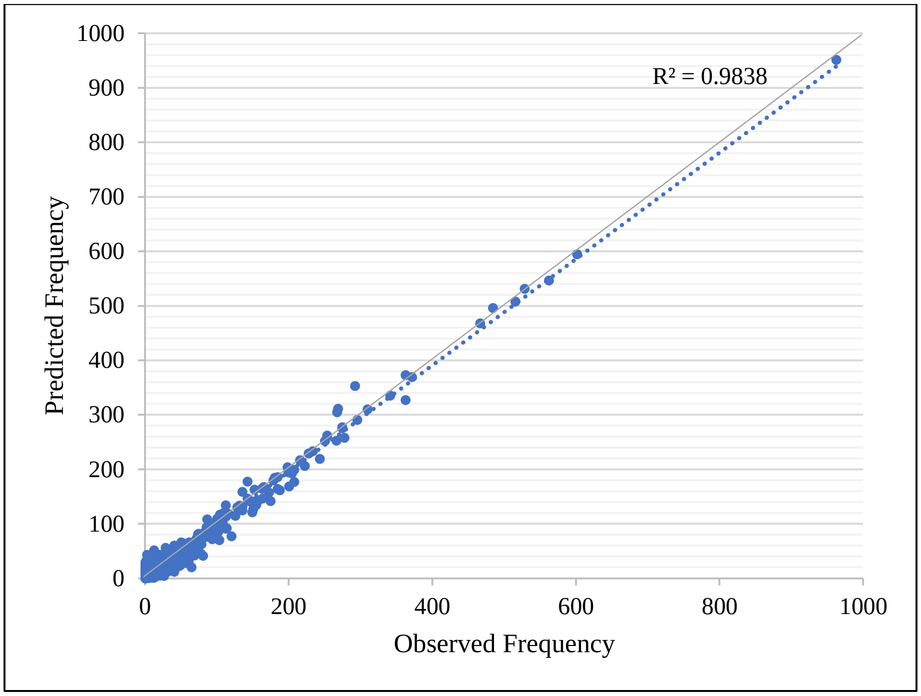


Figure 7-8: Predicted v.s. observed frequency plot for Alternative model 2

Table 7-8 and Figure 7-8 both show improvement of the model in terms of ρ^2 and R^2 values with the inclusion of the dummy variable accounting for under-prediction.

A close examination of the 15 identified under-predicted routes shows the majority of them to be related to Peel Region, which is a relatively large trip generator in Ontario. This explains the large number of trips associated with these routes and subsequent under-prediction by more than 40 trips by the model. Some of these errors are also caused by specific routes having higher observed frequencies due to their nearness to an origin or destination point within a zone despite having lower utility among other routes. This reasoning is further described in Section 9.2 and illustrated in Figure 9-1.

While the model estimated in Table 7-8 apparently performs better, the remainder of this thesis uses the final model shown in Table 7-4.

Chapter 8: Scenario Testing using Route Choice Model

In this section, the C-Logit model from the previous chapter shown in Table 7-4 is tested by introducing several different scenarios. These scenarios will test the impact of a change to the minimum travel time. The emphasis on travel time was selected due to the high elasticity of this variable, which indicates a strong effect on route choice when compared to all model variables.

8.1 Scenario #1: Increased Travel Time on Highway 401

It has been previously established in Section 5.3 that Highway 401 is the busiest roadway in Canada. Furthermore, truck drivers have been regarded as first and foremost time-minimizers, which was confirmed through the elasticity analysis in Section 7.4. Based on these aspects, a plausible scenario to explore is one where travel time along Highway 401 were to increase due to conceivable reasons such as construction delays.

In theory, severe delays along Highway 401 may divert drivers onto other routes which offer lower travel times and hence greater utility. To simulate this scenario, the TTMIN variable was adjusted to capture a proportional increase of travel time along Highway 401, as shown in Equation 8-1.

$$TTMIN_{s1} = TTMIN_{existing} \times (1 + \alpha(FWP401)) \tag{8-1}$$

Where

- $TTMIN_{s1}$ is the adjusted minimum travel time along a route identified for scenario 1
- $TTMIN_{existing}$ is the current minimum travel time along a route in the current model dataset
- *FWP401* is the proportion of Highway 401 comprising a given route.
- α is a variable utilized to increase the TTMIN based on the proportion of Highway 401 for a given route.

For this scenario testing, the variable α is set to 100% resulting in TTMIN effectively doubling in time for the portion of a route traveling along Highway 401. Note that this is a simplification that can be conducted quickly on all routes due to the aggregate nature of the calculation. A more accurate calculation could be performed by calculating and adjusting the travel time for each route segment individually, but is beyond the scope of the thesis.

The resulting utilities under the adjusted travel time are calculated along with the route choice frequencies predicted by the C-Logit model. The new frequencies are then compared to the previously estimated route selections.

To visualize the results of this scenario, some example OD-Pairs are shown with their existing and future scenario truck route frequencies as predicted by the model. Individual routes between OD-Pairs are color coded, and their frequencies are visualized by the thickness of the route. Note how the thickest (most popular) route changes in the existing and future scenarios in each example.

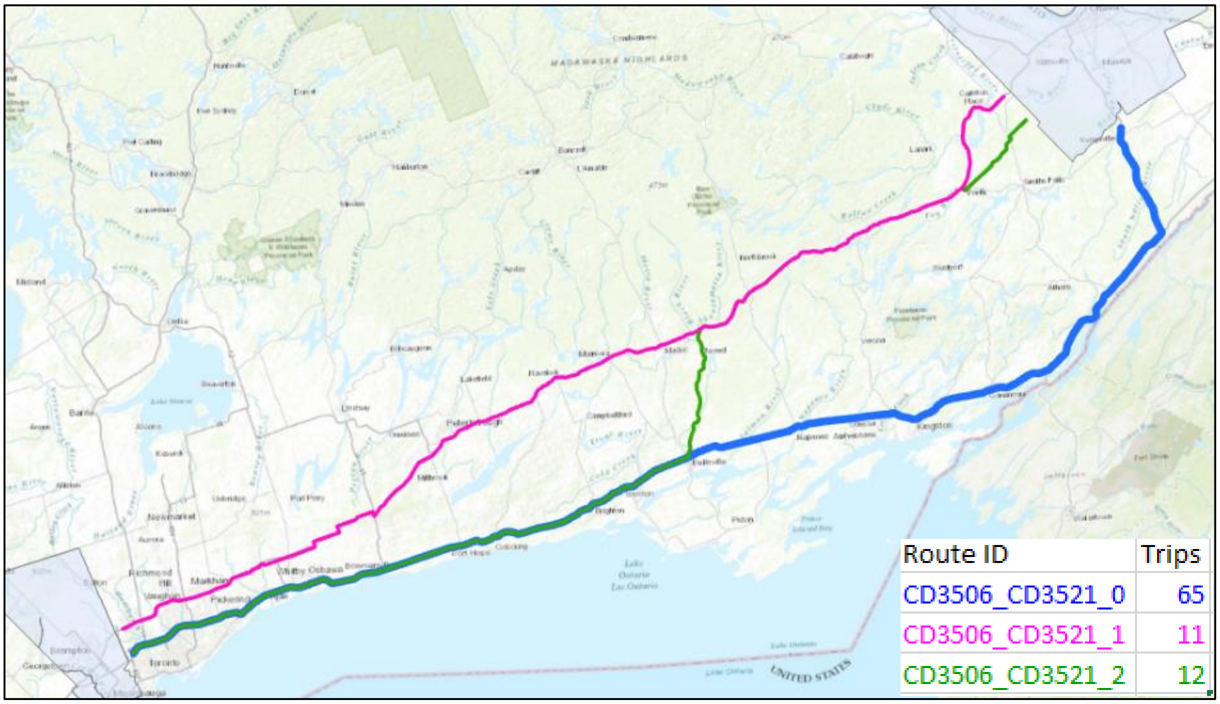


Figure 8-1: Observed route choice between Peel Region and Ottawa

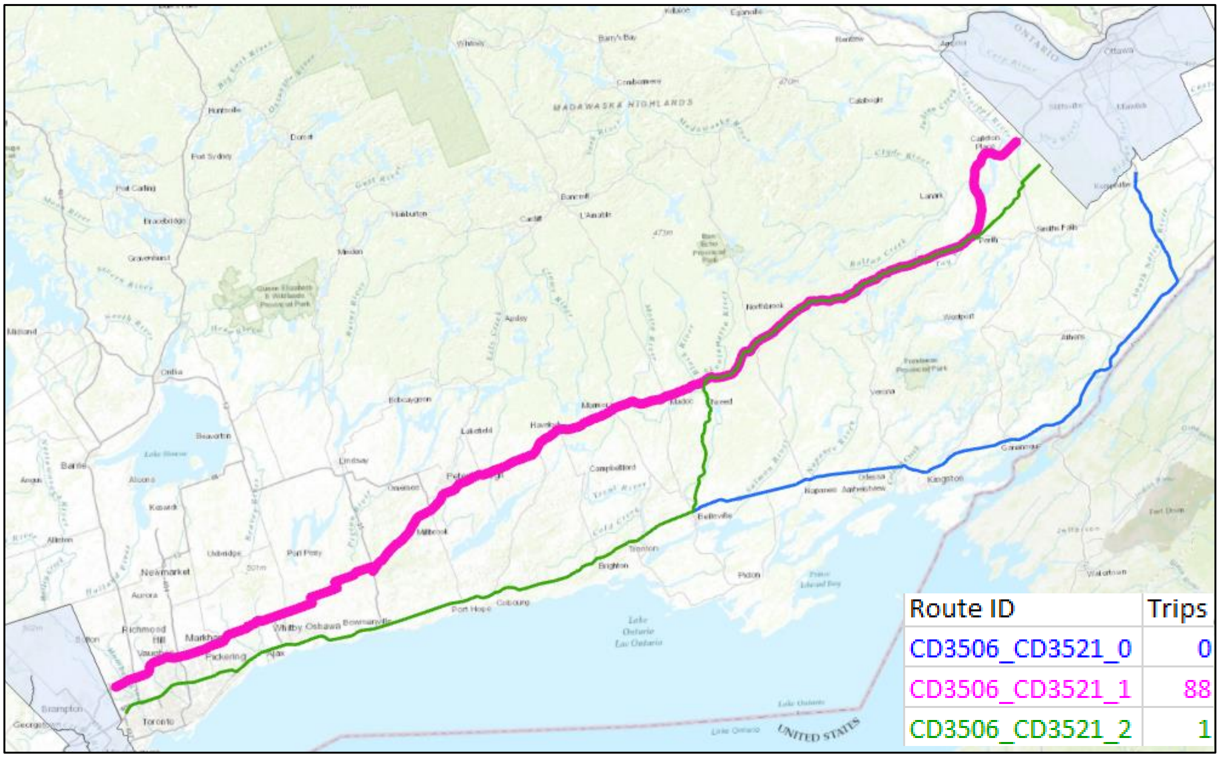
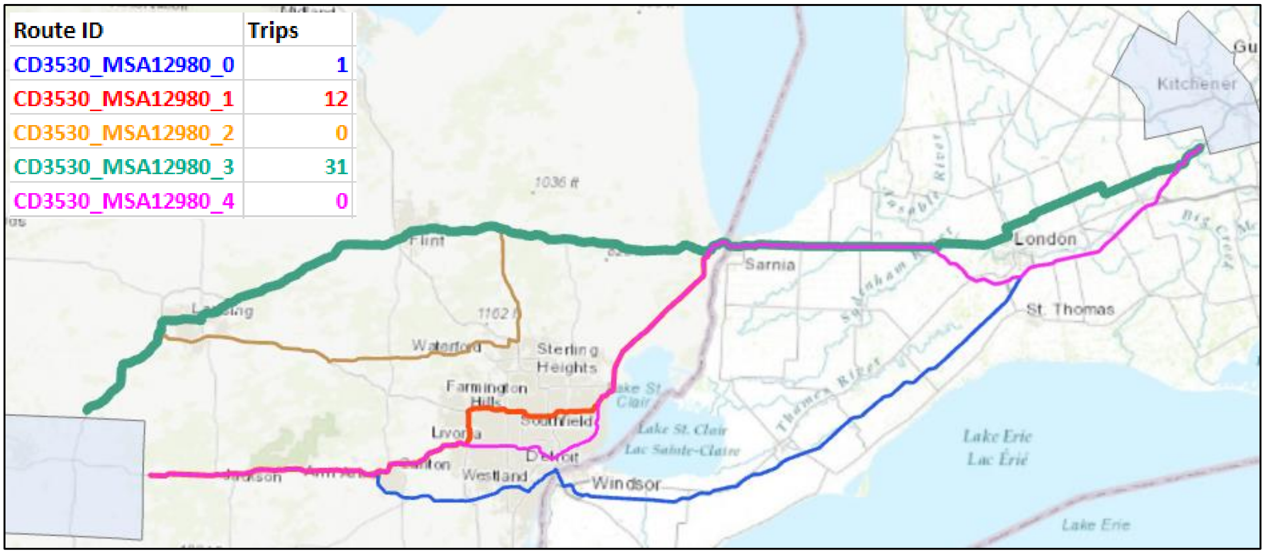


Figure 8-2: Scenario 1 route choice between Peel Region and Ottawa

Figure 8-1 shows the existing route choice between Peel Region and Ottawa, where the most popular route is along Highway 401 and has a frequency of 65 truck trips. In Scenario 1, shown in Figure 8-2, the route along Highway 7 exhibits the highest frequency due to the additional travel time on Highway 401.

Route ID	Trips
CD3530_MSA12980_0	29
CD3530_MSA12980_1	4
CD3530_MSA12980_2	0
CD3530_MSA12980_3	10
CD3530_MSA12980_4	1

Figure 8-3: Observed route choice between Region of Waterloo and Battle Creek, MI



Route ID	Trips
CD3530_MSA12980_0	1
CD3530_MSA12980_1	12
CD3530_MSA12980_2	0
CD3530_MSA12980_3	31
CD3530_MSA12980_4	0

Figure 8-4: Scenario 1 route choice between Region of Waterloo and Battle Creek, MI

Another example is shown for a cross border OD-Pair with a relatively long trucking distance between the Region of Waterloo, Ontario, and Battle Creek, Michigan. For the observed trips shown in Figure 8-3, the most popular route is along Highway 401 between Region of Waterloo and the Windsor-Detroit Border as shown by the blue path. In the Scenario 1 (Figure 8-4), the majority of truck trips completely bypass Highway 401 in favor of Regional Roads and Highway 402 to use the Sarnia border.

A macroscopic illustration of the routing changes incurred due to Scenario #1 is shown in Figure 8-5. Each OD-Pair is represented by a straight line connecting the relevant origin and destination zones in the route choice dataset. The color shading of the zones along with the thickness of the connecting lines indicate the percent of trips re-routed between the OD-Pairs due to additional travel time along Highway 401 in Scenario 1. For example, the thick line representing the OD-Pair of York Region and Columbus, Ohio shows a 90% change in truck routing. The darker shade of these two zones also suggests that they are subject to high degrees of routing changes. Similarly, the thin line connecting Peel Region and Kawartha Lakes indicates a low change in routing decisions under Scenario #1. Note that Peel Region is much darker compared to Kawartha Lakes, suggesting that there are other zones that incur substantial changes in route choice for Peel Region related trips. These can be identified by thicker lines, such as the ones between Peel Region and Levis, Quebec, or Peel Region and Sherbrook, Quebec.

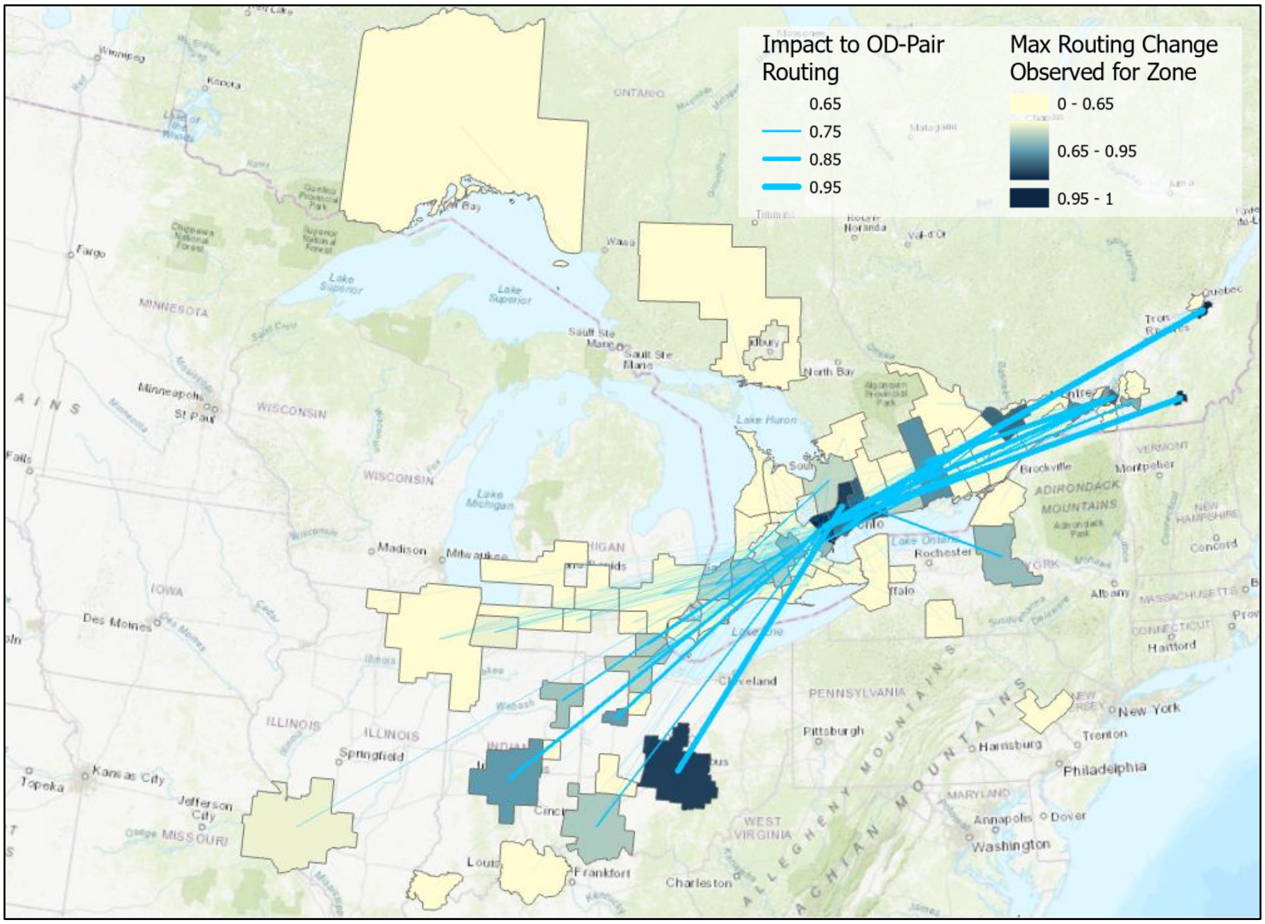


Figure 8-5: Macroscopic illustration of Scenario #1 effects on route choice

Figure 8-5 suggests the zones that are most affected by Highway 401 as illustrated by the change in route when travel times are increased.

While this scenario is hypothetical, the model responds appropriately to the changes incurred to the established routes and reasonably predicts future trucking patterns under such circumstances. In the same manner, this model can be used to predict trucking patterns in other possible scenarios, such as one where the entire freeway network of Ontario was subject to higher travel times, with all other variables remaining qual. This scenario is shown in the next section.

8.2 Scenario #2: Increased Travel Time on all Freeways

Another, more extreme scenario to explore is one where the entire freeway network of Ontario experiences higher travel times compared to current conditions, forcing truck drivers to divert to more local arterial routes. Using a similar method as in Scenario #1, Equation 8-1 can be adjusted to include the proportion of other freeways to the factor used to increase the travel time of a route and its subsequent change in popularity. The adjustment for travel time in this scenario is shown in Equation 8-2.

$$TTMIN_{s2} = TTMIN_{existing} \times (1 + \alpha(FWP + FWP401)) \tag{8-2}$$

Where

- $TTMIN_{s2}$ is the adjusted minimum travel time along a route identified for Scenario 2.
- $TTMIN_{existing}$ is the current minimum travel time along a route in the current model dataset.
- *FWP* is the proportion of Freeway comprising a given route.
- *FWP401* is the proportion of Highway 401 comprising a given route.
- α is a variable utilized to increase the TTMIN based on the proportion of Freeways and Highway 401 for a given route.

For comparison, the existing truck route frequencies for the example OD-Pair of Ottawa and Longueuil are shown in Figure 8-6. The figure shows that majority of observed truck trips occur along Highway 417. There are also five truck trips using the route along Highway 401.

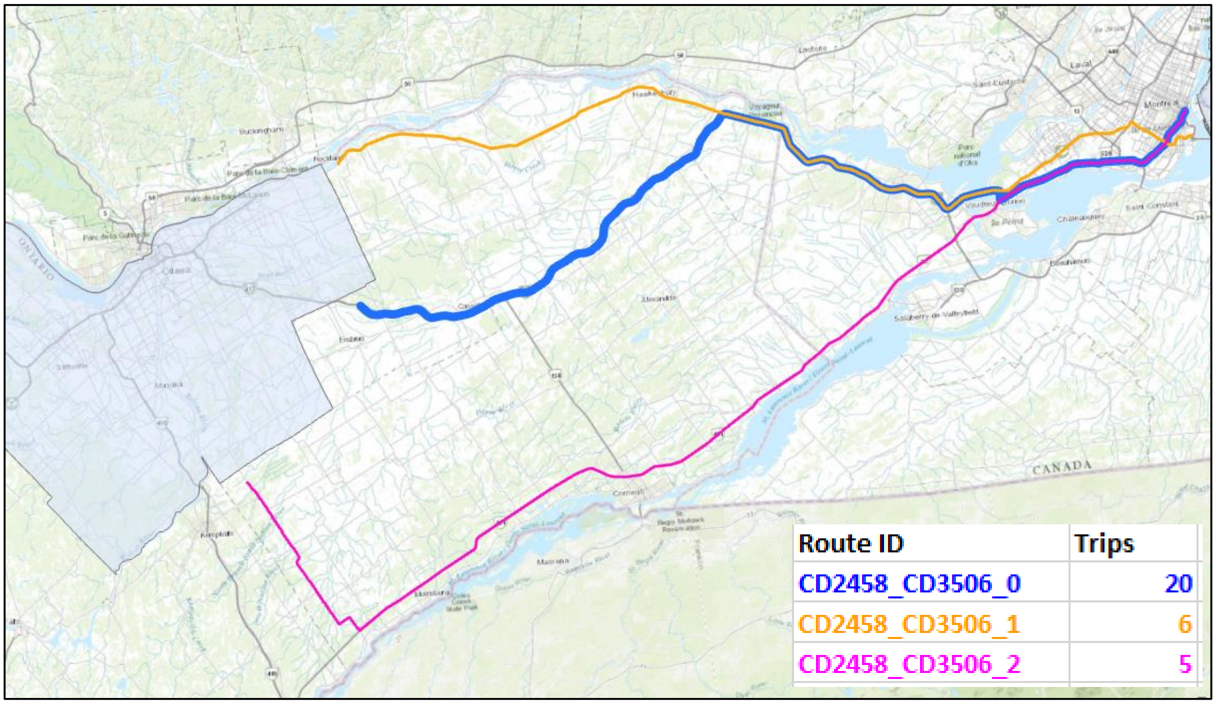
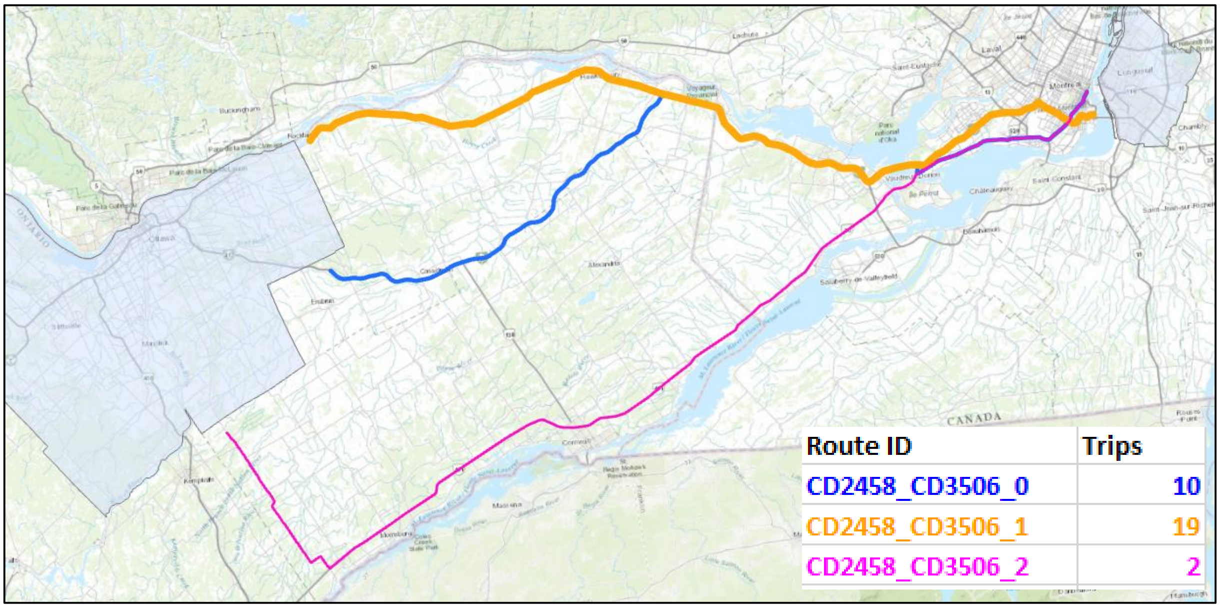


Figure 8-6: Observed route choice between Ottawa and Longueuil, QC

Figure 8-7 shows the truck frequencies along the routes between Ottawa and Longueuil under Scenario 2 conditions. As expected, the most dominant route becomes the one with the least proportion of Freeways and Highway 401, and is indicated by the thick golden line, while the truck frequencies for the other two routes have reduced.



Route ID	Trips
CD2458_CD3506_0	10
CD2458_CD3506_1	19
CD2458_CD3506_2	2

Figure 8-7: Scenario 2 route choice between Ottawa and Longueuil, QC

As with Scenario #1, a macroscopic illustration of the routing changes incurred due to Scenario #2 is shown in Figure 8-8. The color shading of the zones along with the thickness of the line's connecting them indicate the percent of trips re-routed between the OD-Pairs due to increases in travel time along Ontario's freeway network.

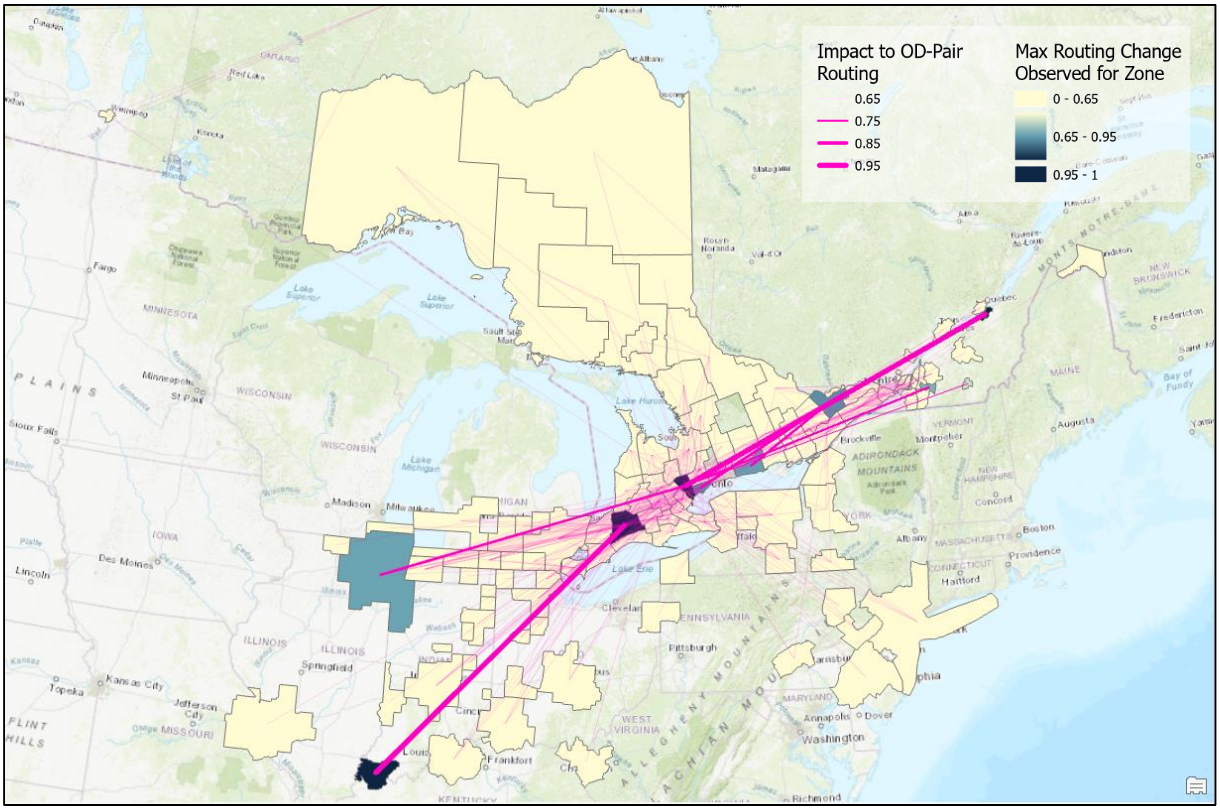


Figure 8-8: Macroscopic illustration of Scenario #2 effects on route choice

As expected, more OD-Pairs are being impacted in Scenario #2 than in Scenario #1 in terms of re-routing due to the larger coverage of the freeway network.

However, the amount of routing changes observed in Scenario #2 are lower than in Scenario #1. This suggests a lack of viable alternatives for drivers to re-route to if the freeway network is subject to severe delays, indicating that most zones are connected by freeways. Scenario #1 exhibited more routing changes as other freeways served as viable alternatives between zones, whereas few OD-Pairs could offer alternatives with lower freeway usage in Scenario #2.

The Ontario zones most impacted by the two scenarios presented in this thesis are Peel Region, City of Toronto, Middlesex County, York Region, Durham Region and Ottawa, and Hastings County. All these zones exhibit more than an 80% re-routing of truck trips for select OD-Pairs in either scenario. This is expected as these zones lie along Highway 401 or other freeways. These zones have been ranked in Table 8-1 based on their highest exhibited routing change in either scenario.

Table 8-1: Ranking of Ontario zones impacted by scenarios

Rank	Zone	Scenario 1 Impact	Scenario 2 Impact
1	Peel Region	97%	96%
2	Middlesex County	72%	95%
3	York Region	92%	39%
4	Durham Region	88%	69%
5	Ottawa	86%	81%
6	Toronto	86%	81%
7	Hastings County	81%	54%

Two potential scenarios have been simulated in this chapter using the model to demonstrate its practical usage in long term transportation planning. Other types of scenarios involving proposed route alternatives can also be considered by adding them to choice sets between OD-Pairs and determining their utilities and probabilities of selection. This would provide useful insight on the adequacy of such proposed roadways and help identify their shortcomings.

Chapter 9: Conclusions

9.1 Thesis Summary

This thesis presents the development of a route choice model for long-haul truck trips connected to the Province of Ontario by origin or destination.

Fundamentally, strong planning can help promote the safety of vehicles by focusing on truck corridors and ensure that the roadway system allows businesses to stay competitive and increase the country's economic output. The 93% growth in AADTT observed on provincial highways over the past 29 years is expected to continue due to the recent trend for e-commerce shipping increasing demand for long-distance freight. This justifies a study of Ontario's freight route choice to better understand long-haul truck vehicle movements and ensure that future planning aligns appropriately to expected shipping patterns.

A literature review was completed to better understand of process of route choice modelling, including the following main themes:

- Data requirements such as map-matched GPS data
- Choice set generation methods and the Commonality Factor (CF);
- Route choice model estimation methods; and
- Influential factors that explain route choice such as route performance, reliability, geometry, and facilities.

A map matched truck trip dataset of Ontario consisting of 50,431 trips between 840 unique origins and destinations (OD-Pairs) was processed in ArcGIS for choice set generation and analyzed in NLogit for model estimation. Choice set generation was conducted using an iterative algorithm developed in ArcGIS model builder that defined unique routes for every OD-Pair based on the

Commonality factor. Based on a CF threshold of 85%, all truck trips were assigned into 2483 unique routes, where each OD-Pair exhibited up to 16 unique routes. The number of unique routes between an origin and destination was found to decrease with increasing average route length, supporting the findings of Luong et al. (2018) on route choice diversity.

Variables available for this thesis were processed using ArcGIS software due to their spatial nature. Two broad categories of variables were tested including the performance based variables and infrastructure based variables. The former performance variables were originally based on the observed GPS data, but were also later tested using outside data using historical travel time information originally sourced from 'HERE' and made available using ArcGIS Pro software. The model estimation showed that minimum travel time derived from the observed GPS data was found to be the most suitable performance variable to include in the final model (negative relationship). Other performance variables were not included due to high correlations. Additional variables in the final model were infrastructure based factors measured for each given route, including: the number of intersections (negative relationship); the proportion of Highway 401 (positive relationship); the proportion of other freeways (positive relationship); and the number of diesel stations (positive relationship). Elasticities calculated for these variables found the minimum travel time to have the largest impact on route choice as expected. The results for all other variables were relatively inelastic with values well below 1%. The elasticity analysis indicates that travel time is the single most important determinant of route choice as expected.

The choice set data originally included 2,200 routes imported to NLogit for model estimation. The original choice set was developed with a maximum correlation of 85% between alternatives as measured by the commonality factor (CF=0.85). Based on Cascetta et al., (1997), the Commonality

Factor of each route was also input as an independent variable to modify the MNL model into a C-Logit Model.

A sensitivity analysis of the commonality factor was undertaken by regenerating choice sets for each OD-Pair in the trip dataset based on uniqueness threshold values of 0.95, 0.90, 0.80, 0.75, 0.70, and 0.65. C-Logit models were estimated using the final model variables and showed consistency in terms of significance and influence. The preferred C-Logit model based on performance was based on a low commonality threshold value of 0.65 as it exhibited the highest ρ^2 value. This is consistent with Cascetta's recommendation noted in Telgen, 2010 that model performance improves with stricter thresholds of uniqueness. This results from a simpler dataset since fewer alternatives are generally available for a given OD-Pair. However, there did not appear to be any other consistent trends in the model results as the CF threshold variable was changed, indicating that the threshold value had a negligible on the final model outcomes.

Alternative models based on average travel time to account for congestion, and dummy variables accounting for under-predictions have also been presented. While these models may be advantageous from a policy perspective, the preferred final C-logit model is retained based on its practicality and consistency.

Scenario testing was conducted on the final route choice model to test the changes caused by the travel time variable since it was shown to have the most elastic response. Two scenarios were created for this purpose by adjusting the travel time for routes utilizing Highway 401 and other freeways since these are currently popular routes in Ontario. In these hypothetical scenarios, the model reasonably identified changes in trucking patterns between OD-Pairs by predicting that route choice would shift from these roadways to other alternatives in the presence of increased travel times as expected. The results of this analysis also identified the zones most likely to be connected to changing route patterns if the performance of these major highways deteriorates in the future.

9.2 Limitations and Recommendations

Some of the limitations associated with this research are discussed below to provide appropriate context for individuals planning to utilize the results for their own purposes.

This thesis used observational data to derive the choice set for modelling instead of the theoretical constructs of alternatives used by some other studies. A positive outcome of the observational approach is that there is a strong confidence that the observed routes are feasible choices since they have been used in the past. However, one drawback is that zone pairs with only one observed alternative cannot be included in the model since there is no choice available. In this thesis, 263 OD-Pairs accounting for 13,320 trips were removed from the modelling dataset as they only had one route between them. The loss of this information may bias the overall results since cases with an extremely dominant choice are not included. However, situations with only one viable route are easy to forecast; therefore, the complexity of a route choice model should not be needed in such circumstances. These situations may be useful to analyze in the future due to the lack of route redundancy implied by only one feasible route.

Another potential drawback to the utilization of observed routes in developing a choice set is that they must be aggregated to some degree. For this thesis, the trips were aggregated to a given origin-destination zone pair, with different routes defined by the choice set algorithm travelling between the established boundaries of census divisions and therefore limiting the route choice to the long-haul portion of a trip. However, this method does not capture truck trajectories within a given origin or destination zone, which may influence route selection. For example, if an origin facility is located on the southern edge of a start zone as shown in Figure 9-1, a route alternative with convenient access to this portion of the zone is more desirable over a route preferred in terms of performance. Therefore, it is recommended that local origin and destination proximities to established routes should be measured and considered as variables even where the local urban portion of route choice is not within the project scope. This would require the model to be estimated with individual observations instead of the aggregate format used in this thesis. A discussion of the aggregation can be found in Section 6.2.

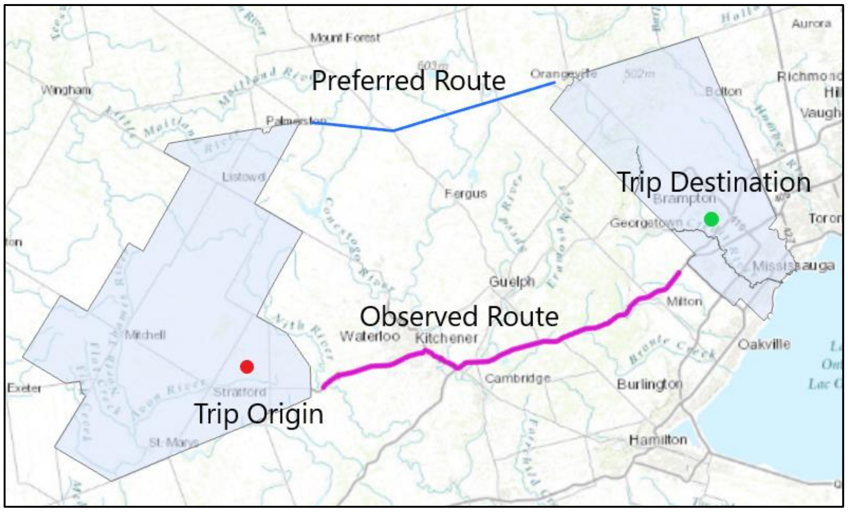


Figure 9-1: Trip local origin proximity to routes

Trip data in each route in the modelling dataset was analyzed to calculate route performance measures including average travel time, minimum travel time, maximum travel time and 95th percentile travel time. To measure route reliability, these metrics were used to calculate each route's Travel Time Index, Buffer Time Index, and Planning Time index to account for expected and unexpected delays as shown by (Russell, 2014). While route variability is important, there was not enough data present to ideally calculate items such as the 95th percentile travel time as many routes exhibited only one trip. Furthermore, the time-based variables included considerable urban routing which is subject to high variability. Consequently, the reliability indices were unrepresentative of the routes defined between the census zone boundaries. However, it is recommended for future studies in route choice to consider one of these reliability measures in the route choice model if appropriate data is available.

9.3 Implications of Research

This thesis adds to the literature concerning long haul freight route choice in Canada. The algorithms developed through this work have been designed to take any observed dataset of truck trip trajectories, and result in the generation of observed choice sets. The methodological procedure is therefore highly valuable. Furthermore, the sensitivity analysis of the Commonality Factor revealed that the user-defined threshold for CF and subsequent choice set ultimately do not have a significant impact on the predictive power of the model while confirming that lower CF thresholds are preferable. This was a novel result not widely available in the academic literature.

This research additionally has practical applications for freight operations and road network planning.

For example, there are several applicable findings of this thesis that can inform various kinds of decisions made my transportation authorities such as the Ontario Ministry of Transportation. Firstly, the choice set generation process for a given origin and destination can immediately reveal the adequacy of connectivity between the given zones. A lack of multiple feasible routes can imply the need for additional route options. The observed routes from this thesis could also be used to determine appropriate emergency detours for trucks in cases where a single route is disrupted.

The factors arising from the route choice model in this thesis can help develop adequate route in the future. For instance, if a roadway needs to be designated as a freight corridor, it should exhibit the significant characteristics such as freeway composition and availability of diesel pumps. But most importantly, the travel time should be sufficient since this is the most dominant variable in the model as determined by variable elasticities.

When conducting safety assessments, collision data can be overlaid onto routes identified in the choice set for this thesis to provide initial screening of dangerous routes for long-haul trips. Routes exhibiting high frequencies of truck related collisions can then be prioritized for safety improvements.

Finally, the route choice model can be instrumental for long term planning as it is developed using real observed truck trip data. By applying appropriate growth statistics, new or existing routes can be assessed to forecast future travel patterns for long-haul trucks. When integrated with short-haul trucks and passenger cars, such planning is integral to ensuring that our infrastructure is capable of satisfying the transportation needs of future generations.

References

Abkowitz, M., & Cheng, P. D.-M. (1988). Developing a risk/cost framework for routing truck movements of hazardous materials. *Accident Analysis & Prevention*, 39–51.

Apronti, D. T., Saha, P., Moomen, M., & Ksaibati, K. (2019). Truck safety evaluation on Wyoming Mountain Passes. *Accident Analysis & Prevention*, 342–349.

Aultman-Hall, L., Hill, M. L., & Agent, K. R. (1999). Methodology for Evaluating Large Truck Access to Intermodal and Other Facilities. *Transportation Research Record*, 61–68.

Bauer, K., & Harwood, D. (1998). Statistical Models of At-Grade Intersection Accidents—Addendum. *Virginia: Federal Highway Administration.*

Ben-Akiva, M., Toledo, T., Santos, J., Marzano, V., Lee, Y. J., Zhao, F., & Cox, N. (2015). Freight data collection using GPS and web-based surveys: Insights from US truck drivers' survey and perspectives for urban freight. *Case Studies on Transport Policy*, *4*, 1–15.

Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B*, *36*, 577–592.

Cascetta, E., Russo, F., & Vitetta, A. (1997). STOCHASTIC USER EQUILIBRIUM ASSIGNMENT WITH EXPLICIT PATH ENUMERATION: COMPARISON OF MODELS AND ALGORITHMS. *8th IFAC/IFIP/IFORS Symposium on Transportation Systems 1997 (TS '97), Chania, Greece, 16-18 June*, 1031–1037.

Dalumpines, R., & Scott, D. M. (2011). GIS-based Map-matching: Development and Demonstration of a Postprocessing Mapmatching Algorithm for Transportation Research. *Advancing Geoinformation Science for a Changing World*, *1*, 101–120.

Daniel, J., & Chien, S. I.-J. (2004). Truck Safety Factors on Urban Arterials. *Journal of Transportation Engineering*, 742–752.

Dijkstra, A., & Drolenga, H. (2008). Safety effects of route choice in a road network: Simulation of changing route choice. *Leidschendam : SWOV Institute for Road Safety Research*.

Dong, C., Richards, S. H., Huang, B., & Jiang, X. (2013). Identifying the factors contributing to the severity of truck-involved crashes. *Injury Control and Safety Promotion*, *22*(2), 116–126. http://dx.doi.org/10.1080/17457300.2013.844713

Fleger, S. A., Haas, R. P., Trombly, J. W., III, R. H., Noltenius, J. E., Pécheux, K. K., & Chen, K. J. (2002). *Study of Adequacy of Commercial Truck Parking*.

Freitas, L. M. de. (2018). *A recursive logit multimodal route choice model* [Swiss Federal Institute of Technology Zurich]. https://ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/publications/students/501-600/sa600.pdf

Gingerich, K. (2017). *Studying Regional and Cross Border Freight Movement Activities with Truck GPS Big Data* [University of Windsor]. https://pdfs.semanticscholar.org/b16d/d0d69470f13accf20a07de18a5e0dd337b45.pdf

Gingerich, K., & Maoh, H. (2017). Big Data Analysis to Measure Delays of Canadian Domestic and Cross-Border Truck Trips. *Transportation Research Board 96th Annual Meeting*. https://trid.trb.org/view/1437678

Gingerich, K., & Maoh, H. (2019). The role of airport proximity on warehouse location and associated truck trips: Evidence from Toronto, Ontario. *Journal of Transport Geography*, *74*, 97–109. https://doi.org/10.1016/j.jtrangeo.2018.11.010

Golob, T. F., & Regan, A. C. (2003). Truck-Involved Crashes and Traffic Levels on Urban Freeways. *UC Berkeley: University of California Transportation Center*. https://escholarship.org/uc/item/6w37f3jd

Harwood, D. W., Vinerfl, J. G., & Russell, E. R. (1993). Procedure For Developing Truck Accident And Release Rates For Hazmat Routing. *Journal of Transportation Engineering*, 189–199.

Hasnine, M. S., & Habib, K. N. (2018). What about the dynamics in daily travel mode choices? A dynamic discrete choice approach for tour-based mode choice modelling. *Transport Policy*, *71*, 70–80. https://doi.org/https://doi.org/10.1016/j.tranpol.2018.07.011

Jan, O., Horowitz, A. J., & Peng, Z.-R. (2000). Using GPS Data to Understand Variations in Path Choice. *Journal of the Transportation Research Board*, *1725*(1), 37–44.

Joshua, S. C., & Garber, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation Planning and Technology*, *15*, 41–58.

Kaneko, N., Oka, H., Chikaraishi, M., Becker, H., & Fukuda, D. (2018). Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-based Recursive Logit Model Featuring Link Awareness. *Transportation Research Procedia*, *34*, 251–258. https://doi.org/10.1016/j.trpro.2018.11.039

Knorring, J. H., He, R., & Kornhauser, A. L. (2005). Analysis of Route Choice Decisions by Long-Haul Truck Drivers. *Journal of the Transportation Research Board*, *1923*, 46–60.

Kunchev, L. (2017). Methodology for selection the truck route. *Engineering for Rural Development*, *16*, 263–272. https://doi.org/10.22616/ERDev2017.16.N052

Li, J., & Lai, X. (2019). Modelling travellers' route choice behaviours with the concept of equivalent impedance. *Transportation*, *46*, 233–262.

Li, Y., Le, J. T., Middleton, D. R., & Quiroga, C. A. (2012). Mapping Oversized and Overweight Truck Routes with Procedure Based on Geographic Information Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 8–16.

Luong, T. D., Tahlyan, D., & Pinjari, A. R. (2018). Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida. In *Transportation Research Record*. https://doi.org/10.1177/0361198118784175

Meyer, M. D., & Chu, H.-C. (2009). Estimated Crash Benefits Related to Truck-Only Toll Lanes. *Transportation Research Board 88th Annual Meeting*.

Ministry of Transportation. (2019). *Preliminary 2019 Ontario Road Safety Annual Report Selected Statistics*. http://www.mto.gov.on.ca/english/publications/pdfs/preliminary-2019-orsar-selected-statistics.pdf

Morley, C. L. (1994). Experimental destination choice analysis. *Annals of Tourism Research*, *21*(4), 780–791. https://doi.org/https://doi.org/10.1016/0160-7383(94)90083-3

MTO iCorridor. (2019). *Historical AADT & AADTT*. Ontario Ministry of Transportation ICorridor. https://www.arcgis.com/home/item.html?id=https://services.arcgis.com/6iGx1Dq91oKtcE7x/arcgis/rest/services/Historical_AADT/FeatureServer

Mussa, R. (2004). Safety and Operational Evaluation of Truck Lane Restriction. *Journal of the Transportation Research Forum*, 117–127.

Nevland, E. A., Gingerich, K., & Park, P. Y. (2020). A data-driven systematic approach for identifying and classifying long-haul truck parking location. *Transport Policy*, *96*, 48–59.

https://doi.org/10.1016/j.tranpol.2020.04.003

Oka, H., Hagino, Y., Kenmochi, T., Tani, R., Nishi, R., Endo, K., & Fukuda, D. (2017). Predicting travel pattern changes of freight trucks in the Tokyo Metropolitan area based on the latest large-scale urban freight. *Transportation Research, Part E*, 1–20.

Ontario Data. (2015). *Ontario Road Network: Road Net Element*. Ministry of Natural Resources and Forestry. https://data.ontario.ca/dataset/ontario-road-network-road-net-element

Oron-Gilad, T., & Ronen, A. (2007). Road characteristics and driver fatigue: A simulator study. *Traffic Injury Prevention*, *8*(3), 281–289. https://doi.org/10.1080/15389580701354318

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, *2*, 65–100.

Prato, C. G., & Bekhor, S. (2007). Modeling Route Choice Behavior: How Relevant Is the Composition of Choice Set? *Transportation Research Record*, *2003*, 64–73. https://journals.sagepub.com/doi/10.3141/2003-09

Private Motor Truck Council of Canada. (2019). *Guidelines For Security In Trucking – A Checklist*. http://www.pmtc.ca/CMFiles/SecurityCheckListJul2018.pdf

Ramming, M. S. (2002). *Network Knowledge and Route Choice* [Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/49797

Robertson, S., & Aultman-Hall, L. (2001). Incorporating Truck Crash Modeling into a Methodology for Evaluating the Relative Need for Truck Route Improvements. *University of Kentucky*.

Rowell, M., Gagliano, A., & Goodchild, A. (2014). Identifying truck route choice priorities: the implications for travel models. *Transportation Letters*, *6*, 98–106.

Russell, M. (2014). *Travel Time Reliability and Level of Service* [Louisiana State University]. https://digitalcommons.lsu.edu/cgi/viewcontent.cgi?article=4746&context=gradschool_theses

Samuel, P. W., Robert, J. P., & Holguin-Veras, J. (2002). Truckways: A New Path Toward Safer and More Efficient Freight Transportation. *Reason Public Policy Institute*, *Policy Stu*. http://www.rppi.org/ps294.html

Scholars Portal. (2012). *Scholars GeoPortal*. http://geo2.scholarsportal.info/

Sobhani, A., Aliabadi, H. A., & Farooq, B. (2019). Metropolis-Hasting based Expanded Path Size Logit model for cyclists' route choice using GPS data. *International Journal of Transportation Science and Technology*, *8*(2), 161–175.

Statistics Canada. (2018a). *General Freight Trucking - 4841*. Summary - Canadian Industry Statistics. https://www.ic.gc.ca/app/scr/app/cis/summary-sommaire/484

Statistics Canada. (2018b). *Vehicle registrations*. The Daily. https://www150.statcan.gc.ca/n1/daily-quotidien/180615/dq180615e-eng.ht

Statistics Canada. (2020). *Retail e-commerce sales, unadjusted (x 1,000)*. 20-10-0072-01. https://doi.org/10.25318/2010007201-eng

Tahlyan, D., Luong, T. D., Pinjari, A. R., & Ozkul, S. (2017a). *Development and Analysis of Truck Route Choice Data for the Tampa Bay Region using GPS Data*.

Tahlyan, D., Luong, T. D., Pinjari, A. R., & Ozkul, S. (2017b). *Truck Route Choice Modeling using Large Streams of GPS Data* (Issue July). http://cait.rutgers.edu/files/CAIT-UTC-NC-32-

final.pdf%0Ahttps://cait.rutgers.edu/files/CAIT-UTC-NC-32-final.pdf%0Ahttps://rosap.ntl.bts.gov/view/dot/34888%0Ahttps://trid.trb.org/view/1504914

Telgen, M. G. (2010). *Realistic route choice modeling* [University of Twente]. https://essay.utwente.nl/59718/1/MA_thesis_M_Telgen.pdf

Train, K. (1993). *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. Cambridge: The MIT Press. https://eml.berkeley.edu/books/choice.html

Transport Canada. (2017). *Transportation in Canada 2016*. https://www.tc.gc.ca/media/documents/policy/comprehensive_report_2016.pdf

Transport Canada. (2020). *Lac-Mégantic bypass*. Government of Canada. https://www.tc.gc.ca/eng/lac-megantic-bypass.html

United States Census Bureau. (2016). *TIGER/Line Shapefile, 2016, nation, U.S., Primary Roads National Shapefile*. Department of Commerce. https://catalog.data.gov/dataset/tiger-line-shapefile-2016-nation-u-s-primary-roads-national-shapefile

Wang, Z., & Goodchild, A. V. (2014). GPS Data Analysis of the Impact of Tolling on Truck Speed and Routing. *Journal of the Transportation Research Board*, *2411*, 112–119.

Wright, P. H., & Burnham, A. (1985). Roadway Features and Truck safety. *Transportation Planning and Technology*, 263–269.

Appendices

Appendix A: Route Choice Literature Review Summary

Citation:
Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B*, *36*, 577–592.
Topic:
Choice Set Generation
Research Question:
Which choice set generation and route assignment methods are best
Data:
Literature
Methods:
Analysis of Route choice models Multinomial Logit and C-Logit, Empirical Analysis ,Stochastic User Equilibrium Assignment
Factors (and significance):
Commonality Factor, Goodness of Fit, C-Logit outperforms Probit Model

Citation:
Dalumpines, R. & Scott, D.M. 2011, "GIS-based map-matching: Development and demonstration of a post-processing map-matching algorithm for transportation research", Lecture Notes in Geoinformation and Cartography, pp. 101.
Topic:
To develop a post-processing map-matching algorithm in a GIS platform
Research Question:
Develop a map matching algorithm for transportation research applications such as route choice analysis
Data:
GPS-assisted time-use survey that involved nearly 2,000 households. Actual routes taken by household members who travelled to work by car were extracted using GPS data and GIS-based map-matching algorithm.
Methods:
Geometric, buffer, and network functions in GIS. Python scripting language used in developing the GIS-based map matching algorithm.
Factors (and significance):
The algorithm also generates a travel time, route distance, and number of left and right turns for each observed route used as independent variables in route choice models. Accuracy of the GIS-based map-matching algorithm is sensitive to the buffer distance.

Citation:
Gingerich, K. & Maoh, H. 2019, "The role of airport proximity on warehouse location and associated truck trips: Evidence from Toronto, Ontario", Journal of Transport Geography, vol. 74, pp. 97-109.
Topic:
Examines the relationship between land use and transportation from the perspective of warehouses in Toronto
Research Question:
Impact of warehouse location on freight transportation trips
Data:
Dataset purchased from InfoCanada containing 1.4 million firms. 489 firms pertaining to Toronto warehouses. GPS dataset of trucks was used to explore the relationship between warehousing land use and transportation.
Methods:
A location choice model is developed to reproduce the decision process of firms selecting suitable locations to operate warehousing facilities. Multinomial logit (MNL) and mixed logit (MXL) models were estimated to identify the factors affecting the location decision of warehouses. Models were estimated using NLOGIT software with maximum likelihood estimation to determine the parameters that provide the best fit.
Factors (and significance):
Proportion of industrial land use provides the largest role in the model. Other factors include proximity to Pearson Airport, proximity to rail facilities, proportion of industrial land use, highways within a 200m radius, cost of land, and rural locations.

Citation:
Kaneko, N., Oka, H., Chikaraishi, M., Becker, H. & Fukuda, D. 2018, "Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-based Recursive Logit Model Featuring Link Awareness", Transportation Research Procedia, pp. 251.
Topic:
Develop an alternative Recursive Logit model that considers the probability of awareness of the next link that improves the stability of model estimations
Research Question:
An accurate and detailed route choice model is needed to evaluated travel plans and policies such as congestion tax and toll systems
Data:
Used a digital road map as a basic network data. Model was estimated using vehicle trajectory data from Electronic Toll Collection dataset
Methods:
Chose 70 downtown destination points associated with heavy traffic flow divided all trips into long, medium, and short trips. Ranked trips length by facility type. Used a recursive logit model. Driver awareness was estimated using a binary logit (BL) model Created a utility function
Factors (and significance):
The final model included the following significant parameters that influenced route choice: Travel Time, Link Length cost, Roadway Width Number of U-turns, Number of Right turns, Route awareness

Citation:
Knorring, J.H., He, R. & Kornhauser, A.L. 2005, Analysis of route choice decisions by long-haul truck drivers.
Topic:
Conduct an empirical analysis of long-haul truck drivers' route choice decision making
Research Question:
What decisions do drivers make to determine perceived speeds on alternate routes? How rational are those decisions? How accurately do logistic models predict driver behavior and how to do we measure driver preferences and risk aversion.
Data:
Revealed preference data set consisting of about 250,000 trucks over a 13-day period.
Methods:
Used Revealed preference method. logistic model was constructed to describe route choice behavior when truck drivers are faced with alternate routes
Factors (and significance):
Significant findings were Trade-offs between distance and time. Perceptions are generated from a number of factors including past experience on the route, time of day, current traffic conditions, and knowledge of the route.

Citation:
Kunchev, L. 2017, "Methodology for selection the truck route", Engineering for Rural Development, pp. 263.
Topic:
Methodology for selection the route of a truck moving between two points by taking into account multiple factors relevant to transport
Research Question:
To employ a complex scientific approach for choosing the route of transportation of cargo by trucks and road trains.
Data:
Considered Data and Road network from Sofia to Varna. The pairwise comparison of the criteria and sub-criteria is carried out by a group of eleven experts related to road transport, managers of transport companies, and drivers.
Methods:
The methods of Analytic Hierarchy Process (AHP) and Cost/Benefits have been applied. defining and ranking the criteria by means of the multi-criteria analysis method and the cost/benefit method for route selection. Super-Decision software calculates the consistency index (CI). Graphical Sensitivity Analysis
Factors (and significance):
Route choice criteria were separated in to Main Criteria and Sub Criteria. Main criteria were, infrastructure road conditions, geographic characteristics of the road, roadside conditions, weather conditions, fuel consumption. Sub criteria included, road surface, longitudinal and transverse unevenness of the road, inclines on the road, number of turns, restaurants, petrol stations. Results showed Infrastructure and road conditions were most important. Important Sub-criteria were road surface condition, uneveness of the road, road sections under repair, road category, availability of places to rest and parking lots.

Citation:

Li, J. & Lai, X. 2019, "Modelling travellers' route choice behaviours with the concept of equivalent impedance", Transportation, vol. 46, no. 1, pp. 233-262.

Topic:

To capture route choice behaviors and to account for the correlation of routes in the logit model

Research Question:

Route choice models are affected when irrelevant alternatives are included in the choice set. The model cannot interpret the correlation degree of routes due to the independence of the irrelevant alternatives property (IIA) and leads to enlarged probabilities for correlated routes.

Data:

Real urban network with GPS data GPS devices were installed in urban taxis for monitoring purposes, thus, the route choice behaviors were based on the drivers' judgments

Methods:

MNL model with utility correction was used because it is the most favorable due to its simple form. Concept of equivalent impedance is presented to aggregate a set of links assuming people remember and process road network information at an abstract level. Derived a correction term for the utility of a multinomial logit model in which the advantages of a closed-form structure and easy computation remain unchanged. Cross-validation test to further investigate the forecasting abilities of the models.

Factors (and significance):

Effectiveness for the route choice model were measured by: Choice set generation, Analysis of robustness RMSE, Application in a real road network. The factors for the choice model itself included:Length (km), Artery road ratio, Number of unsignalized intersections, Number of signalized intersections

Citation:

Luong, T.D., Tahlyan, D. & Pinjari, A.R. 2018, Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida.

Topic:

Comprehensive exploratory analysis of truck route choice diversity

Research Question:

How to measure diversity in the routes trucks use to travel between an OD pair? What factors influence the diversity of truck route choice between an OD pair?

Data:

73,000 truck routes derived from 200 million GPS records. Produced database of diversity metrics for 277 OD pairs for long-haul travel in Florida and 527 OD pairs for short-haul travel in Tampa Bay.

Methods:

Negative binomial regression models were estimated to explore the influence of the factors on the number of unique routes traveled between an OD pair. Fractional response models were estimated to determine average path size (overlap) and standardized Shannon entropy (evenness) of route usage.

Factors (and significance):

The following metrics were used to quantify diversity: number of unique routes, average commonality factor, average path size, non-overlapping index, standardized variance of route usage,standardized Shannon entropy of route usage.

Citation:
Musolino, G., Polimeni, A. & Vitetta, A. 2018, "Freight vehicle routing with reliable link travel times: a method based on network fundamental diagram", Transportation Letters, vol. 10, no. 3, pp. 159-171.
Topic:
Develop solution of the Vehicle Routing Problem (VRP) based on reliable link travel times.
Research Question:
Link travel times may be estimated also by taking into account spatially aggregated traffic conditions estimated by Network Fundamental Diagram NFD.
Data:
Road network and partitioning of urban area of Villa San Giovanni (Italy).
Methods:
Introduction of spatially aggregated data, which are estimated using a Network Fundamental Diagram (NFD). The solution of the vehicle routing generates the best routes for freight vehicles. Genetic Algorithm (GA) was used to solve the Vehicle Routing Problem
Factors (and significance):
Reliability in a motorway as the probability that vehicular speed decreases in the short term, which is a function of the link density limit. Introduce a reliability term, which depends on the fundamental diagram of the link and on the NFD of the homogeneous cluster of adjacent links.

Citation:
Oka, H., Hagino, Y., Kenmochi, T., Tani, R., Nishi, R., Endo, K. & Fukuda, D. 2018, "Predicting travel pattern changes of freight trucks in the Tokyo Metropolitan area based on the latest large-scale urban freight survey and route choice modeling", Transportation Research: Logistics and Transportation.
Topic:
Predicting travel pattern changes of freight trucks
Research Question:
Evaluate freight policies based on a route choice model calibrated by a GPS Dataset obtained from a large-scale freight survey
Data:
Data collected by Tokyo freight survey. Two Datasets were formed: 1 included GPS records of trucks obtained from the digital tachograph data collected by an on-board unit. 2 is detailed truck trajectories sourced from GPS devices on 84 trucks with international maritime containers and 188 large trucks with a maximum load capacity of 10 tons. The data concerns the trajectories of each truck over a 7-day period.
Methods:
Route choice model with the concept of the maximum route-overlapping ratio was developed. This study applies the MROR model to analyze the route choice behavior of trucks.
Factors (and significance):
Travel patterns changed significantly depending on the type of trucks in terms of the arrival and departure locations, time of day, and by road type. The following parameters were used in the route choice model. Value of travel time Non-weight designated road (dummy), Expressway (dummy), National arterial road (dummy), Specific general road (dummy), Other types of road (dummy), Overlapping ratio for the whole sample (%), Number of route observations

Citation:
Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, *2*, 65–100.
Topic:
State of the Art is route choice analysis
Research Question:
What are the current methods of choice set generation and route selection modelling
Data:
Literature
Methods:
Literature Review
Factors (and significance):
C-Logit Path Size Logit,Pair Combinatorial Logit, Cross Nested Logit, multinomial Probit,

Citation:
Prato, C. G., & Bekhor, S. (2007). Modeling Route Choice Behavior: How Relevant Is the Composition of Choice Set? *Transportation Research Record*, *2003*, 64–73. Retrieved from https://journals.sagepub.com/doi/10.3141/2003-09
Topic:
Modelling Route Choice Behavior in terms of choice set
Research Question:
How relevant is the composition of the choice set in route choice modelling.
Data:
Urban network of Turin Italy. 276 Web based survey responses choosing 236 routes over 182 origin destination pairs.
Methods:
Branch and Bound Algorithm C-Logit and Path size logit models Biogeme software Gauss Matrix Programming Language
Factors (and significance):
Commonality Factor, Distance, Travel Time of Experienced and Inexperienced driver, Landmarks, Habits, Navigating Ability

Citation:
Rowell, M., Gagliano, A. & Goodchild, A. 2014, "Identifying truck route choice priorities: The implications for travel models", Transportation Letters, vol. 6, no. 2, pp. 98-106.
Topic:
Identifies the truck routing priorities of freight companies through a survey of shippers, carriers, and receivers
Research Question:
Main objectives of this paper were: To show that groups of the trucking industry have distinct route choice priorities; to identify the differences; to propose a method to incorporate the results into freight models
Data:
Data collection was done through a survey distributed to various freight companies. The survey consisted of 15 questions to expose routing priorities. The factors were identified through literature review and studying existing software programs that companies use for routing decisions.
Methods:
Respondents had to rate 15 items they believed to affect route choice decision making. Item response theory and latent class analysis highlighted common and differing priorities among survey respondents. R was used to identify the items that discriminate between companies
Factors (and significance):
Minimizing cost and meeting customer requirements. road grade, hours of service limits, and driver availability depended on Delivery type suggested that truck routing priorities are not constant and uniform but dependent on trip length. Key priorities from the survey were: minimizing travel distance, minimizing travel time, minimizing total cost, planning multiple stop loads, meeting customer requirements, avoiding highway congestion, avoiding highway tolls, refueling locations; availability of truck parking to wait for scheduled delivery or pickup times; road grade and curvature; hazardous material considerations; size or weight limits; truck driver availability; truck driver hours of service limitations; availability of support in case of problems

Appendix B: Sample Calculations for Commonality Factor

Figure 2-1

Length of Route i Li = 413518 metres

Length of Route j Lj = 467117 metres

Length of overlapping portion between Route i and Route j Lij = 0 metres

$$CF_i = \sum_j \frac{l_{ij}}{\sqrt{L_i L_j}} = \frac{0}{\sqrt{413518 \times 467117}} = 0$$

Figure 2-2

Length of Route i Li: 414763 metres

Length of Route j Lj: 449299 metres

Length of overlapping portion between Route i and Route j Lij: 194790 metres

$$CF_i = \sum_j \frac{l_{ij}}{\sqrt{L_i L_j}} = \frac{194790}{\sqrt{414763 \ \times 449299}} = 45\%$$

Figure 2-3

Length of Route i Li: 461779 metres

Length of Route j Lj: 467496 metres

Length of overlapping portion between Route i and Route j Lij; 415366 metres

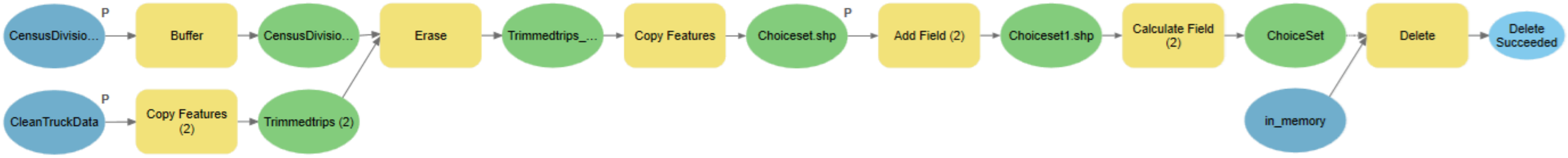
$$CF_i = \sum_j \frac{l_{ij}}{\sqrt{L_i L_j}} = \frac{415366}{\sqrt{461779 \times 449299}} = \ 91\%$$

Appendix C: Models Built in ArcGIS

Preprocessing of Trip Data

This model used inputs from observed truck trips and census division zones (for Canada) to organize all trips by origin and destination and create new shapefiles with the appropriate fieldnames for subsequent analyses; FID, TripID, OD-Pair, TripLength, MaxCF.

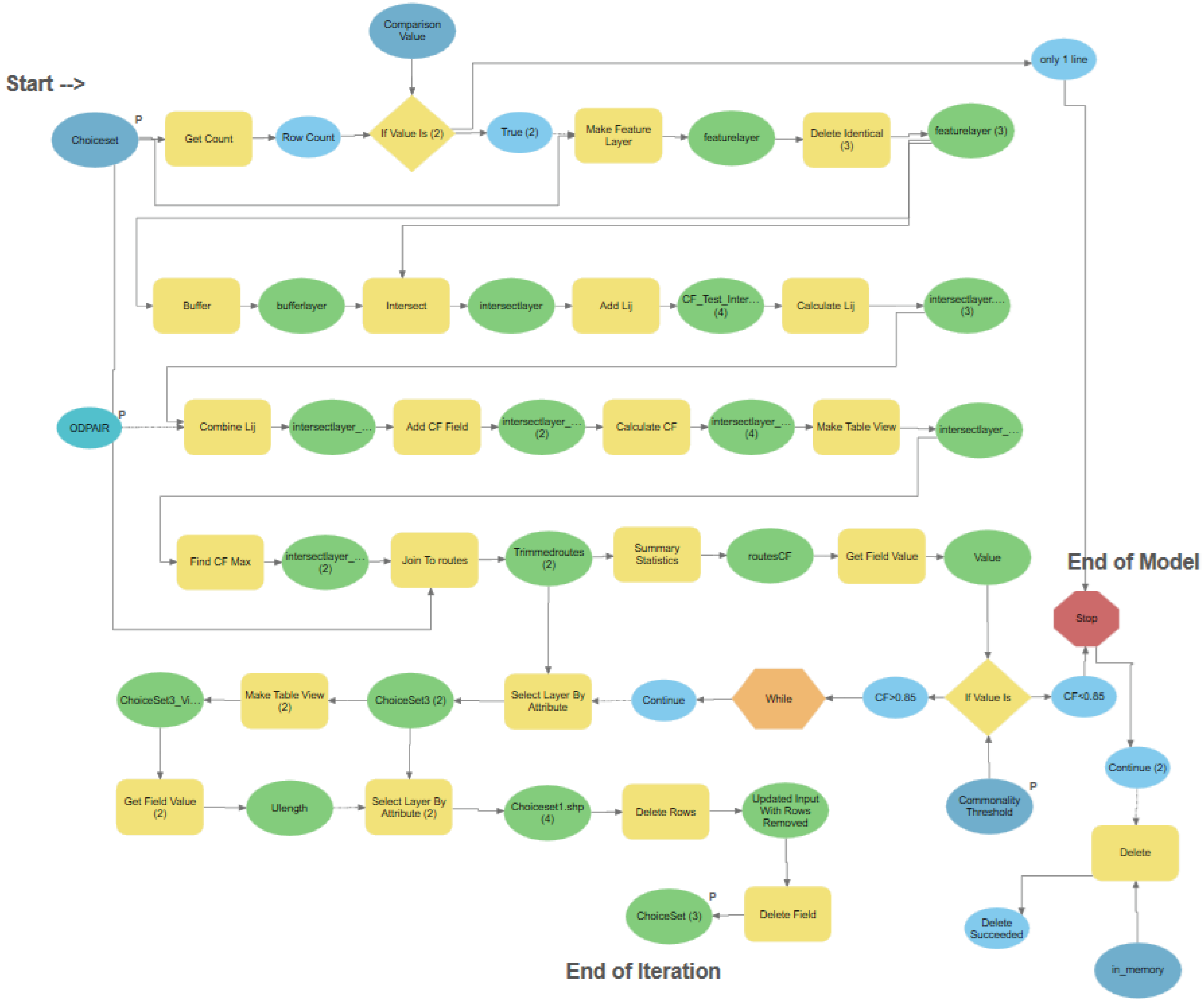
Figure B-1: Preprocessing Algorithm Diagram



Choice Set Generation

The Choice Set Generator model compares polylines representing trips in a given OD-Pair to determine their Commonality Factor (CF). This is computed through a series of geo-processing tools such as, buffer, intersect, summary statistics, and calculate field. Polylines exhibiting CFs above a defined threshold (e.g. 85%) share substantial overlap and may not be Iteration is necessary since deletion of one trip impacts the CF for other pairs of trips, hence requiring re-comparison and recalculation. This was achieved using a “while iterator”. Upon completion, the shapefile for a given OD-Pair would contain the set of routes dissimilar enough to be considered unique and independent from one another.

Figure B-2: Choice Set Generator Algorithm Diagram



Route Allocation Model

The Route Allocation Model uses a similar series of tools as in the Choice Set Generator model to calculate the CFs between trips and routes.

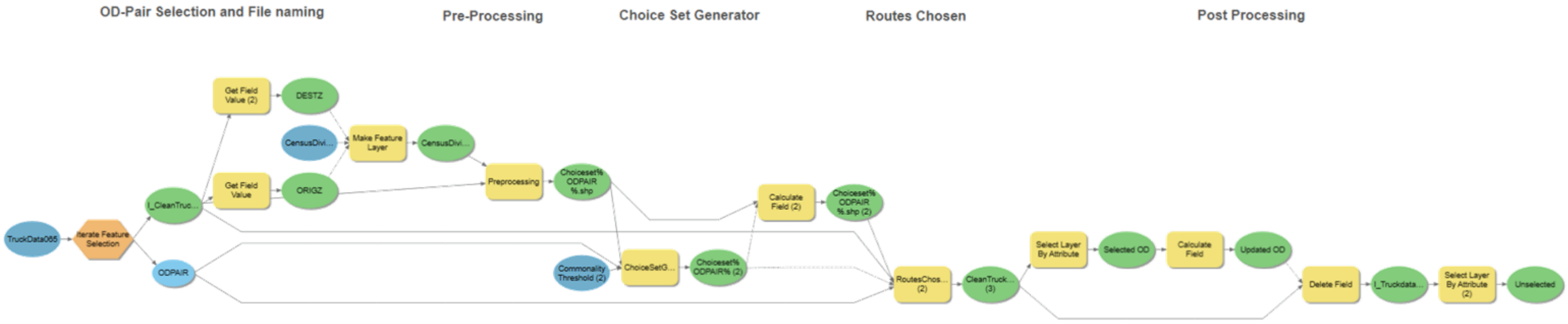
Each trip is assigned the most similar route identified by the highest CF value.

Figure B-3: Route Allocation Algorithm Diagram

Combined Model with Iteration for all OD-Pairs

The Pre-processing, Choice Set Generator, and Route Allocation models are combined into a nested model that uses the “Iterate Feature Selection” iterator to group trip records OD-Pairs and use them as input parameters to the Pre-processing model. The output shapefiles of the Pre-processing model become inputs to the Choice Set Generator, whose output along with the selected trip records became inputs to the Routes Chosen model. ArcGIS’s in_memory workspace was utilized for several intermediate outputs in the interest of computational time.

Figure B-4: Combined Route Identifier Algorithm Diagram



Appendix D: Code Used in NLogit

Descriptive Statistics on Variables

```
DSTAT

    ;Rhs=CSET,TTIME,AVGSPD,TTMIN,TTMAX,SIMMINHR,SIMTTIME,SIMMAXHR,SIMSPD, FWP, FWP401, FWP4XX, FWP407, DIESEL, WEIGH, INTRSCT, CF$
```

Logit Modelling for CF=0.95

```
NLOGIT

  ;Lhs = Choice, Cset, AltID

  ;Choices= 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22

        ;Effects: TTMIN(*)

        ;Effects: FWP(*)

        ;Effects: FWP401(*)

        ;Effects: DIESEL(*)

        ;Effects: INTRSCT(*)

  ;RHS=  TTMIN, FWP, FWP401, DIESEL, INTRSCT, CF$
```

Appendix E: Intermediate C-Logit Models

Table E-1: Model results for observed minimum travel time

Variable	Coefficient	T-Statistic
TTMIN	-2.84300***	-126.76
CF	1.23296***	24.34
LL(0)	-52494.888	
LL(β)	-30246.668	
ρ^2	0.423	
No. of Observations	37,111	

******* *indicates the parameter is statistically significant with 99% confidence*

Table E-2: Model results for average travel time

Variable	Coefficient	T-Statistic
TTIME	-0.947***	-69.18
CF	2.217***	50.5
LL(0)	-52494.89	
LL(β)	-46786.25	
ρ^2	0.109	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-3: Model results for observed maximum travel time

Variable	Coefficient	T-Statistic
TTMAX	1.178***	120.39
CF	2.322***	50.21
LL(0)	-52494.89	
LL(β)	-37131.64	
ρ^2	0.292	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-4: Model results for observed average speed

Variable	Coefficient	T-Statistic
AVGSPD	0.102***	99.93
CF	1.89***	41.15
LL(0)	-52494.88	
LL(β)	-43533.55	
ρ^2	0.171	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-5: Model results for simulated minimum travel time

Variable	Coefficient	T-Statistic
SIMTTMIN	-0.727***	-52.47
CF	2.75***	60.61
LL(0)	-52494.88	
LL(β)	-48566.16	
ρ^2	0.075	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-6: Model results for simulated average travel time

Variable	Coefficient	T-Statistic
SIMTTIME	-.248***	-27.37
CF	2.646***	60.55
LL(0)	-52494.88	
LL(β)	-49992.02	
ρ^2	0.048	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-7: Model results for simulated maximum travel time

Variable	Coefficient	T-Statistic
SIMTTMAX	-.025***	-4.21
CF	2.51***	57.61
LL(0)	-52494.89	
LL(β)	-50368.41	
ρ^2	0.041	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-8: Model results for simulated average speed

Variable	Coefficient	T-Statistic
SIMSPD	.109***	86.62
CF	1.95***	42.47
LL(0)	-52494.89	
LL(β)	-45895.18	
ρ^2	0.126	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Table E-9: Model results for route performance and infrastructure variables

Variable	Coefficient	T-Statistic
TTMIN	-1.675***	-74.75
AVGSPD	-0.002	-1.4
FWP	1.189***	22.58
FWP401	1.989***	41.91
DIESEL	.269***	57.19
INTRSCT	-.006***	-17.04
CF	.693***	11.7
LL(0)	-52494.88	
LL(β)	-25020.70	
ρ^2	0.523	
No. of Observations	37,111	

**** indicates the parameter is statistically significant with 99% confidence*

Appendix F: Variable Elasticities (NLogit Output)

TTMIN	1	2	3	4	5	6	7	8	9	10	11	12	13
1|	**-3.386**	2.581	2.581	2.581	2.581	2.581	2.581	2.581	2.581	2.581	2.581	2.581	2.5809
2|	1.825	**-4.813**	1.825	1.825	1.825	1.825	1.825	1.825	1.825	1.825	1.825	1.825	1.8251
3|	0.583	0.583	**-2.342**	0.583	0.583	0.583	0.583	0.583	0.583	0.583	0.583	0.583	0.5832
4|	0.251	0.251	0.251	**-1.19**	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.2511
5|	0.092	0.092	0.092	0.092	**-0.656**	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.0923
6|	0.033	0.033	0.033	0.033	0.033	**-0.361**	0.033	0.033	0.033	0.033	0.033	0.033	0.0325
7|	0.022	0.022	0.022	0.022	0.022	0.022	**-0.209**	0.022	0.022	0.022	0.022	0.022	0.0221
8|	0.005	0.005	0.005	0.005	0.005	0.005	0.005	**-0.148**	0.005	0.005	0.005	0.005	0.0049
9|	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	**-0.055**	0.005	0.005	0.005	0.0054
10|	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	**-0.03**	0.006	0.006	0.0057
11|	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	**-0.029**	3E-04	0.0003
12|	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	**-0.016**	0.0003
13|	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	**-0.015**

FWP	1	2	3	4	5	6	7	8	9	10	11	12	13
1|	**0.232**	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239	-0.239
2|	-0.15	**0.294**	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15
3|	-0.04	-0.04	**0.127**	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
4|	-0.021	-0.021	-0.021	**0.071**	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021
5|	-0.008	-0.008	-0.008	-0.008	**0.04**	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008
6|	-0.004	-0.004	-0.004	-0.004	-0.004	**0.018**	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
7|	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	**0.009**	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
8|	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	**0.005**	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04
9|	-8E-04	-8E-04	-8E-04	-8E-04	-8E-04	-8E-04	-8E-04	-8E-04	**0.002**	-8E-04	-8E-04	-8E-04	-8E-04
10|	-6E-04	-6E-04	-6E-04	-6E-04	-6E-04	-6E-04	-6E-04	-6E-04	-6E-04	**0.001**	-6E-04	-6E-04	-6E-04
11|	0	0	0	0	0	0	0	0	0	0	**9E-04**	0	0
12|	0	0	0	0	0	0	0	0	0	0	0	**7E-04**	0
13|	0	0	0	0	0	0	0	0	0	0	0	0	**0**

FWP401	1	2	3	4	5	6	7	8	9	10	11	12	13
1|	**0.171**	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321	-0.321
2|	-0.179	**0.227**	-0.179	-0.179	-0.179	-0.179	-0.179	-0.179	-0.179	-0.179	-0.179	-0.179	-0.179
3|	-0.064	-0.064	**0.147**	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064
4|	-0.027	-0.027	-0.027	**0.054**	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027
5|	-0.013	-0.013	-0.013	-0.013	**0.031**	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013
6|	-0.007	-0.007	-0.007	-0.007	-0.007	**0.027**	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
7|	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	**0.014**	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
8|	0	0	0	0	0	0	0	**0.003**	0	0	0	0	0
9|	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	-7E-04	**0.001**	-7E-04	-7E-04	-7E-04	-7E-04
10|	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	**8E-04**	-0.002	-0.002	-0.002
11|	0	0	0	0	0	0	0	0	0	0	**0.001**	0	0
12|	0	0	0	0	0	0	0	0	0	0	0	**3E-04**	0
13|	0	0	0	0	0	0	0	0	0	0	0	0	**0**

DIESEL	1	2	3	4	5	6	7	8	9	10	11	12	13
1|	**0.116**	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242	-0.242
2|	-0.118	**0.126**	-0.118	-0.118	-0.118	-0.118	-0.118	-0.118	-0.118	-0.118	-0.118	-0.118	-0.118
3|	-0.045	-0.045	**0.066**	-0.045	-0.045	-0.045	-0.045	-0.045	-0.045	-0.045	-0.045	-0.045	-0.045
4|	-0.023	-0.023	-0.023	**0.035**	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023
5|	-0.013	-0.013	-0.013	-0.013	**0.021**	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013
6|	-0.006	-0.006	-0.006	-0.006	-0.006	**0.012**	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006
7|	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	**0.008**	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
8|	-3E-04	-3E-04	-3E-04	-3E-04	-3E-04	-3E-04	-3E-04	**0.004**	-3E-04	-3E-04	-3E-04	-3E-04	-3E-04
9|	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	**0.002**	-0.001	-0.001	-0.001	-0.001
10|	-5E-04	-5E-04	-5E-04	-5E-04	-5E-04	-5E-04	-5E-04	-5E-04	-5E-04	**0.001**	-5E-04	-5E-04	-5E-04
11|	0	0	0	0	0	0	0	0	0	0	**4E-04**	0	0
12|	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	-1E-04	**0.001**	-1E-04
13|	0	0	0	0	0	0	0	0	0	0	0	0	**0.0004**

INTRSCT	1	2	3	4	5	6	7	8	9	10	11	12	13
1|	**-0.067**	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.0213
2|	0.022	**-0.099**	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022
3|	0.01	0.01	**-0.062**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.0103
4|	0.004	0.004	0.004	**-0.042**	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.0044
5|	0.004	0.004	0.004	0.004	**-0.022**	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.0038
6|	8E-04	8E-04	8E-04	8E-04	8E-04	**-0.012**	8E-04	8E-04	8E-04	8E-04	8E-04	8E-04	0.0008
7|	6E-04	6E-04	6E-04	6E-04	6E-04	6E-04	**-0.007**	6E-04	6E-04	6E-04	6E-04	6E-04	0.0006
8|	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	**-0.006**	2E-04	2E-04	2E-04	2E-04	0.0002
9|	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	2E-04	**-0.004**	2E-04	2E-04	2E-04	0.0002
10|	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	3E-04	**-0.003**	3E-04	3E-04	0.0003
11|	0	0	0	0	0	0	0	0	0	0	**-0.001**	0	0
12|	0	0	0	0	0	0	0	0	0	0	0	**-9E-04**	0
13|	0	0	0	0	0	0	0	0	0	0	0	0	**-3E-04**