7-7-2020

Modelling Truck Stop Destinations and Durations Within a Truck Tour Micro-Simulation Framework for the GTHA, Ontario, Canada

Ahmed Alshurafa
University of Windsor

MODELLING TRUCK STOP DESTINATIONS AND DURATIONS WITHIN A TRUCK TOUR MICRO-SIMULATION FRAMEWORK FOR THE GTHA, ONTARIO, CANADA

By

Ahmed Alshurafa

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2020

MODELLING TRUCK STOP DESTINATIONS AND DURATIONS WITHIN A TRUCK TOUR MICRO-SIMULATION FRAMEWORK FOR THE GTHA, ONTARIO, CANADA

by

Ahmed Alshurafa

APPROVED BY:

____

J. Tofflemire
Department of Civil and Environmental Engineering

____

W. Anderson
Department of Political Science

____

C. Lee
Department of Civil and Environmental Engineering

____

H. Maoh, Advisor
Department of Civil and Environmental Engineering

May 11th, 2020

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

In recent years, an increasing number of researchers and practitioners have shown an interest in model freight transportation activities. These activities have been growing at a significant rate due to globalization and the dependence on goods that are produced in offshore markets. Prior freight models were often aggregated, which made them less reliable for policy analysis. A remedy to overcome the limitations in aggregate model is to develop agent-based micro-simulation transportation models. These models are more comprehensive, thereby allowing them to calculate more accurate predictions. The current study utilizes data extracted from truck GPS records to model freight movements as the outcome of truck tours. A modeling framework is proposed for use in simulating the tours of individual trucks. The framework starts by predicting the number of tours per individual establishments. This is followed by micro-simulating each tour travel time, duration, and exact starting time. A stop generation model was used to predict the number of stops per tour and then the purpose of all intermediate stops within the tour. Next, the location of truck stops and the dwelling time at each stop are simulated. The focus of this research is to study the destination and duration of truck tour stops, and the analysis of the tours will make use of advanced statistical and geo-spatial modeling techniques. The results allow us to identify the significant factors that impact the movement of heavy trucks on the road network system. The geospatial and statistical results form the basis for developing a more comprehensive understanding of freight movement processes in Ontario. The models were incorporated in the proposed agent-based simulation model and were then used to predict the destination and duration of truck tour stops at the micro-level.

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor, Dr. Hanna Maoh. His continuous and persistent support, guidance, and encouragement gave me the foundation upon which I was able to build my work. His passion for transportation engineering research and his diligence was an inspiration to me. In addition, I would also like to thank my committee members: Dr. Chris Lee, Dr. William Anderson, and Mr. John Tofflemire. The time that each of these people took to review my work offer innumerable benefits and insights that strengthened my thesis. I would also like to extend my thanks to the chair of the committee, Dr. Yong Hoon Kim, for moderating my defense.

I wish to extend a special thanks to Dr. Shakil Khan for managing the Traffic Lab at the Cross Border Institute where I conducted my research. Additionally, I would like to thank my fellow lab mates: Terence Dimatulac, Georgina Madar, Moe Abdo, and Ayat Hussein. Working with this group of talented and bright individuals offered a pleasurable and supportive learning environment that greatly enhanced the quality of my work. I am extremely grateful for their continuous help and support throughout my work. I am also thankful for my friends for listening to my complaints and for their constant support and encouragement. When the pressure of my workload was wearing me down, they provided me with an outlet that allowed me to become revitalized and continue with my work refreshed with a new sense of vigor.

Lastly, and most importantly, I must express my very profound gratitude to my parents, siblings, and family for providing me with unfailing support and continuous encouragement. Without their love and support, this accomplishment would not have been possible. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Overview

The continuous increase in freight transportation activities has encouraged research on freight movements in recent years. However, current freight models are either too aggregate in nature or still too immature or under development to be included in advanced agent-based micro-simulation transportation modeling systems. This is a problem given the fact that freight activities represent a significant portion of the observed traffic congestion on the transportation network. The lack of adequate data and the diversity of carries that interact in the shipping operations have curtailed the progress of developing advanced freight demand forecasting models. In general, freight travel activities are more complicated than personal travel activities. Also, detailed freight demand modeling requires disaggregate data to understand on-ground activities and predict future freight travel demand. Accordingly, the availability of detailed data should enable researchers to develop comprehensive models to be used by decision-maker to assist in informing future transportation plans.

Traditional freight demand models (TFDMs) are focused on modeling trips at the zonal level. These models are still being used by most governmental agencies as part of their existing transportation planning models. TFDMs are usually developed as part of a conventional four-stage model (i.e., trip generation, trip distribution, modal split, and traffic assignment), although they might not include all four stages especially the modal split model. Historically, the four-stage modeling process was intended to deal with passenger vehicle movements but has been modified to model commercial vehicle movement despite key differences between passenger and commercial traffic. Despite their popularity, TFDMs suffer several drawbacks including their aggregate nature and lack of behavioral realism (Tavasszy, 2006).

Researchers during the last 15 years have shifted to developing agent-based models to overcome the drawbacks of traditional freight models (Stefan et al. 2005; Hunt and Stefan, 2007; Kuppam et al. 2014; Greaves and Figliozzi, 2008; You and Ritchie, 2017). However, agent-based models are considered to be data-hungry as they require extensive amount of information about the individual commercial vehicles and their establishments. Such data are traditionally collected via specialized surveys that are costly. For instance, Hunt et al. (2006) conducted an exhaustive survey to collect information from establishments engaged in shipping goods and providing services in Alberta, Canada. The survey also included an extensive set of interview with truck drivers to understand their commercial vehicle movement activities. The collected data considered commercial vehicle movements for 24-hour periods and contained origin, destination, purpose, and commodity information that was associated with the tours.

Recently, the availability of the Global Positioning System (GPS) technology created the opportunity for those who are interested in analyzing travel demand data. As a result, a growing number of studies have conducted research on commercial vehicle movements using big data from trucks' GPS data. Generally, GPS transponders are used by freight carriers to track the movement of their trucks from the start of the trip until the final destination and back. The generated records, specifically, the coordinates (i.e., longitude and latitude) of the truck itself at a specific point in time, can provide critical insights into the current patterns of freight movement. However, since the records forming the GPS data were not originally intended as an input for transportation models and analysis, there is a need for novel methods and techniques to mine truck GPS big data before they can be used in travel demand analysis. When dealing with truck movements at the micro-level, it is important to analyze tours. Truck tours represent a chain of travel activities in which a truck starts from an establishment and then makes several stops—including stops to pick up/deliver goods, refuel, or rest—before returning back to the establishment. The majority of existing freight models focus on trips as opposed to tours.

1.2 Statement of the Problem

The current study contends that the analysis of truck tours within a microsimulation framework will provide a more realistic picture of the true process governing the movement of trucks. This research project is unique because it will make use of truck GPS big-data of Canadian trucks. More specifically, this study will utilize one month of truck GPS data that represent the movement of trucks within the province of Ontario to advance the micro-simulation paradigm in travel demand modeling. Such work has not been done in the past; therefore, the study's results will offer a novel contribution to the transportation engineering literature and contribute to the development of operational integrated micro-simulation models of freight movement. Overall, this research will facilitate the development of operational micro-simulation models that are behavioral and policy sensitive.

1.3 Objectives

The current research has four primary objectives:

1. develop a more comprehensive understanding of freight movement processes in Canada by studying truck tours;
2. specify, estimate, and develop destination choice models to identify the location of a truck stops comprising a tour;

3. specifiy, estimate, and develop truck stop duration models to determine the total length of a truck stopped event.
4. contribute to the development of an agent-based micro-simulation modeling framework for simulating freight movements by focusing on the destination and duration of truck tour stops;

1.4 Thesis Outline

The remainder of this thesis is organized as follows. Chapter two provides an overview of previous studies regarding freight model types, issues, and challenges, while chapter three highlights the methods of analysis and the data that will be used in this project. This will be followed the fourth chapter, which details and discusses the results. The fifth and final chapter outlines the anticipated outcomes and provides conclusions.

CHAPTER 2: LITERATURE REVIEW

Modern supply chain processes rely heavily on goods that are shipped between different manufacturers, wholesalers, and venders. Freight travel activities are at the heart of the supply-chain process. The forecasting of freight activities is a complex process and has not been explored to the same degree as passenger travel activities. Freight in a transportation system can be found in five basic forms: road systems, rail systems, rapid transit systems, marine transport systems, and airline systems. The literature review focuses on road systems, specifically commercial vehicle movments (CVMs). The continuous dependency on road freight activities has influenced the performance of road systems especially in countries like Canada and the US. This, in turn, makes road freight transportation a critical part of the transportation planning process. Failing to plan for freight activities in transportation plans could hamper the performance of transportation networks, resulting in distress in freight movement and economic performance. In recent years, an increasing number of studies have been conducted on commercial vehicle activities to help planners and policy makers improve their ongoing and future Transportation Master Plans (TMPs). Despite the ongoing efforts, most freight transportation models in the urban and/or metropolitan context are either immature or are still under development (Freight Demand Modeling).

A recent study by (Doustmohammadi et al. 2016a) demonstrates how transportaion planning processes can be assisted, but this requires efficient and reliable freight demand forecasting models. Data from these models can be used to predict three important factors: the impact of freight on transportation networks, short- and long-term freight demands, and the interaction between commerical and passenger vehicle travel.

2.1 Freight Modeling Approaches

Developing an effective model necessitates a deep study of CVMs. Researchers were able to apply a variety of methodological techniques to study commercial vehicle activities. Several studies developed various frameworks to analyze and model CVMs. In this context CVMs respresent truck tours. A truck tour is typically defined as a round trip where a truck leaves the establishment to perform one or more stops before returning to the establishment. These stops may be for a variety of reasons, including transferring goods, providing services, or taking a break. Chow et al. (2010) and Fischer et al. (2005) provide a review of different freight forecasting models with respect to input data, model development, and the output of used model. The tree diagram in Figure 2-1 is created based on their reviews to illustrate the distribution of freight forecasting models.

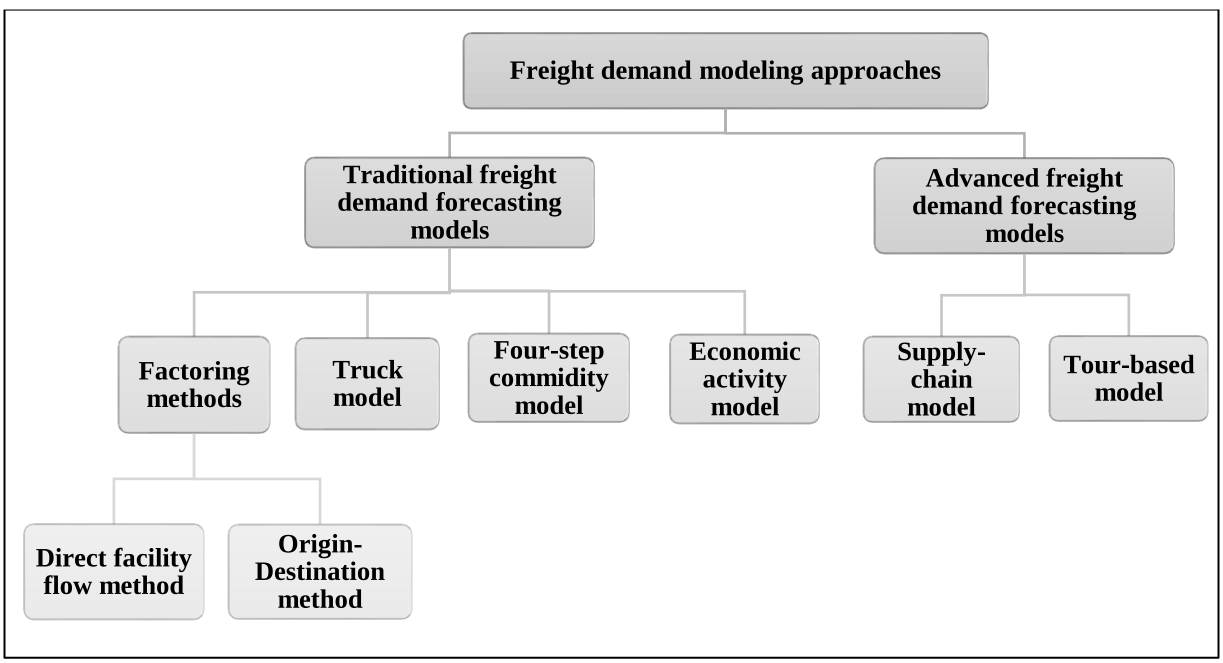


Figure 2-1: Modified freight demand modeling approaches based on Chow et al. (2010) and Fischer et al. (2005)

The following two sub-sections provides a review of models, modeling efforts in freight demand models, the nature of the data required, and the advantages and drawbacks of each approach.

2.1.1 Traditional Freight Demand Forecasting Models

Traditional freight demand models (TFDMs), which are focused on modeling trips at the zonal level, are still being used by most governmental agencies as part of their existing transportation planning models. TFDMs are easy to operate but their data and results are limited. Some studies, use the term “conventional demand models,” among other terminology; however, to avoid any confusion, the current proposal will use the term “traditional”.

TFDMs can be classified into one of the following four approaches: the factoring model, truck model, commidity model, and economic activity model (National Academies of Sciences, Engineering, and Medicine, 2008). The factoring model is considering as one of the simplest and fastest ways to forecast commercial vehicle movements or commodity flows. Since this method relies on growth factors in its calculation, Yang et al. (2010), it has been identified as a growth-factor model in some studies. Factoring models include two approaches: the direct facility flow approach, and the origin-destination approach. The direct facility flow approach predicts truck flows by implementing current and past truck count data to estimate the growth factor and then apply the latter to observed truck traffic volumes to determine truck flows on a link-by-link basis Chow et al. (2010). In contrast, the origin-destination method applies the growth factors to the base year origin-destination matrix to determine truck trips.

TFDMs are usually developed as part of a conventional four-stage model, although they might not include all four stages (i.e., trip generation, trip distribution, modal split, and traffic assignment). Historically, the four-stage modeling proccess was intended to deal with passenger vehicle movements but has been modified to model commercial vehicle movement despite key differences between passenger and commercial traffic. These models are known as Truck models and since the name specifies truck, the modal split stage of the conventional four-stage model is eliminated. Therefore, a truck model is referred to as a three-step model as it features a trip generation, trip distribution and traffic assignment steps. The truck model usually represents three different truck classes: light, medium and heavy trucks based on gross vehicle weight rating. The latter estimates aggregate truck flows and assign them to the road network links. A commodity model in essence is a four-step model. In a nutshell, the first step estimates an origin-destination matrix of freight in terms of weight or value using a set of predefined Traffic Analysis Zones (TAZs). In the second step, trip flow between distinct zones are estimated for the study area. In the third step, a mode choice model is applied to determine what modes should be used to ship the commodities between their respective origins and destinations. Usually, trucks stand as one of the major modes that are used for long-haul commodity logistics. In the last step, assigning the flows to the road netwrok depends on the estimated truck flows from the third step. This stage will provide the traffic volume on each road link of the network.

The economic activity model can be named as an Input-output economic model. This model describes the effects of transportation on the trade and economy Anon (2013). The model relies on a set of technical coefficients as well as trade coeffiecients that describe how economic sectors are tied to each other within the one region and between regions. Examples of this approach can be found in the work of Maoh et al. (2008). Huang and Kockelman (2007) developed a Random Utility Based Multi-regional Input-Output (RUBMRIO) models to study how economic linkages and trade translates into freight trips. Another example of such model at the urban level can be found in the work of Abraham and Hunt (1999) who developed the MEPLAN model. The latter is an Integrated Urban Model that models the process of trading between markets (here

TAZs) as the outcome of consuming land uses by various land use activities. The interactions of these actors in response to changes in land prices translate into passenger and commercial trips that can then be assigned to the road network.

To conclude, traditional freight demand forecasting models, with the exception of economic models, are usually straightforward and easy to implement since their data requirements are minimal and not hard to obtain. The challenges and issues of this approach will be discussed in the last section.

2.1.2 Advanced Freight Demand Forecasting Models

An investigation done by Doustmohammadi et al. (2016a) revealed that the two most recent models—the tour-based model and the supply-chain model—demonstrated greater promise with regard to addressing current and future freight forecasting needs. Another study done by Doustmohammadi et al. (2016b) showed that the tour-based approach was more effective than the trip-based approach since the tour models have the ability to capture the true movement of commercial vehicles. However, unlike the trip-based approach, tour model development requires observed GPS data or detailed survey data.

The following sections are organized on the basis of two basic advance freight demand forecating models: 1) tour-based model and 2) supply-chain model.

Tour-Based Model

The tour-based models have the capability to predict the movement patterns of commercial vehicles at the micro-level. However, this class of models is considered to be data-hungry as they require extensive amount of information about the individual vehicles and their establishments. The first dissagregate tour-based model was developed by Hunt and Stefan (2007) for Calgary, Alberta. This model used data generated from an extensive set of interviews as a primary source for representing commercial vehicle movements. The collected data considered commercial vehicle movements for 24-hour periods. The collected data also contained origin, destination, purpose, and commodity information that were associated with the tours. The Calgary model offered a novel technique for modeling commercial vehicle tours as a process that consisted of six stages. In the first stage, a tour generation model is developed to estimate the number of tours per employee for a given establishment category in a given zone per day using a regression models. In the second stage, single-level logit models are developed to find the tour purpose and vehicle type for each tour per each zone based on establishment category. The utility functions of the logit models include zonal-level land uses variables, as well as, establishment location and accessibility attributes. In the third stage, a single level logit model is used to determine the time period of the day when the tour will take place. Here, the model divides the 24 hours into five major periods (i.e. five alternatives). Next, a Monte Carlo simulation process is used to determine the precise start-time for each tour with respect to the simulated period based on the period specific logit model. In the fourth stage, single-level logit models are used again to specify the next stop purpose, such as business, return to establishment, or another. In the fifth stage, a different set of logit models are used to determine the next stop location, which depends on the purpose of the next stop. In the sixths and final stage, Monte Carlo simulations are applied for a second time to determine the duration of a given stop. The Calgary model offered a more promising approach compared to the conventional trip-based approach. This was demonstrated by the work conducted by Ferguson et al. (2012), which transferred the Calgary's modeling framework for the Greater Toronto and Hamilton Area (GTHA).

Recently, Kim and Park (2017) developed a tour-based model similar to the Calgary model. The type of data used in the model was a nationwide commodity flow survey. The framework consists of four sub-modules; departure time choice, next-stop destination, stop duration, and next-stop purpose choice. The first sub-module, departure time choice, has been modeled using the Monte Carlo process which considered the establishment category and time period. The second sub-module, next-stop choice destination, was based on the multinomial logit (MNL) model and included an accessibility variable to capture different levels of economic agglomeration. In the third sub-module, Monte Carlo simulations were employed to predict stop duration. However the simulation made use of travel distance, existence of an industry, total number of employees, and accessibility variables. The fourth sub-module in the modeling framework was focused on next-stop purpose choice. This was modeled as the last stage unlike the case of models in other studies such as Hunt and Stefan (2007) and Gliebe et al. (2007). The next-stop purpose choice was modeled in the same fashion as the next-stop choice destination. However, including the accessibility variable in the MNL model was not helpful.

Gliebe et al. (2007) proposed a dynamic activity choice model as a disaggregate tour-based freight modeling scheme for the pattern genearation for Ohio, USA. The model incremently assigns activities and activity locations to the traveler, as shown in Figure 2-2. This model was developed based on an establishment survey data.

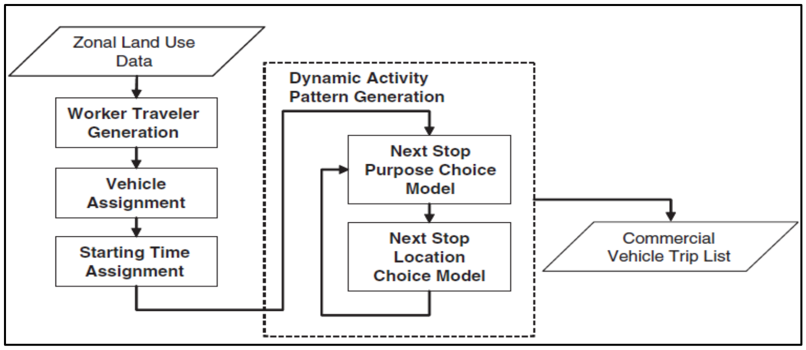


Figure 2-2: Structure of Dynamic activity choice model (Gliebe et al., 2007)

Several tour-based modeling frameworks have been developed based on raw Global Positioning System (GPS) data to gain information about truck movement activities. The study by Joubert and Axhausen (2009)for South Africa provided a novel technique to process truck GPS data and extract commercial vehicles tours. This study was followed up by Joubert et al. (2010), who developed an agent-based model that combines and simulates the commercial vehicles chains and passenger vehicles. Other efforts can be found in the works of Stephen et al. (2008), Kuppam et al. (2014) and You and Ritchie (2017). Recently, Gingerich (2017) devised novel techniques for mining truck GPS data to identify detailed truck movement activities between Canada and USA. For instance, Gingerich et al. (2014) developed an entropy classification method to differentiate stop types as primary and secondary stops. Building on the pioneer effort of preivous studies, Doustmohammadi et al. (2016b) were able to compare the results from a truck tour-based model to field truck counts at a regional scale. Their work showed that the tour-based modeling approach is capable of producing relatively close results to real life movements.

Supply-Chain Model

The supply-chain model is fashioned by linking different activity parties, such as consumers or factories, to the distribution channel list (Disserta, 2011). These class of models are considered dynamic as they can investigate the interface between freight transport and economic processes. Based on the litreature review, this class of models is classified as comprehensive as it uses commodity flow data to capture new shipping behaviours, such as the adoption of outsourcing, e-commerce and Just-in-Time (JIT) delivery systems Samimi et al. (2010) to better understand the CVMs . Also this approach pushes forward the predictive-ability of the developed modeling framework as it improves sensitivity of the calibrated models in terms of the economics of commodities when decisions are made to ship goods Fischer et al. (2005). In summary, the supply-chain framework is shaped to allow the user to analyze, estimate, and forecast the logistical choices for a variety of stakeholders. A study conducted in Chicago has produced a freight forecasting framework (see Figure 2-3) that includes supply-chain and tour-based methods at a national and regional scales, respectively, to forecast goods movement and commercial vehicles for regional planning purposes Outwater et al. (2013). The latter study integrated the two methods in a single framework. This integration provides a linkage between short-haul (regional) and long-haul (national) shipements. However, this framework has been developed to work with disaggregate data.

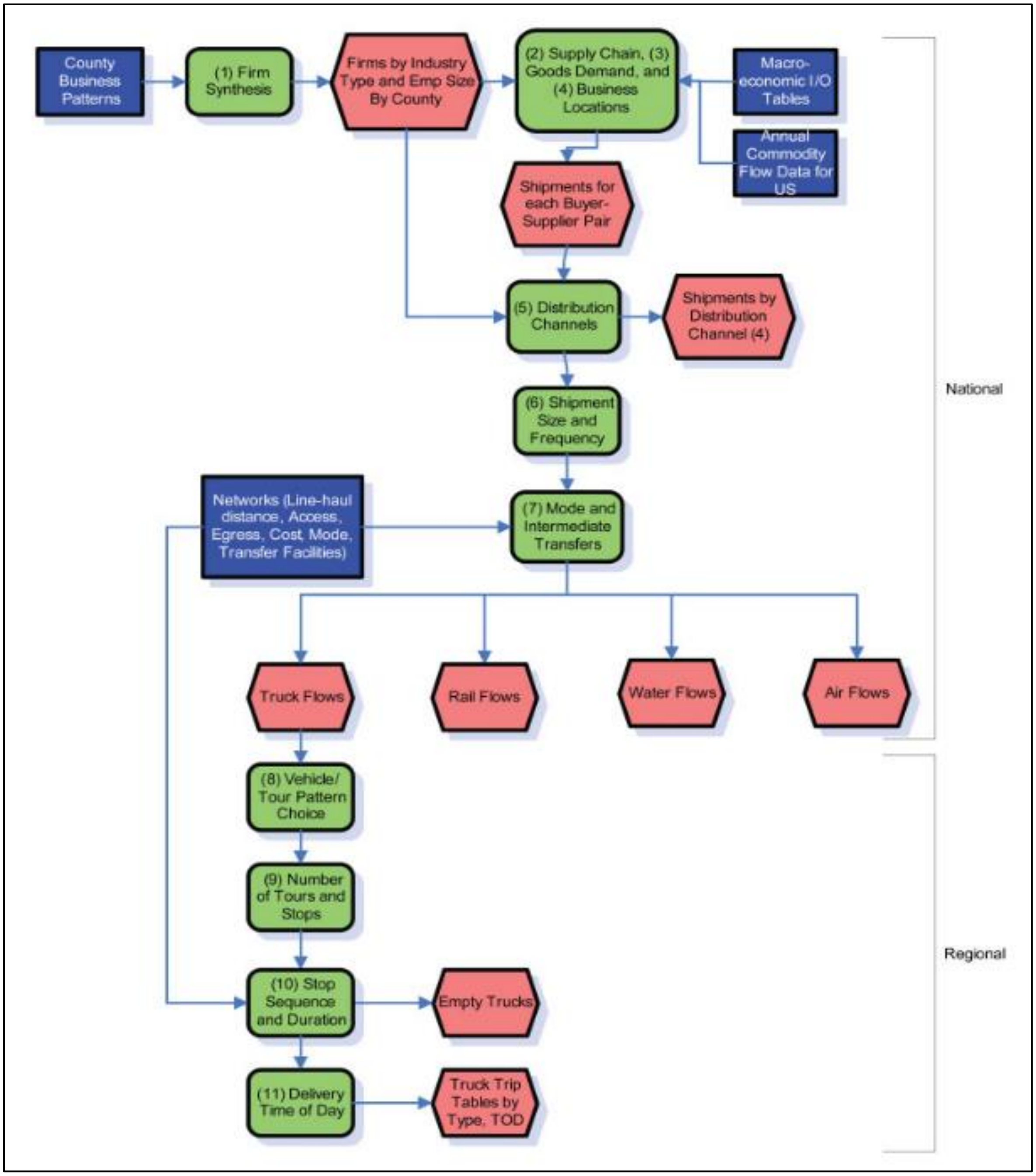


Figure 2-3: Chicago supply-chain and tour-based modeling framework (Outwater et al. 2013)

2.2 Freight Forecasting Modeling Issues and Challenges

Despite the progress made on developing freight forecasting models, researchers are still facing challenges in modeling freight movements. These challenges can be classified as follows: economic, political, geographic, and commerce operation concerns (Outwater et al. 2013). Besides these challenges, data availability plays a vital role in the development of freight models. In this context, data can be obtained from different sources starting from the most common type —Travel survey data — and ending at the advanced type —GPS tracking data—. As shown in Table 2-1, each of these sources has their pros and cons.

Table 2-1: Pros and Cons for different data type

Data Type	Pros	Cons
Driver Surveys	Firsthand accounts of routes and route choice	Lacks certain details due to privacy concerns; often contain mistakes
Establishment Surveys	Contains employment information, commodity production, and building details	Lacks certain details due to privacy concerns
Zonal-Level Data	Contains demographic, business sector, and land use data	Allows for general estimation
GPS Data	Contains accurate route data (unless signal loss or user error occurs)	Lacks truck type, commodity type, and pertinent business and route choice information; biased because it's obtained from a sample

Source: (Moore, 2017)

TFDMs rely heavily on travel survey data for their calibration. When dealing with TFDMs that uses the conventional 4-stage modeling approach, the data requirements are usually not as intensive as in the case of agent tour-based models. The 4-stage modeling approach is fairly simplistic since the use of the gravity model for producing the Origin-Destination flows depends on two factors as described by Wang and Holguin-Versas (2008): the zonal attributes and the travel impedance. A recent report by Moore (2017) showed that TFDMs are not suitable for modeling freight tours for urban areas since TFDMs fail to capture information about the interdependency of multiple trips within truck tours.

Generally, carriers are usually reluctant to disclose or share their fleet movement information with other agencies. Fortunately, in recent years, GPS data depicting the movement of trucks across various geographies started emerging. As such, several researchers have been able to develop their freight models based on such rich data (You and Ritchie, 2017; Kuppam et al. 2014; Stephen et al. 2008). Despite their ability to provide details about the movement of trucks, GPS tracking information available to researchers usually suffers from drawbacks such as, lack of details on the types of commidities being shipped, class of trucks used for the shipments, the location of the establishments, etc. Also, the utilized GPS data usually pertain to a smaple of trucks whose carriers subsribed to track their trucks. Therefore, unless the obtained GPS data covers a fairly large sample, the derived trip information might be biased. These drawbacks can be attributed to the fact that GPS tracking records were not originally intended as an input for transportation models and analysis. Thus, in turn, requires a need for novel methods and techniques to mine truck GPS big-data before they can be used in travel demand analysis. Research on travel demand models has placed a stronger emphasis on passenger vehicles rather than commercial vehicles. Despite the existing efforts, these modeling approaches suffer from several serious drawbacks for both passengers and commercial vehicles (Mladenovic and Trifunovic, 2014). On the freight side, a recent study shows that the development of freight forecasting methods still lags behind the development of passenger transportation forecasting, in both theoretical and simulation modeling analyses Jansuwan et al. (2017). According to Tavasszy (2006) these lags can be classified into two challenges: policy issues and modeling needs. Issues related to these two types of challenges are highlighted in Table 2-2.

Table 2-2: Key policy issues and modeling needs

Policy Issue	Modeling Needs
• A doubling of freight flows worldwide by 2050	• Forecasting international freight growth • Decoupling freight and the economy • Sensitivity to cost changes
• Growing volumes and shares of freight traffic on roads, due to both increased flows and greater numbers of smaller trucks	• Truck traffic behavior • Influence of greater freight shares on car drivers
• Creation of seamless multimodal networks	• Linking sea, inland waterway, and land transport models
• Concerns about international economic competitiveness; relation between worldwide networks and global trade; determining the costs and benefits of freight investments	• Suitable worldwide and continental models • Improved relation between spatial computable general equilibrium (freight—economy) and network models
• Pricing	• Response to cost changes by truck type, road type, time of day
• Ascertaining the performance of advancing logistics concepts such as hybrid supply chains, collaborative networks, e-logistics (business-to-consumers and business-to-business), and return logistics	• Differentiating between goods with different logistics characteristics • Making detailed statistics available
• Changes in vehicle type/mix—growth is surpassing all other categories and appears more difficult to capture in measurement or policy	• Forecasting choice of vehicle type, as well as causes and impacts
• Noise and emissions regulations, environmental damage, investments in new technology	• More accurate prediction of freight impacts and level of detail
• 24-hour economy—spreading out operations to deal with congestion	• Explaining shift of flows to different time-of-day periods
• New concepts for urban goods distribution	• Forecasting of tours at urban level • Sensitivity to time of day
• Safety and security	• Modeling of critical commodity movements by contents and origin

Source: Tavasszy (2006)

2.3 Modeling Tour Stop Destination Choices

Truck destination is the main influential character that governs truck tours (Hunt and Stefan, 2007). This makes truck destination one of the most important stages in freight demand forecasting models. However, the higher complexity in truck destination behavior demands greater insight into the modeling techniques used in freight demand forecasting models. Also, the availability of data determines the appropriate modeling techniques to model truck destinations. For instance, Hunt and Stefan (2007) developed a different set of logit models to determine the next stop location that depends on the purpose of the next stop as a part of a tour-based microsimulation for the city of Calgary. A total of 13 models were developed to model the stop location. Each model differentiates the behaviour of stop loaction based on a combination of industry category, vehicle type, and next stop purpose. Among the used covariates, the developed utility function included the average household income for a given zone, the travel utility between pair of zones (e.g., travel distance and travel time), the number of population and employment for a given zone, the accessibility to population and emloyment for a given zone, an attractor score for a given zone which provides an additional factor attracting a given type of stop, and enclosed angle for a given zone. The enclosed angle investigates whether the next stop gets closer or further from the establishment in a physical sense.

Kuppam et. al (2014) developed a destination choice model that predicts the location of each stop for a given tour. The used variables in this model are land use type of the establishment, previous stop location, current stop purpose, number of stops for a given tour by purpose. Also, the model accounted for the employment and population which are represnted in terms of zonal area type. The model also considered two types of accessibility variables: direct zone-to-zone or travel time between locations of two consecutive stops and aggregate accessibility measures;

which describes the accessibility of a stop location to employment. Among the key findings of this model, travel time has a very high negative impact on location choice utility. However, this impact increases as the number of stops increases for a given tour. Gliebe et al. (2007) proposed a dynamic activity pattern generation that includes the next stop location choice model. Like other models, this model considered the next stop purpose as a pre-step to model the next stop location. What distinguish this model is that, every five minutes, the driver is asked whether to stay at the current location or leave to start a new location, though restrictions were applied.

Througout the litreature, destination choice models have been explored similarly regarding the next stop purpose. However, the developed models were grouped regardless of the tour class. For example, these models did not distinguish a tour that has two stops from one that has four stops. The tour was grown until the next stop destination was the home establishment. In this thesis, we decided to develop separte models based on the class of the tour to gain a more accurate picture of the destination choice behaviour. Having separate models will also provide a more comprehensive understanding of the movements associate with of each class type.

2.4 Modeling Tour Stop Duration

Tour stop duration plays a vital role in determining the overall tour duration since the latter is the sum of the stop duration of each individual stop within the tour and the travel time spent by the commercial vehicle on the road. However, unlike stop duration, travel time can be easily predicted by calculating the shortest path time between any two points on the network. Several studies attempted to model the stop duration using rudimentary Monte Carlo techniques (Kim and Park, 2017; Gliebe et al. 2007; Hunt and Stefan, 2007). However, the approach followed in these studies was not adequately capturing the behavioral variation of stops duration. By comparison, econometric hazard duration models have the ability to capture those behavioral variations as they study the factors affecting the lifespan of the stop duration. There are three types of hazard duration models: parametric model, non-parametric model and semi-parametric model.

- The parametric modeling approach starts by assuming that the used data is follow a known statistical distribution such as the Log-normal, Exponential, Gompertz, and Gamma to name a few. The model is specified based a number of fixed variables with respect to the sample size. However, to maintain a superior performance, specification has to be parsimonious; that is, the number of variables should not be too large (Wheatley-Price et. al 2012) ;
- The non-parametric modeling approach is useful for graphical assessment as it provides a better picture of how the survival function looks like before a parametric or semi-parametric approach is used. However, unlike parametric and semi-parametric models, the non-parametric model does not account for independent variables in its calculation/graph (Katchova, 2013). As a result, this model does not lend itself well for predictive purposes.

- The semi-parametric model is a mix of parametric and non-parametric components. Oakes (1977) used the maximum likelihood techniques in a semi-parametric model to provide a more efficient estimation of the parameters of the model. The Cox-proportional hazard model is the most widely used semi-parametric model. The model consists of two multiplicative components: the underlying baseline hazard function $\lambda_0(t)$ and the parametric part $\exp(\mathbf{x}'\boldsymbol{\beta})$. In the model, $\lambda_0(t)$ describes how the risk of an event per time changes over time at a baseline levels of the specified covariates, while $\exp(\mathbf{x}'\boldsymbol{\beta})$ describes how the hazard changes in response to the explanatory variables.

According to the freight transportation literature, Sharman et al. (2012) is the only study to date that developed hazard-based stop duration models. The authors developed two models: an accelerated failure-time parametric hazard model, and a proportional non-parametric hazard model. The developed models made use of the following variables: 1) Arrival time of a stop whether in a dense region or not. Dense region is defined as a gross combined population and employment density of greater than 3,500 per square kilometer, 2) Travel distance whether on the inbound or outbound trip, 3) Number of stops on the tour, 4) Total sales value of all firms located on the property parcel, and 5) Type of establishment industry. The results of both models agreed on the following: longer travel distance on the inbound or outbound trips were associated with longer stop duration, and higher number of stops for a given tour decrease the stop duration for each stop. After comparing the results for the two models, the authors concluded that the accelerated failure-time parametric hazard model provided more interesting insights compared to the non-parametric model.

CHAPTER 3: STUDY AREA AND DATA DESCRIPTION

3.1 Study Area

The study area in this research is the Greater Toronto and Hamilton Area (GTHA). The GTHA consists of the following six key regions: Toronto, Peel, Durham, York, Hamilton and Halton (Figure 3-1). The GTHA is the economic heart of the province of Ontario and Canada. According to the most recent Canadian census, the GTHA was house for 6,954,433 people, 2,415,181 households, and 5,718,120 jobs in the year 2016. Table 3-1 provides the distribution of people and jobs by region.

Table 3-1: Distribution of people and jobs in GTHA regions

Region	Population	Jobs
Peel	1,381,739	1,119,400
York	1,109,909	905,545
Toronto	2,731,571	2,294,785
Durham	645,862	523,485
Hamilton	536,917	441,060
Halton	548,435	433,845

Table 3-2 provides the distribution of workers by economic sector. Out of the all industries in the GTHA, the six most prominent demands were in manufacturing, retail trade, professional, scientific and technical services, educational services, and health care and social assistance. Among the highest regions of these industries, Toronto tended to be the biggest host. This is not surprising given the fact that Toronto is one of the most populous and largest cities in Canada and the fourth largest city in North America. Within respect to hosting industries in the GTHA, Peel and York came in second after Toronto, while the rest of regions—Durham, Halton, and Hamilton—were third.

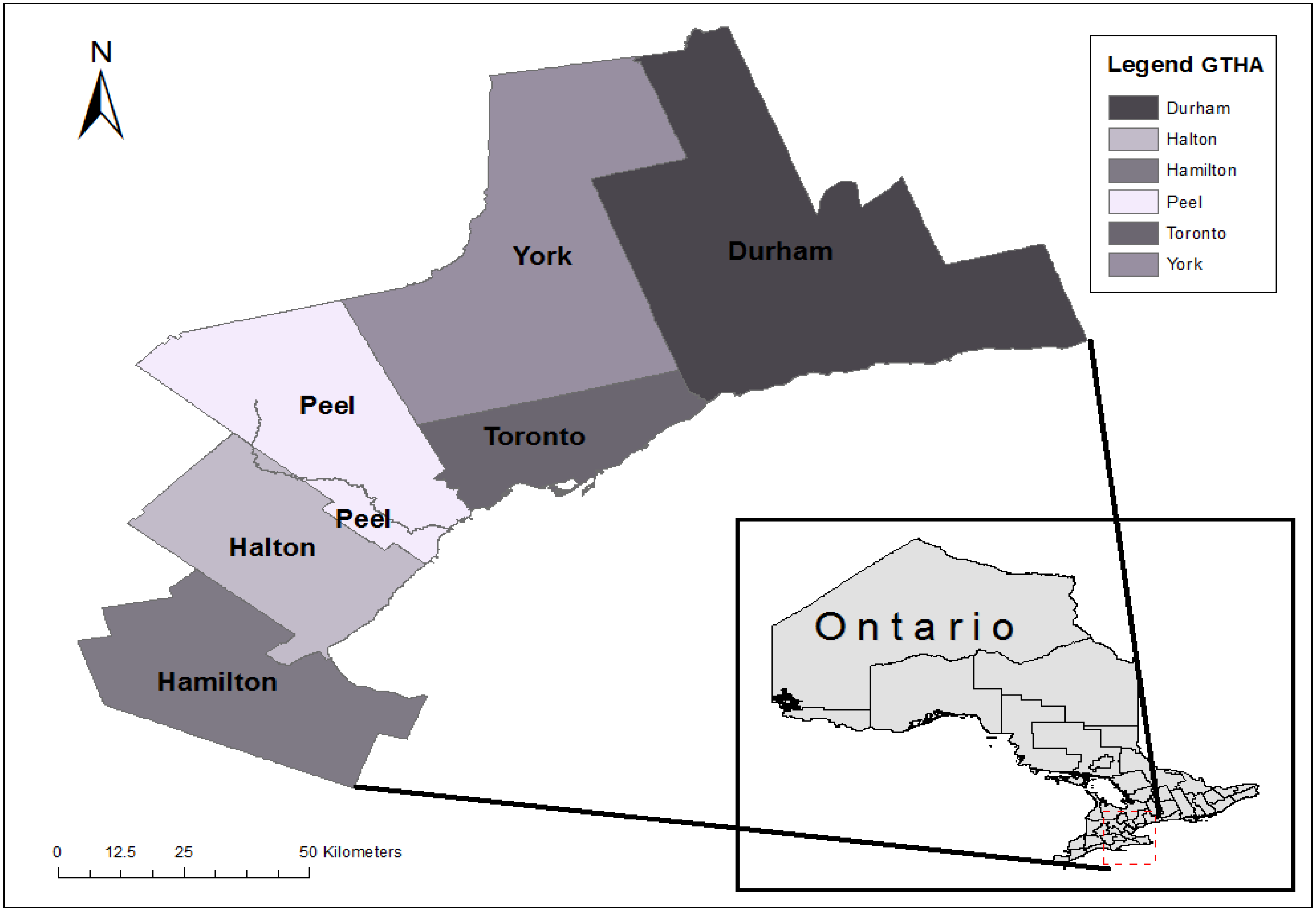


Figure 3-1: Geographic distribution of the Greater Toronto Hamilton area

Table 3-2: Distribution of workers by economic sector

Industry/ Region	Durham	York	Toronto	Peel	Halton	Hamilton	Total
Agriculture, forestry, fishing and hunting	3,000	2,265	2,095	2,175	1,470	2,865	13,870
Mining, quarrying, and oil and gas extraction	630	785	2,040	860	705	320	5,340
Utilities	7,560	3,320	5,915	3,055	1,835	1,315	23,000
Construction	27,260	43,055	76,480	44,755	16,790	20,115	228,455
Manufacturing	28,645	54,190	105,285	90,490	28,415	33,155	340,180
Wholesale trade	13,320	31,265	50,120	41,925	18,220	11,105	165,955
Retail trade	39,960	68,000	141,540	85,425	34,640	32,195	401,760
Transportation and warehousing	16,610	21,225	57,910	69,920	13,505	11,465	190,635
Information and cultural industries	9,895	16,790	61,345	17,780	8,235	5,545	119,590
Finance and insurance	22,355	49,310	120,005	47,495	23,510	10,835	273,510
Real estate and rental and leasing	6,615	18,415	39,940	15,495	7,125	4,870	92,460
Professional, scientific and technical services	24,150	66,445	175,685	61,500	31,020	15,705	374,505
Management of companies and enterprises	835	1,465	4,085	1,815	1,160	395	9,755
Administrative and support, waste management	17,425	24,035	78,890	41,985	11,900	14,130	188,365
Educational services	26,935	46,900	110,280	43,005	24,385	24,300	275,805
Health care and social assistance	36,610	50,510	143,250	59,265	27,390	36,280	353,305
Arts, entertainment and recreation	7,600	11,230	35,000	9,720	5,840	4,840	74,230
Accommodation and food services	20,630	33,525	106,910	42,200	17,150	18,325	238,740
Other services (except public administration)	13,660	25,045	67,385	27,625	10,705	11,740	156,160
Public administration	20,050	22,865	53,395	24,375	13,755	12,490	146,930
Total	343,745	590,640	1,437,555	730,865	297,755	271,990	3,672,550

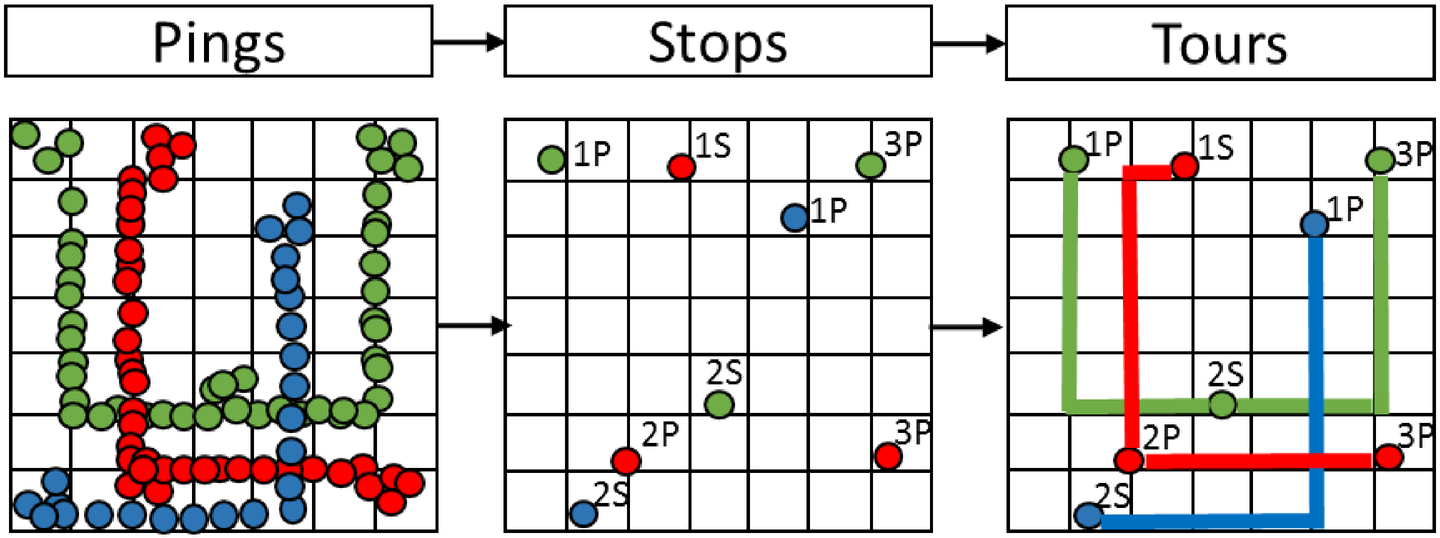
3.2 Data Description

This study utilizes GPS truck data to explore the tours of trucks originating from the Greater Toronto and Hamilton Area (GTHA) and destined within the province of Ontario for the month of March 2016. These tours are derived from GPS data that represent the movement of trucks. GPS transponders are used by freight carriers to track the movement of their trucks from the start of the trip, to the destination, and back. The records generated from the transponders are referred to as pings. These pings provide the coordinates (longitude and latitude) besides a unique ID of the truck itself at a specific point in time. The March 2016 GPS data were acquired from Shaw Tracking, which is a Canadian fleet management company that was acquired by Omnitracs in 2017. The acquired raw GPS pings captured the movement of 43,142 individual Canadian registered trucks across all of North America. These trucks were owned by 569 Canadian carriers.

The truck ID plays a vital role in determining the belonged pings for a given a tour. To that extent, truck tour can be visualized by plotting the recorded pings on a map. Here, each tour starts when a truck leaves the establishment to perform one or more intermediate stops that fall under one of three classifications: primary stops, which usually occur when goods are subjected to loading/unloading; secondary stops, which include refueling stops, rest, and others stops that do not involve the goods; or return stops, which are when a truck returns to the yard after visiting the last stop of a tour.

Figure 3-2 is used to visualize the idea of processing pings to identify stops and tours. In the example, three colors are used to present the pings belonging to the same Truck ID. The dwell time associated with the pings are then used to determine the stops for the trucks following the approach presented in Gingerich et al. (2015). Gingerich et al. (2016) also developed a model for differentiating primary stops from secondary ones. Calculating the dwelling time, travel time and travel distance between stops can be done using the timestamp, longitude and latitude associated with the pings representing the stops.

The processed GPS pings resulted in a total of 13,482 valid tours that represent the movement of trucks within the province of Ontario. These tours were generated by a total of 2,314 trucks that belonged to 42 Canadian carriers. Figure 3-3 presents the spatial distribution of these tours by origin. As the figure shows, the Peel region by far generated the largest number of tours (26%) within the province. Overall, the GTHA accounted for 45% of all the generated tours (i.e. 6,113 tours) observed during March 2016.



Truck	Pings
Red	32
Blue	15
Green	35

Truck	Stops
Red	3
Blue	2
Green	3

Truck	Tour
Red	1
Blue	1
Green	1

Figure 3-2: Transformation of pings

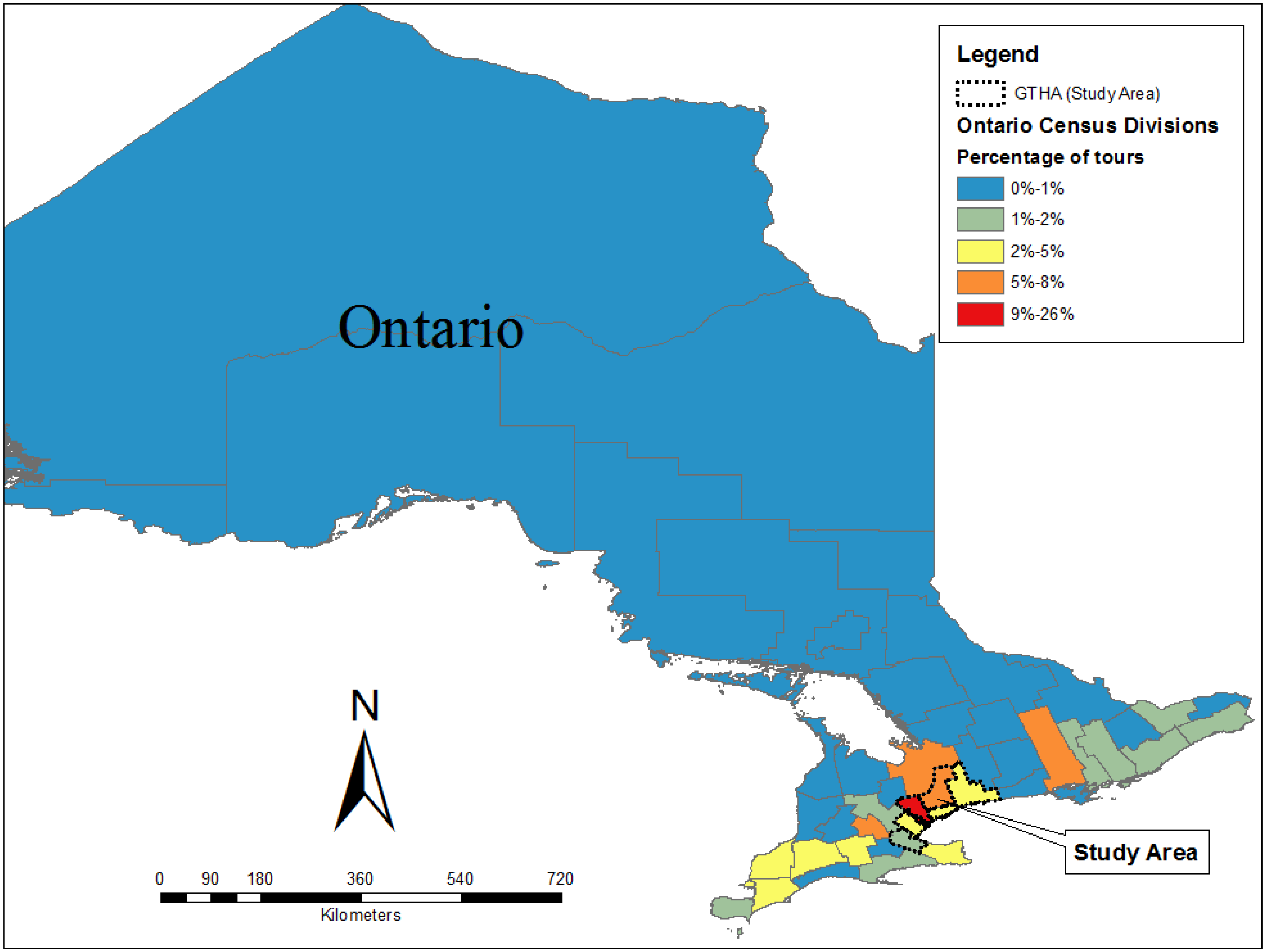


Figure 3-3: Spatial distribution of tours by origin census division

3.3 Preliminary Data Exploration and Manipulation

In this study, a total of 4,111 tours were derived from the 6,113 tours dataset, such that these tours originate from a location within the GTHA, return to establishment, and has a duration of one hour to thirteen hours. The analyzed tours included 13,401 intermediate stops that are categorized as follows: 64% primary, 31% secondary, and 5% return to establishment (see figure 3-4). The 4,111 tours have been classified based on number of stop/s in each tour (e.g. if a tour has 2 stops, then it will be classified as a 2-stop tour). Four different tours were classified; 1-stop, 2-stop, 3-stop, and 4-stop tour.

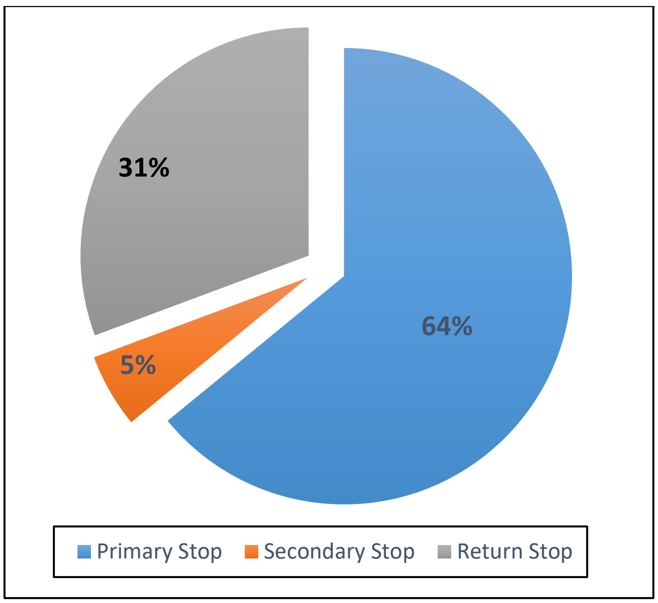


Figure 3-4: Proportion of stops by type

Majority of the 4,111 derived tours were 1-stop tour (77%). While the rest of tours accounted for (23%). Figure 3-5 shows a high demand on 1-stop tour compared to other classifications of tours. The 1-stop tour is known as 2-leg trip where the truck leaves the establishment to perform one stop and then goes back to the same establishment.

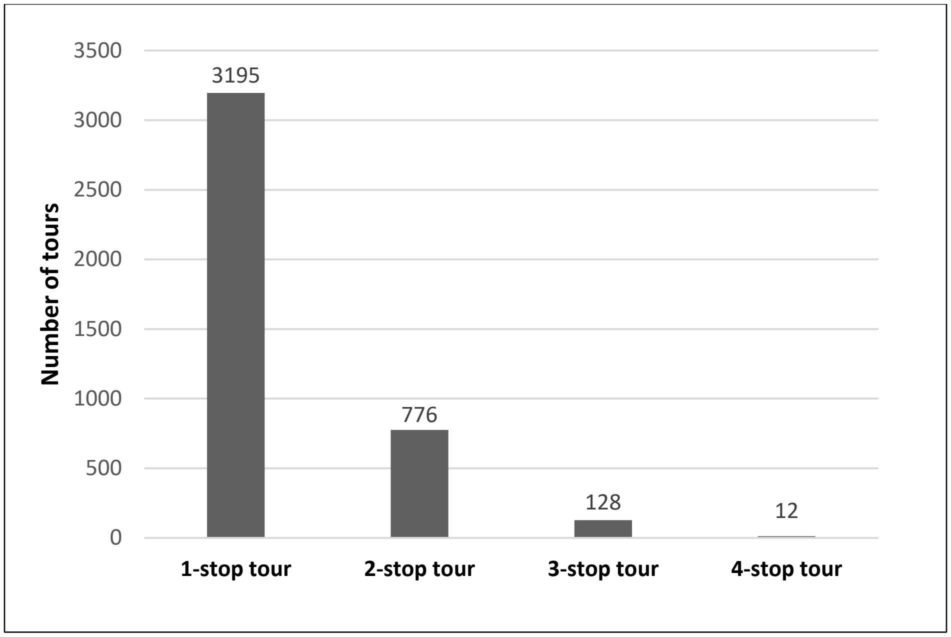


Figure 3-5: **Distribution of tours based on tours classification**

The following section will explore in more details the distribution of intermediate stops in each tour's class. As shown in Figure 3-6, the percentage of primary stops decreases as the classification of tours increases. While, the percentage of secondary stops increases as the classification of tour increases.

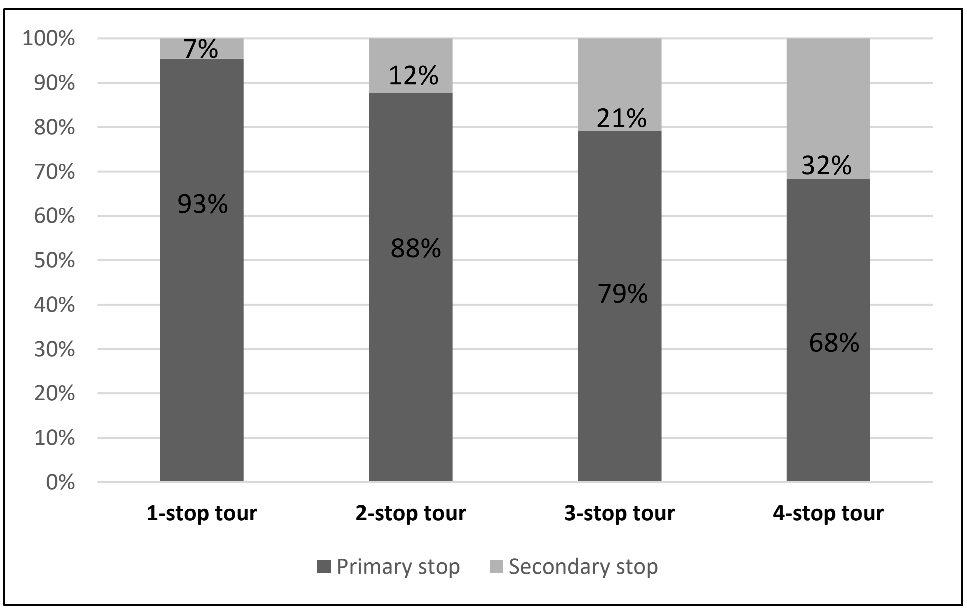


Figure 3-6: Percentage of intermediate stops per k-stop tour

In term of the geographic locations of the derived tours, Peel region generated 59% of these tours. However, York, Halton, Toronto, Durham, and Hamilton generated 42% in total (see Figure 3-7). More details about the percentage of Origin/k-stop tour are provided in Table 3-3. Each tour in the dataset includes information about its intermediate stops, the distance travelled by the truck to complete the tour, and the time it took the truck to complete the tour. The tour duration includes both of the travel time between stops and the dwelling time at each stop.

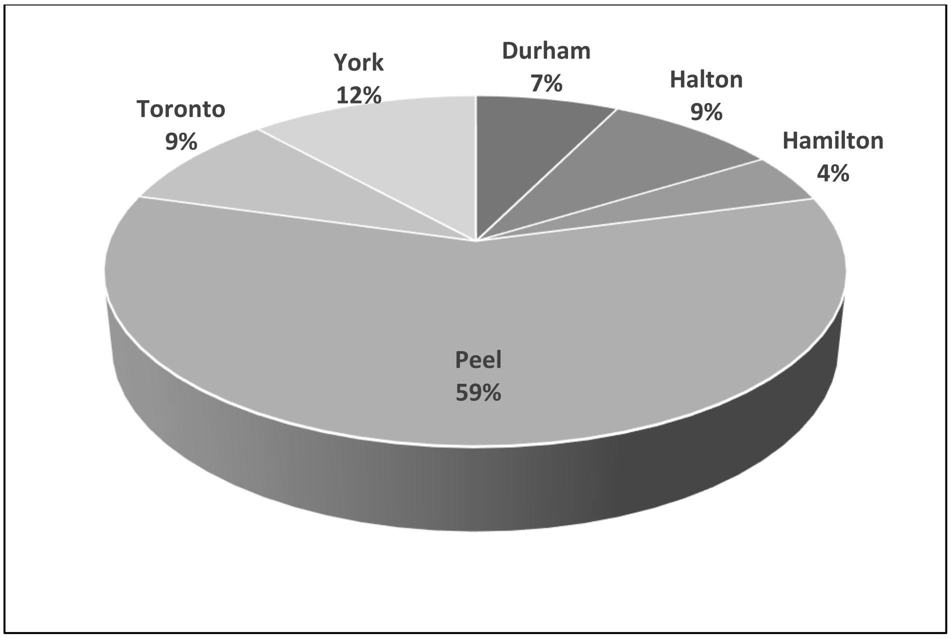


Figure 3-7: Percentage of geographic location of tour by origin

Table 3-3: Percentage of k-stop tour per origin

k-stop tour	Durham	Halton	Hamilton	Peel	Toronto	York
1-stop	6.03 %	6.91 %	3.45 %	44.15 %	7.30 %	9.88 %
2-stop	1.07 %	1.73 %	1.02 %	11.89 %	1.48 %	1.68 %
3-stop	0.17 %	0.24 %	0.10 %	2.41 %	0.10 %	0.10 %
4-stop	0.05 %	0.02 %	0.00 %	0.17 %	0.00 %	0.05 %
Total	**7.32 %**	**8.90 %**	**4.57 %**	**58.62 %**	**8.88 %**	**11.70 %**

The geographic distribution of tours by starting location of tours in GTHA vary. Peel region tends to have most of the tours because it houses a multimodal rail facility (i.e., CN yard) and the Toronto Pearson International Airport. Followed by York, which is also has a CN yard. Then, both of Halton, Toronto, and Durham have approximately same number of tours. Hamilton came last in this geographic distribution (see Figure 3-8).

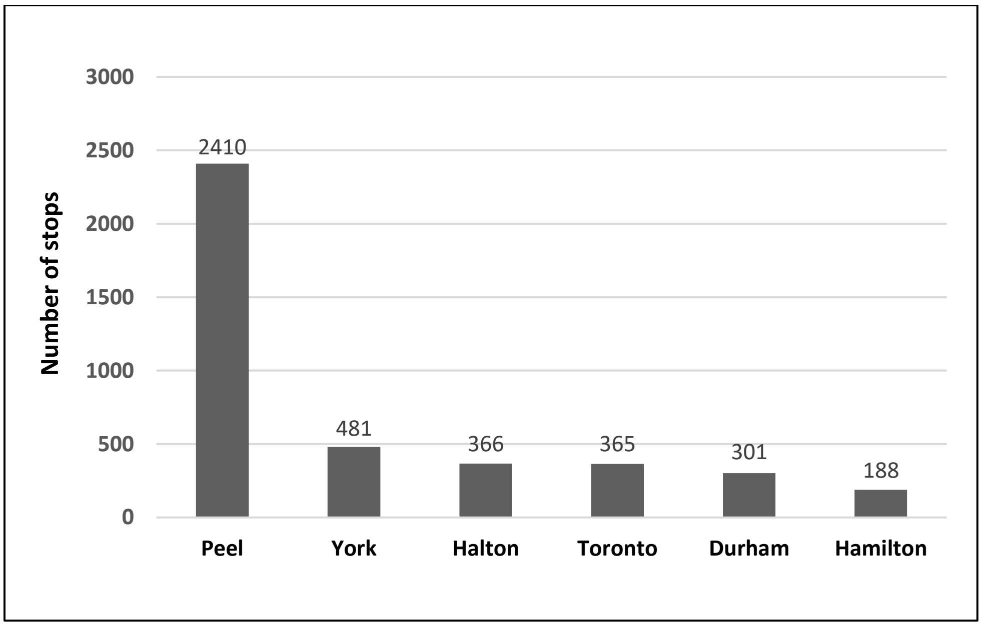


Figure 3-8: Number of stops based on the origin

Since the generated tours have no information about the stops purposes, the GPS Tracker software was used. The latter will take the coordinate system of a stop and based on the built-in geodatabase, the software will determine the stop's purpose. The GPS Tracker software categorized the stops into nine different industries. A summary of stop distribution by type of industry is presented in Figure 3-9. Out of all the industries, over 33% were service's stops, while transportation accounted for 21%. On the other hand, retail trade, manufacturing, and wholesale trade accounted for about 14%, 12%, and 10% respectively. As for construction, finance, public administration and agriculture, they pertained to about 9% of the stops.

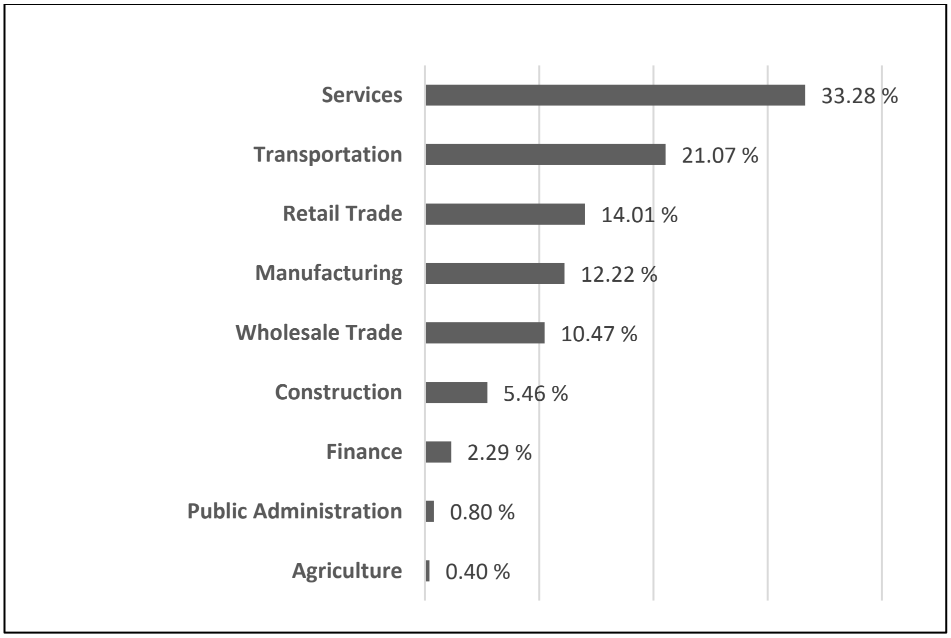


Figure 3-9: Distribution of stops based on industry

A total of 5,179 primary and secondary stops were associated with tours that started in the GTHA. Table 3-4 shows the distribution and percentage of these stops by industry and region.

Table 3-4: **Distribution of industry's stops based on origin**

Industry	Peel		York		Halton		Toronto		Durham		Hamilton	
	Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio	Count	Ratio
Services	1004	**32**	204	**36**	122	**27**	132	**30**	132	**36**	86	**36**
Transportation	713	**23**	124	**22**	96	**21**	76	**18**	83	**23**	13	**5**
Retail Trade	412	**13**	70	**12**	71	**15**	55	**13**	40	**11**	69	**29**
Manufacturing	342	**11**	67	**12**	104	**23**	67	**15**	22	**6**	26	**11**
Wholesale Trade	313	**10**	58	**10**	47	**10**	73	**17**	50	**14**	28	**12**
Construction	199	**6**	15	**3**	14	**3**	18	**4**	33	**9**	13	**5**
Finance	84	**3**	19	**3**	2	**0**	8	**2**	3	**1**	1	**0**
Public Admin	37	**1**	2	**0**	1	**0**	3	**1**	2	**1**	0	**0**
Agriculture	14	**0**	5	**1**	3	**1**	2	**0**	0	**0**	2	**1**
Total	3118	**100%**	564	**100%**	460	**100%**	434	**100%**	365	**100%**	238	**100%**

To illustrate the spatial nature of the analyzed stops, desire-line maps were generated to connect the origins of the tours (i.e. Peel, York, Halton, Toronto, Durham and Hamilton) to the destinations of the stops (i.e. any area in Ontario) as shown in Figures 3-10 to 3-15 (next page). The generated maps have a density for each line which reflects how many times a given destination stop was visited by a trucks from a specific GTHA origin. The average trip length for these destinations were obtained for each GTHA origin. This length represents the average travel distance between a GTHA origin and all corresponding stop destinations (see Table 3-5).

Table 3-5: Average trip distance based on origin

Origin	Average trip length (km)	Count of trips
Peel	121	3,118
York	116	564
Halton	98	460
Toronto	128	435
Durham	150	366
Hamilton	103	238

As shown in table 3-4, Peel region is associated with 60% (=3118/5179) of the total stops. Among the 3,118 stops, a total of 1,004 stops or 32% of the stops linked to Peel were associated with the service sector. Also, a total of 713 stops or 23% of the stops linked to Peel were associated with the Transportation sector. Peel is considered as one of the major freight hubs in North America and as such has direct access to three major Transportation modes including: (1) The Toronto Pearson International Airport, the largest freight air-hub in Canada; (2) two of the biggest CN yards and (3) seven major expressways. Therefore, the Peel region is truly an intermodal region that has the ability to handle freight movements between Southern Ontario and rest of the world.

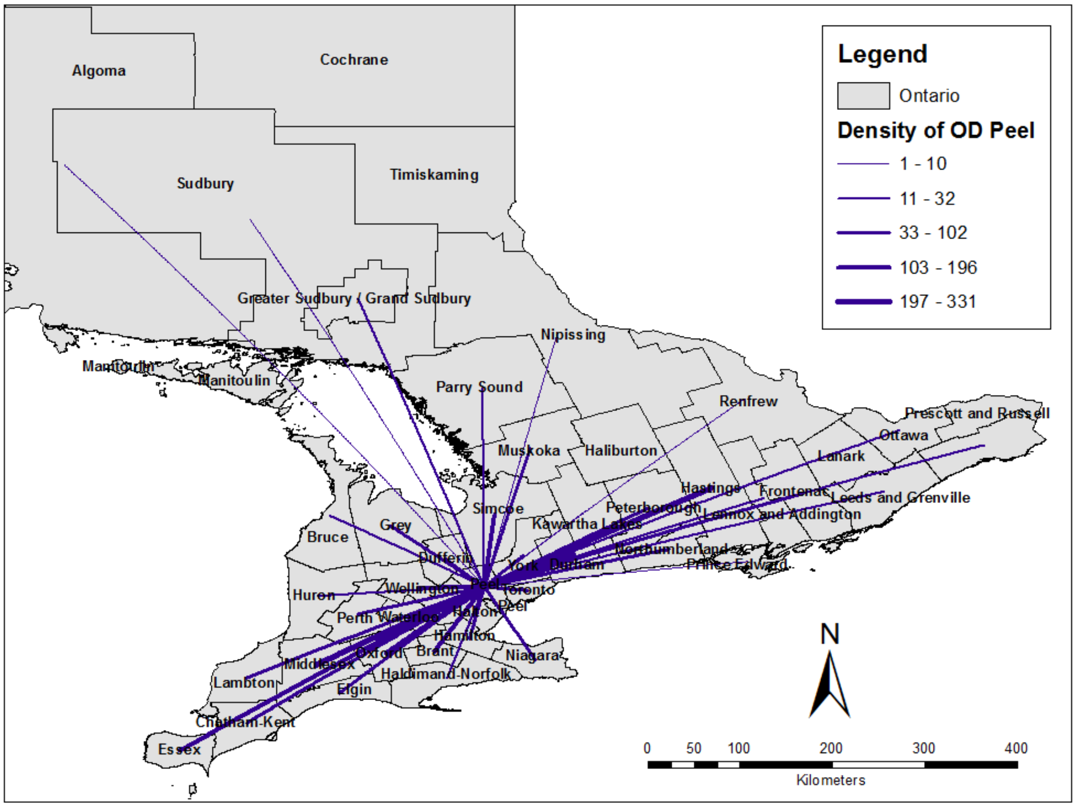


Figure 3-10: Origin-Destination trip of Peel

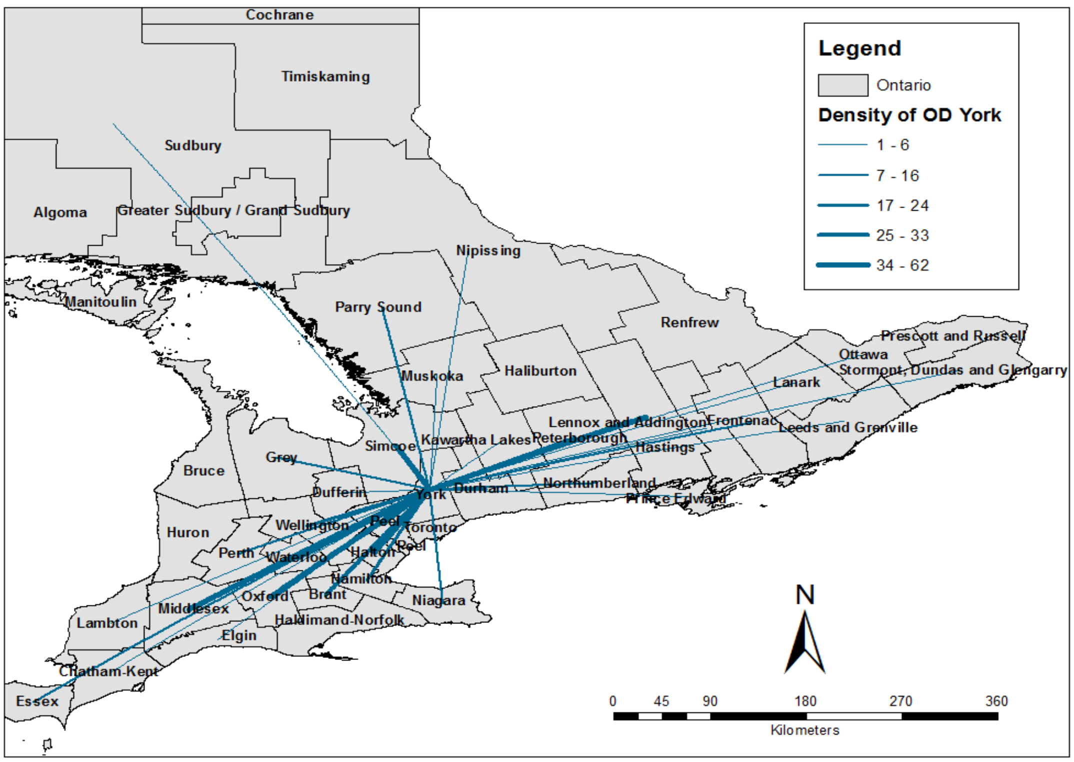


Figure 3-11: Origin-Destination trip of York

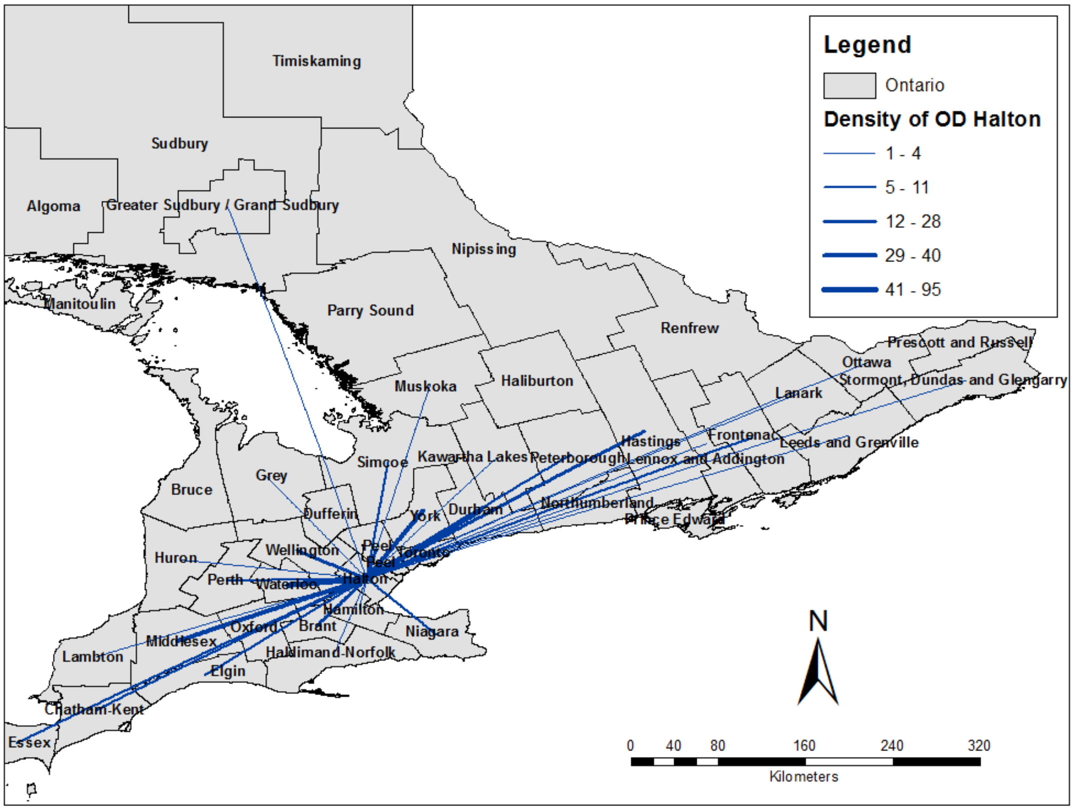


Figure 3-12: Origin-Destination trip of Halton

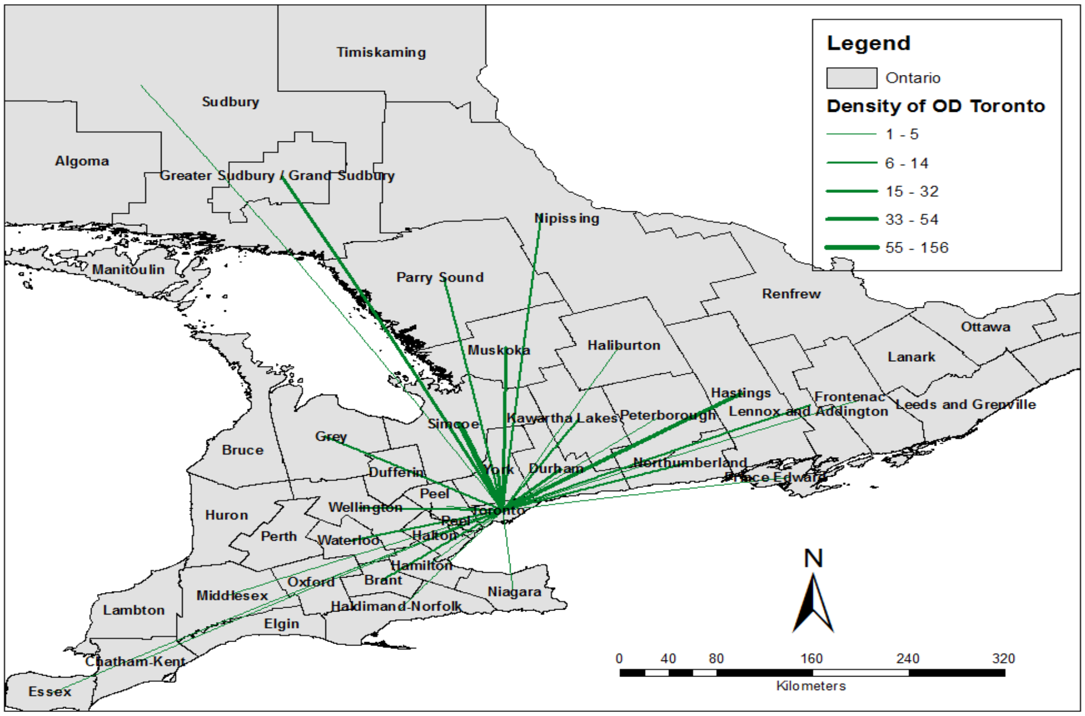


Figure 3-13: Origin-Destination trip of Toronto

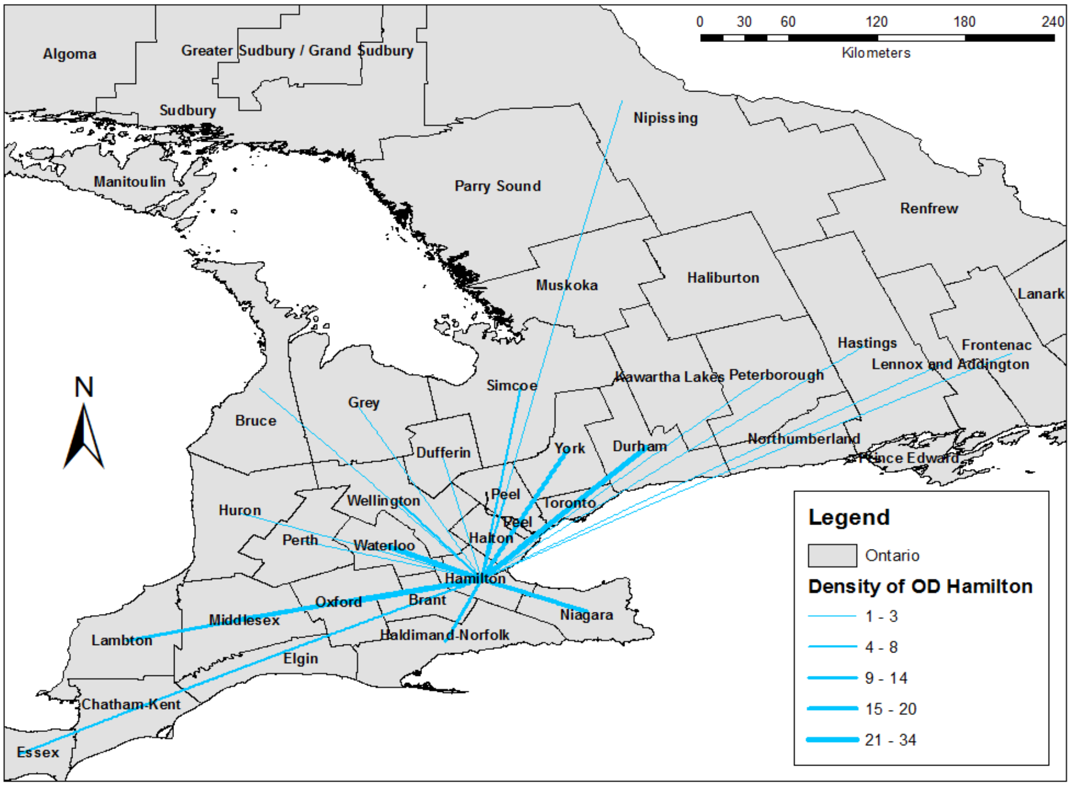


Figure 3-14: Origin-Destination trip of Hamilton

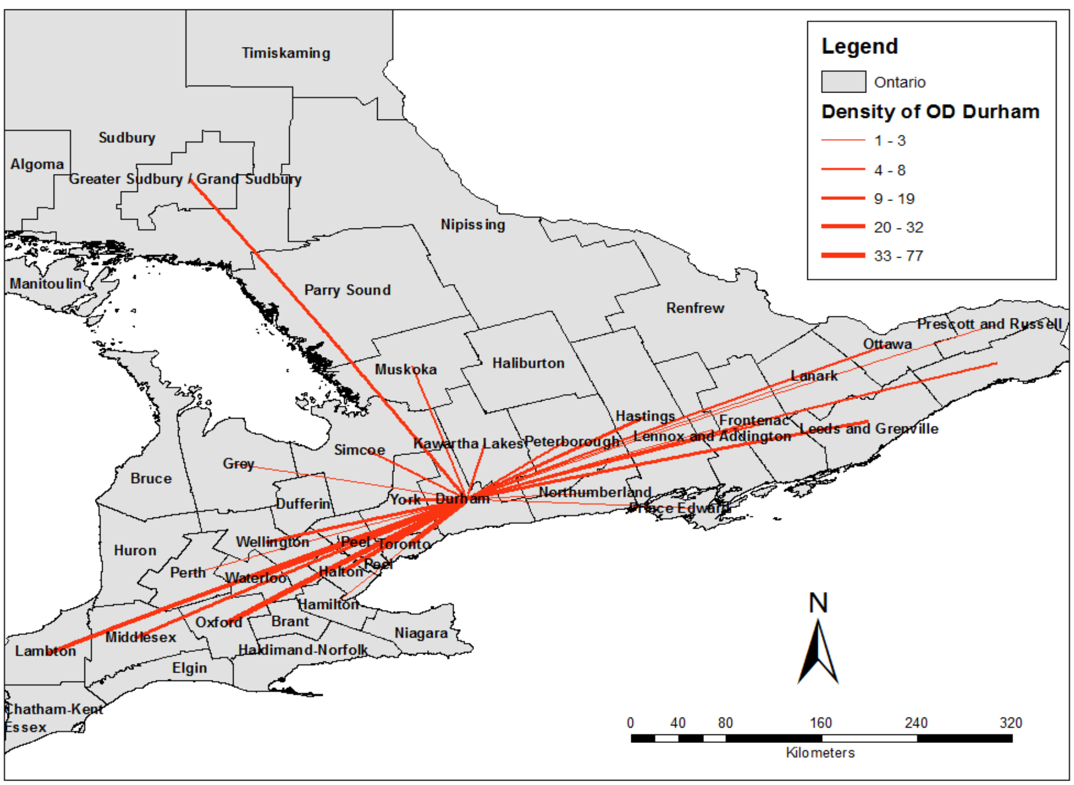


Figure 3-15: Origin-Destination trip of Durham

As mentioned earlier, four types of tours are explored in this study, where each tour has a number of primary/secondary stops. The table below shows the distribution of these stops based on the classification of tours (i.e. k-stop tour). in 1-stop tour are being primary stops (93%), For instance, 2-stop tour has two stops, where these can be a combination of primary, secondary or mix stops. As shown in table 3-6, the stops distribution for 2-stop tour is; one primary and one secondary stops (32%), two primary stops (66%) or two secondary stops (2%).

Table 3-6: Distribution of stop's type per k-stop tour

k-stop tour	Number of tours	One primary stop (%)	One secondary stop (%)	Two primary stops (%)	Two secondary stops (%)	Three primary stops (%)	Three secondary stops (%)	Four primary stops (%)	Four secondary stops (%)
1-stop	3,195	2968 **(93%)**	227 **(7%)**	0	0	0	0	0	0
2-stop	776	0	251 **(32%)**	510 **(66%)**	15 **(2%)**	0	0	0	0
3-stop	128	0	0	55 **(43%)**	23 **(18%)**	48 **(38%)**	2 **(2%)**	0	0
4-stop	12	1 **(8%)**	0	5 **(42%)**	0	2 **(17%)**	0	3 **(25%)**	1 **(8%)**

CHAPTER 4: METHODS OF ANALYSIS

4.1 Modeling Framework

The work in this thesis introduces a tour-based micro-simulation modeling framework that incorporates a number of integrated sub-modules for generating full tours for individual establishments and associated trucks. The framework, shown in Figure 4-1, consists of three modules: (1) Tour Generation Module, (2) Tour Time Module, and (3) Tour Stop Module.

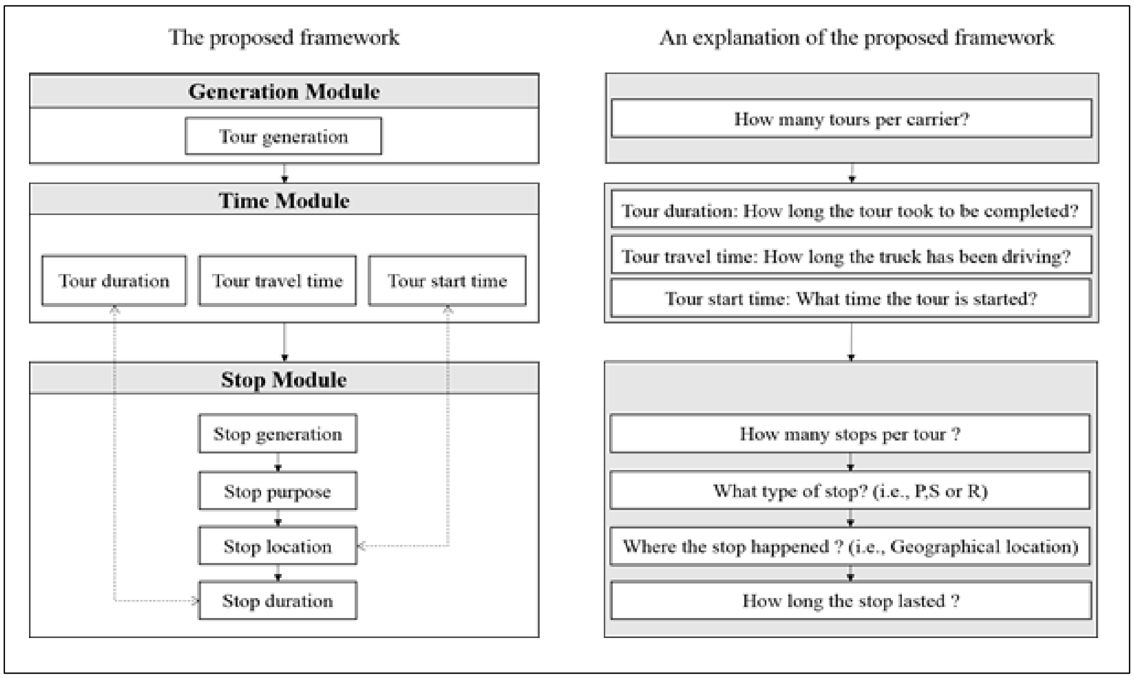


Figure 4-1: Proposed modeling framework for tour-based micro-simulation model

The proposed framework extends the general approach presented in Gingerich et al. (2015), which was only focused on two components of the third module; namely stop generation and stop purpose models. In essence, the extended framework incorporates the two models from Gingerich et al. (2015) in a more comprehensive system that starts by determining the number of tours per carrier or establishment. This is the first module of the proposed framework. Once the number of tours has been determined for each establishment, the second module (i.e. time module) is engaged. Within this module a total of three models are proposed: a tour duration model, a tour travel time model and a tour start time model. The tour duration model will be used to determine the length of the tour from start to end in minutes. A tour travel time model will also be used to predict the total time spent for driving. This part represents the *Service Area* covered by the tour and will become instrumental in the tour stop location model. Finally, within the second module, a tour start time model will be used to predict the start time of the tour. This could be done following a similar approach to the one used in Hunt and Stefan (2007).

The third module of the framework is concerned with the stops comprising the tours. Within this module, the stop generation model developed by Gingerich et al. (2015) will be employed. The latter is an ordered logit model which determines the number of stops per tour (i.e. 1-stop, 2-stops, 3-stops, etc.). Upon determining the number of stops per tour, the type of the generated stops will be predicted using a multinomial logit (MNL) model. Such model, as shown by Gingerich et al. (2015), predicts the probability of the stop being a primary, secondary or return to establishment.

Next, a stop location model is proposed to predict the exact location of the generated stops. Given the predicted information from the previous two steps and the travel time (i.e., service area) associated with the truck making the tour from the second module, a list of potential stops from the universe of all stops of a particular type in the study area will be selected and used to form the choice set for possible destinations. Tours will be classified as 1-stop, 2-stop, 3-stop or 4-stop tours. Next, MNL models will be developed and used to predict the destination location of the stops forming the tours. For instance, if the modeled tour is a 3-stop tour where the stop purpose model has determined that the first two stops are primary, the third is secondary, then the stop location model will engage a MNL model for primary and secondary stops. A MNL-1 to determine the location of the first primary stops based on the formulated choice set. Next, another MNL-2 will be engaged to determine the location of the second primary stop. This is followed by engaging a third MNL for the secondary stop MNL-3 to determine the location of the stop based on the list of potential stops within the determined service area.

Finally, a stop duration model is introduced as the last stage to predict the length of each stop made by the truck within the tour. As a starting point, information from the second module; namely tour duration and tour travel time, will be utilized here. In essence, the difference between the tour duration time and the tour travel time represents the dwell time at all visited stops. Therefore, the total dwell time resulting from the second module will be used as a constraint in the tour duration model. Here, a hazard duration model could be used to predict the stop duration.

The proposed micro-simulation framework will enable us to produce refined Origin-Destination (OD) matrices that could be used as input to freight transportation planning models. Conventional models use the gravity model to estimate OD matrices. We contend that the microsimulation approach is more suitable for capturing the complexities of spatial interactions between the traffic analysis zones forming an urban area. Therefore, the research in this thesis will mainly focus on the stop location and stop duration models of a micro-simulation freight tour model. It will also model the tour travel time process highlighted in the second module.

4.1.1 Modeling Tour Travel Time

Overview

The extracted 4,111 tours contain information about travel distance and travel time for each tour. Travel distance represents the total distance of a particular truck tour, where cumulative travel time represents the total time of a particular truck tour including the dwell time. However, the purpose of developing a Travel Salesman Problem (TSP) model is twofold: 1) calculate the travel distance and travel time. The calculated travel time will be used to create the service area for a given tour. This step is very vital in modeling the stop destination of a truck, and 2) Observed tours extracted from the GPS data to determine if trucks generally optimize their tours when making multiple stops.

4.1.1.1 Travel Salesman Problem Model (TSP)

The use of visual programming models within a Geographic Information Systems (GIS) environment facilitate the automation of geo-spatial processes by connecting visual models to each other or by modifying geo-processing workflows that are native to the utilized software. In this study, a GIS extension called “ModelBuilder” within the ArcGIS 10.6.1 software was used to automate the generation of trucks tours. The purpose of this exercise was to explore the observed tours extracted from the GPS data to determine if trucks generally optimize their tours when making multiple stops. Such process is referred to as the Travel Salesman Problem (TSP). The TSP assumes that a truck starting from the establishment will make several stops before returning to the establishment such that once a stop location is visited the truck will not return to that location. In this model, the location of the establishment along with the location of the different stops for each tour is provided as an input to run the TSP within ArcGIS 10.6.1.

ModelBuilder is an integral component of ArcGIS. It is a visual programming language for creating, automating or modifying geo-processing models by combining different tools from the Arc Toolbox of ArcGIS (ESRI, 2019). The ModelBuilder features an iterator which can save time for repeated or recursive processes. A model is usually represented as a diagram that chains together sequences of processes and geo-processing tools. The advantage of using the ModelBuilder is the ability to build a complex model visually without the need to engage in computer program coding. While running the TSP could be done for each tour manually, the problem becomes very time consuming when dealing with a large number of tours. Therefore, a model for automating the creation of trucks tours was created using the ModelBuilder in this study. The model diagram in Figure 4-2 represents a geoprocessing workflow with multiple processes strung together. The model was developed in such a way that can be used by any users and in future applications given the fact that each step is parameterized. The developed model consists of five tools/steps: 1) three core tools, and 2) two formating tools. Each step utilizes different tool, and the output for each service is used as the input for the step that follows (Table 4-1).

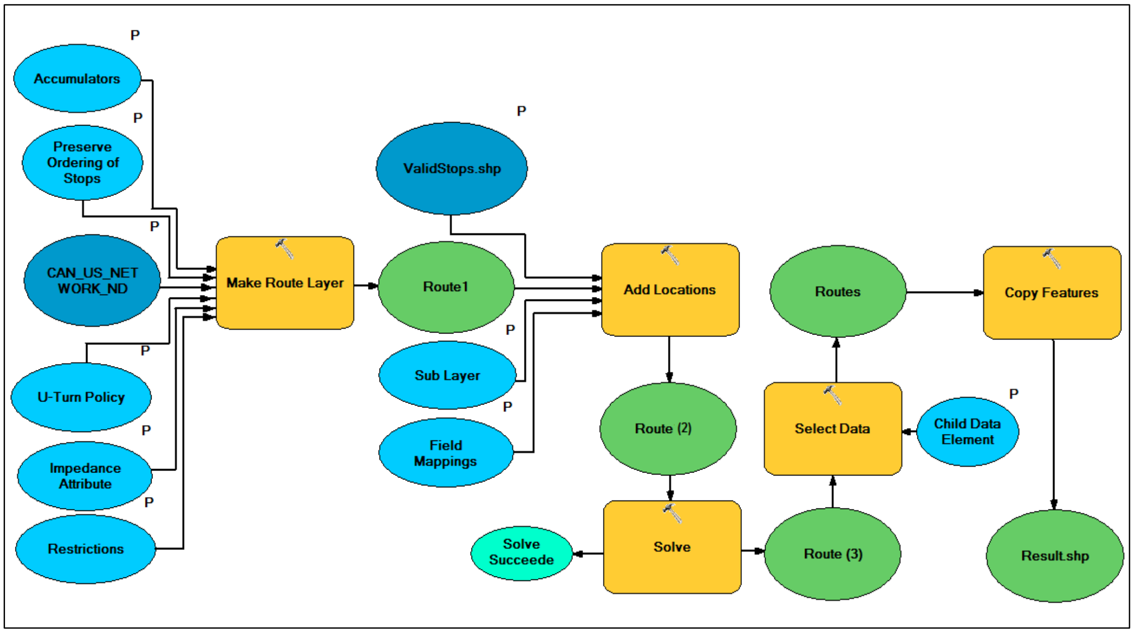


Figure 4-2: Travelling Salesman Problem (TSP) model

As mentioned earlier, the model consists of three core steps:1) make route layer; 2) add locations and 3) solve. For example, step one utilizes a "Make Route Layer" tool using the Travel Salesman Problem (TSP) method from the Network Analyst ToolBox, where the "Network_Dataset.nd*" is used as an input to creat a route layer in conjunction with the five parameters: "accumulators", "Preserve ordering of stops", "U_Turn Policy", "Imperdence" and "Restrictions". This creates "Route 1", which is then inputted into the "Add Locations" tool along with "Valid_Stops.shp*" to allocate the stops on the "Route 1" with the help of two parameters: "Sub Layers" and "Field Mappings". This is used to create "Route 2", which is then inputted into the "Solve" tool to determine "Route 3". The final two steps—Select Data and Copy Features are used to put the output consistent format to make subsequent analysis easier for the Results section.

Table 4-1: Breakdown of the TSP model Components

Steps	Tools	Inputs	Parameters	Output
1	Make Route Layer	Network_Dataset.nd*	**Accumulators** *(i.e., travel time and travel distance)* **Preserve Ordering of Stops** *(i.e., follows the order of stops for a given tour)* **U-Turn Policy** *(i.e., follows the network dataset)* **Impedance Attribute** *(i.e., travel time)* **Restrictions** *(i.e., no restrictions)*	Route(1)
2	Add Locations	Valid_Stops.shp* Route(1)	**Sub Layer** *(i.e., stops)* **Field Mappings** *(i.e., Route_ID)*	Route(2)
3	Solve	Route(2)	**NA**	Route(3)
4	Select Data	Route(3)	**Child Data Element** *(i.e.,Routes)*	Routes
5	Copy Features	Routes	**NA**	Result.shp**

*Input file provided by the user ** An output in a shapefile format

In short, the model was developed in such a way that it can recognize each stop in a tour. For example, Tour_ID 25 has 2 stops, where each stop has the same tour identification. The model will create a route for these stops and record the travel distance and travel time for that tour. The green pins in Figure 4-3 illustrates the locations of all the stops recorded in the data collected from Tour_ID 25 before implementing the model, while the red line outlines a suggested optimal route for that tour based on the model. All stops have been inputted in an automated fashion, as have their outputs (i.e., TSP routes for tours).

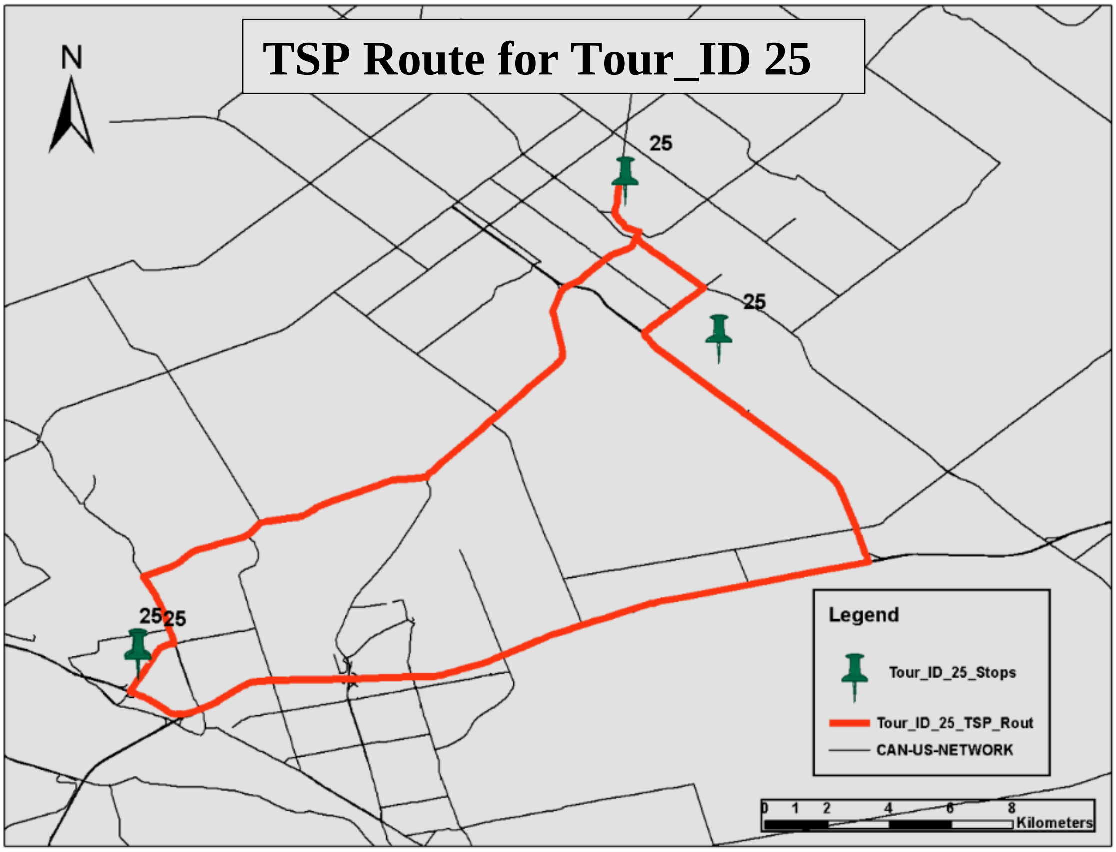


Figure 4-3: Suggested TSP route for tour_ID 25

4.1.2 Modeling Tour Stop Destination Choices

As noted earlier, the stop location component within the third module of the tour modeling framework entails identifying the locations of the stops forming the tour. The problem could be modeled as a destination choice process. Here, a truck must visit one or more stop locations before returning to the establishment. A 30% sample (i.e., 959 tours) was randomly selected from the 3,195 1-stop tours to be used in the modeling work. The random sample was chosen to reduce the number of tours given the long computational time that was required to process these tours in ArcGIS, while the rest of other type of tour classes (i.e. 2-stop, 3-stop and 4-stop tours) were not exposed to random sampling due to the small number of tours in these tours (see Table 4-2).

Table 4-2: Random sample of k-stop tour

k-stop tour	Number of tours	Random sample
1-stop	3,195	959 **(30%)**
2-stop	776	776 **(100%)**
3-stop	128	128 **(100%)**
4-stop	12	12 **(100%)**

4.1.2.1 Modeling Tour Service Areas

We start by developing a model to determine the service area containing the truck tour. Figure 4-4 provides an example of a service area for an establishment (green dotted point) where a truck tour started and visited three stops (red triangles) before returning to the establishments. This step is vital to defining the list of potential stops *N* (blue and red triangles) that will be used in the formulation of the stop destination choice model. Failing to do so will result in an extremely large number of stops to choose from, which will be unrealistic and computationally prohibitive.

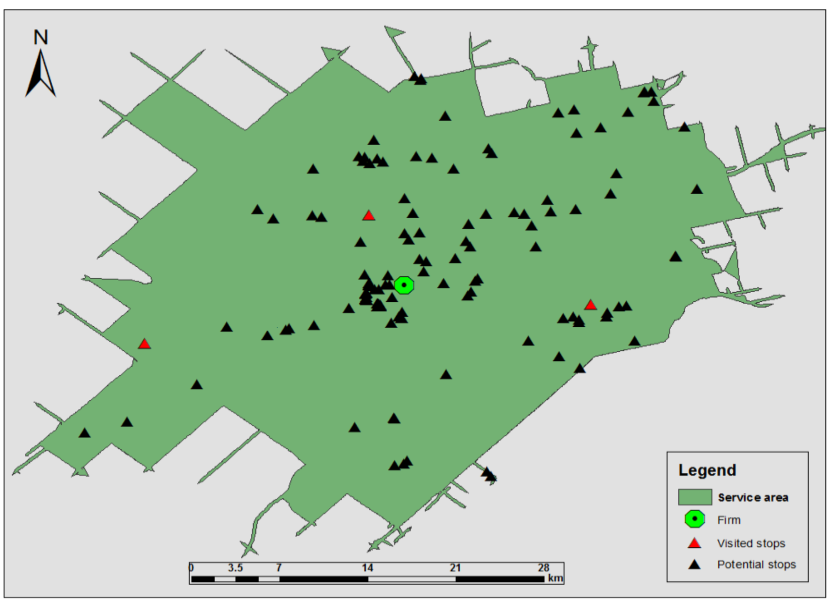


Figure 4-4: An example of a service area of one tour

The tour service area can be delineated on the transportation network based on the calculated travel times required to visit the stops that belong to the tour, not including dwelling time. The travel time is obtained by running the TSP model of the Network Analyst of ArcGIS. Since the TSP time is in minutes (i.e., continuous and non-negative value), regression would be a sensible technique to predict the TSP time when running micro-simulations especially that the stop would not be known at that stage of the micro-simulation. Accordingly, the TSP travel time t_r can be used as the dependent variable in a multivariate regression model, which can be formulated as follows:

$$\ln(t_r) = \beta_0 + \beta_1 X1_r + \cdots\cdots\cdots + \beta_R XR_r \tag{1}$$

In the above formula, each β is a parameter that will be estimated for the specified covariates X that pertains to the trucks, tours, and/or types of industry serviced by the truck. Once t_r is determined by the model, the service area per truck establishment can be calculated using ArcGIS. Table 4-3 lists the key variables used in the model. While some of these variables are used as is, some are combined in the form of interaction terms to account for observed heterogeneity. It should be noted that TSP travel time is directly related to the size of the service area and as such the specified covariates in equation 1 will either contribute to increasing or decreasing the size of the service area depending on the achieved sign of the parameters associated with these covariates.

Table 4-3: Explanatory variables for the tour service area model

Variable Name	Description
Tour class	Class of the available tours:1= 1-stop tour; 2= 2-stop tour; 3=3-stop tour; and 4=4-stop tour
Metropolitan(i)	1 if tour starts from metropolitan *i*; 0 otherwise
Industry(n)	1 if stop pertains to industry type *n*; 0 otherwise
Time of Day	Start hour of the tour
AM Peak	1 if tour starts in morning rush hour (6am - 8am); 0 otherwise
Morning	1 if tour starts in morning hours (9am - 11am); 0 otherwise
Afternoon	1 if tour starts in afternoon hours (12pm – 2pm); 0 otherwise
PM Peak	1 if tour starts in afternoon rush hour (3pm – 5pm); 0 otherwise
Evening	1 if tour starts in evening hours (6pm – 8pm); 0 otherwise
Night	1 if tour starts in night hours (9pm – 5am); 0 otherwise

The tour class variable represents the number of stops for a given tour (e.g. if a tour has three stops, then the tour class value is 3). It is expected a higher tour class to be associated with a larger service area. The majority of the tours used in this model are 1-stop tour (77%). This simplified the process of defining the industry associated with the tour. In the model, the tour was assigned the industry associated with the first primary stop serviced by the truck. We do not have a clear direction with respect to the impact of the industry on the size of the service area, but we generally believe that some industries are more likely to have smaller services area compared to other. The starting time of the tour was also used in the model. Here, we generally anticipate that tours starting in the morning and mid-day to have a smaller service area because they have a higher probability to start new tours after those tours. As for the origin, the location of the starting zone of the first stop is also used. Similar to the industry, we do not have a specific direction about the relationship between the origin of the tour and the size of the service area but we expect certain municipalities within the Greater Toronto and Hamilton Area (GTHA) to have smaller service areas relative to other municipalities.

4.1.2.2 Modeling Tour Stop Destination

The discrete choice modeling technique namely, the MNL model, is used to develop the tour stop destination component of the tour microsimulation framework. The MNL is used to model the choice behavior associated with the selection of the destination stops forming the tours made by truck. The econometric analysis for the discrete choice modeling is performed in the NLOGIT 5 software. Separate MNL models will be developed for 1-stop, 2-stop, and 3-stop tours. For instance, if the truck is going to make a 2-stop tour, then two separate model will be engaged; an MNL model for the first stop and another MNL for the second stop. The probability that stop ***i*** is the chosen destination from a set of alternative stops $j = 1, 2, \ldots, N$ can be estimated as follows:

$$P(i) = \frac{e^{Vi}}{\sum_{j=1}^{N} e^{Vj}} \tag{2}$$

Where $\boldsymbol{V_j}$ is a linear-in-parameters systematic utility function that depends on the characteristics of the stop and the attributes of the truck and/or establishment. Table 4-3 provide the list of explanatory variables that are used in the specification of the utility $\boldsymbol{V_j}$**.** The type of stop (i.e., primary or secondary) is expected to impact the choice probability with primary stops having a higher probability of being chosen. This positive relation can be attributed to the fact that there is at least one primary stop in a given tour regardless of its number of stops. Further, 1-stop tours are more likely to take place for deliver or pickup goods from primary stops. The cumulative time involved to complete a stop is considered in the model specification. This time represents the time spent driving the truck between the establishment and a potential stop plus the time spent at the stop (i.e. stop duration). We hypothesize that longer cumulative time for a stop will increase its chance of being chosen in the case of 1-stop tour. This is particularly the case since the modeled tours are long-haul tours by nature and as such further stops are more likely to be plausible destinations. However, when the number of stops for a given tour is higher than one stop, then a different assumption is imposed. More specifically, we hypothesize that the parameter of the cumulative time variable in the 1st stop model of the 2- and 3-stop tour models will have a negative sign because the 1st stop is usually a starting-up stop and as such it is more likely to be much closer to the establishment compared to 2nd or 3rd stops.

Besides the cumulative time, a travel distance variable is also introduced in the model specification. Travel distance represents the distance between the establishment and a potential stop. We hypothesize that the travel distance will have a positive impact on the choice probability in the 1-stop tour model again due to the long-haul nature of the modeled tours. In the case of 2- and 3-stop tours, the probability is expected to decrease for the 1st stop since many of such stops in a multi-leg tours are more likely to be secondary stops for the purpose of fueling. Such activity is more likely to happen at a location closer to the establishment. Further, the parameter of the distance variable in the 2nd stop model of 2- and 3-stop tours is expected to have a positive sign because there is a higher chance that this stop will be a primary stop. The positive affiliation could be attributed to the long-haul nature of the modeled tours. However, the impact of travel distance on the choice of the stop in the 3rd stop model for the 3-stop tour is difficult to be expected. While the cumulative time and travel distance are introduced in the model specification, they are also used to derive a number of categorical (i.e. dummy variables) to represent different ranges of time and distance. These variables are introduced to account for any possible heterogeneity in the destination choice behavior.

A total of ten industry dummy variables were created and used in the specification of the models. Each of these industry dummies specify the purpose of the stops and as such are expected to have different impacts on the modeled choice probabilities. Location dummy variables were also considered in the model specification. Stops in the choice set were categorized based on the metropolitan area they fall in. Again, we expect the choice probability for stops to differ based on the location of these stops. That is, stops at certain locations are more likely to be chosen, while this is not the case for certain stops at other locations. Starting time dummy variable are also introduced in the model. Here, a total of six variables are derived to represent the six key periods of the day: AM peak, morning, afternoon, PM peak, evening and night. We hypothesize that tour will usually start in the morning. Therefore, the AM peak variable is expected to have a positive sign in the case of the 1-stop tour model. The same is also expected for the 1^{st} stop of the 2- and 3-stop tour models.

The enclosed angle (Theta) between the chosen stop and the previous location from where the truck started that leg of the tour is introduced in the model specification. Figure 4-5 provides a visual depiction of the enclose angle, which is measured counter clockwise from the x-axis where the start location is. The start location could be the establishment or a previously visited stop. The value of theta will range from 0^o to 359^o. The enclosed angle is introduced to capture the impact of tour leg directionality on the choice probability. Hunt and Stefan (2007) used the enclosed angle between the zones of two consecutive stops in their stop destination choice model. Besides the enclosed angle, we introduce four dummy variables to represent the quadrants for which the chosen stop belongs to with respect to the start location of the tour leg, as shown in Figure 4-5. These variables along with the enclosed angle are used to capture the spatial orientation of the chosen stops. We do not have prior knowledge about the impact of these variables on the choice probabilities.

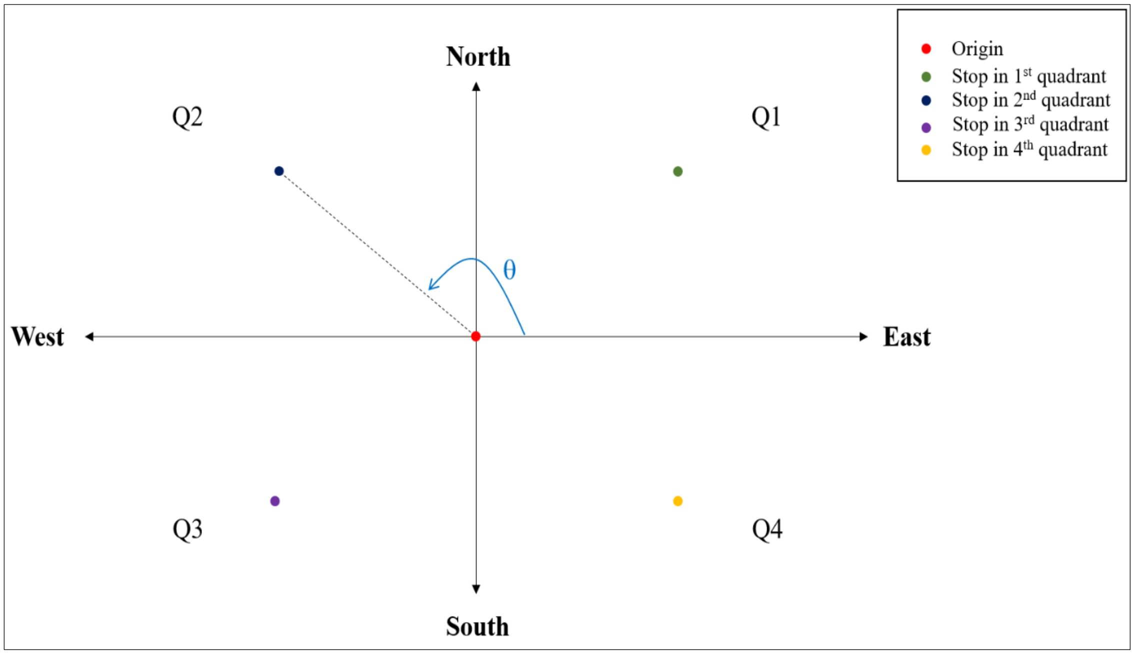


Figure 4-5: Depiction of the angle between a starting point and a stop of a truck

It is worth noting that many of the dummy variables introduced in Table 4-4 are used in the model specification as interaction terms to account for the observed heterogeneity in the modeled stops. In many instances, the hypothesis is focused on testing how certain combined characteristics influence the choice probabilities of the modeled stops. The creation of interaction terms was done by using one of three options: (1) the metropolitan housing the stops, (2) the industry the stops are serving, or (3) the direction of the visited stop with respect to the location where the tour-leg started.

Table 4-4: Explanatory variables for the stop destination model

Variable Name	**Description**
Primary stop	1 if the stop is primary; 0 otherwise
Cumulative time (CT) *CT Less than* 50_{min} *CT Between* 50_{min} *and* 150_{min}	Cumulative time (minutes) involved in completing a stop 1 if cumulative time is less than 50 minutes; 0 otherwise 1 if cumulative travel time is between 50 and 150 minutes; 0 otherwise
Travel distance (TD) *TD Less than* 50_{km} *TD Between* 50_{km} *and* 100_{km} *TD Between* 100_{km} *and* 150_{km} *TD Between* 150_{km} *and* 200_{km} *TD Between* 200_{km} *and* 250_{km} *TD Over* 400_{km}	Travel distance (km) from the origin of stop to the destination 1 if travel distance less than 50 km; 0 otherwise 1 if travel distance between 50 and 100 km; 0 otherwise 1 if travel distance between 100 and 150 km; 0 otherwise 1 if travel distance between 150 and 250 km; 0 otherwise 1 if travel distance between 200 and 250 km; 0 otherwise 1 if travel distance over 400 km; 0 otherwise
Industry(n)	1 if stop pertains to industry type *n*; 0 otherwise
Metropolitan(i)	1 if stop is located in the same metropolitan *i* as the previous stop; 0 otherwise
Metropolitan(ij)	1 if stop is located in the metropolitan *j* which is not the same as previous stop's metropolitan *i* (i.e., $i \neq j$); 0 otherwise
AM Peak *Morning* *Afternoon* *PM Peak* *Evening* *Night*	1 if stop is visited in morning rush hour (6am - 8am); 0 otherwise 1 if stop is visited in morning (9am - 11am); 0 otherwise 1 if stop is visited in afternoon (12pm – 2pm); 0 otherwise 1 if stop is visited in afternoon rush hour (3pm – 5pm); 0 otherwise 1 if stop is visited in evening (6pm – 8pm); 0 otherwise 1 if stop is visited in night (9pm – 5am); 0 otherwise
Enclosed Angle (*Theta*) Q_1 Q_2 Q_3 Q_4	The direction of the stop with respect to the origin 1 if stop going to first Cartesian quadrant; 0 otherwise 1 if stop going to second Cartesian quadrant; 0 otherwise 1 if stop going to third Cartesian quadrant; 0 otherwise 1 if stop going to fourth Cartesian quadrant; 0 otherwise

4.1.3 Modeling Tour Stop Duration

Overview

This study aims to model truck stop duration as a survival analysis problem for trucks dwelling time. Survival analysis is suitable in our context because it is used to estimate the lifespan of an event. According to the literature, survival analysis is also known as time-to-event analysis. The terms “stop duration” and “dwell time” are identical; however, to avoid any confusion, the current study will use the term “dwell time”. The dwell time represents the time it took a truck to load/unload the shipment at the stop. The dwell time was derived and calculated from the dataset. The dataset has information (i.e., a time stamp) just before the stop has happened and after the stop has took place. The difference between the two-time stamps represents the stop duration. In this study, time-to-event represents the dwell time from the moment the truck stops at a location until it departs.

Modeling the survival function can be done through one of three techniques: parametric, semi-parametric, and non-parametric. Since this research will focus on the semi-parametric and non-parametric modeling techniques, the following two sub-section will clarify why these two techniques were utilized.

4.1.3.1 Semi-Parametric Modeling

As the name suggests, the semi-parametric technique is a mix of parametric and non-parametric components. In parametric hazard duration models, the analyst starts by assuming that the used data follow a known statistical distribution such as the Log-normal, Exponential, Gompertz, and Gamma to name a few. However, this is not the case in semi-parametric models. Oakes (1977) used the maximum likelihood techniques to estimate a semi-parametric model to obtain more efficient estimation compared to the parametric technique. The Cox proportional hazard (CPH) model is the most extensively used semi-parametric model since it allows incorporating covariates to explain and predict the dwell time. The CPH is also popular because it avoids the possible misspecification of the hazards functional form. Thus, it is easier to formulate the effects of time-independent covariates through this model. The CPH model consists of two multiplicative parts as shown in the following formula:

$$h_i(t) = \lambda_0(t) \exp(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \tag{3}$$

In this formula, $h_i(t)$ is the hazard for truck *i* to dwell for a duration *t* at a given stop, $\lambda_0(t)$ is the unspecified baseline hazard which describes how the risk of completing the stop (i.e., finish dwelling) changes over time at a baseline levels of the specified covariates ($\lambda_0(t) > 0$), and, exp(.) is the parametric part which describes how the hazard changes in response to the explanatory variables; it also represents the hazard ratio (HR). Here, β_i is a set of parameters that are associated with the x_i explanatory variables. The $\beta_i's$ in this model can be estimated using the partial likelihood method by formulating and maximizing the following function:

$$PL(\beta) = \prod_{t_k:event\ at\ t_k} \frac{h_0(t)e^{\beta x(t_k)}}{\sum_j h_0(t)e^{\beta x(t_j)}} \tag{4}$$

The hazard in the CPH model is calculated to happen at a certain time *t*, say t_k. If we divide time into small intervals *j* = {1, 2, 3, 4, ..., *k*, ..., *J*}, then the probability that the hazard will occur at time interval t_k is:

$$\frac{h_0(t)e^{\beta x(t_k)}}{\sum_j h_0(t)e^{\beta x(t_j)}}$$

Also, the probability that the hazard will happen at time t_1, t_2, t_3 ..., and t_J is:

$$\frac{h_0(t)e^{\beta x_{(t_1)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}}, \frac{h_0(t)e^{\beta x_{(t_2)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}}, \frac{h_0(t)e^{\beta x_{(t_3)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}}, \dots, \text{and } \frac{h_0(t)e^{\beta x_{(t_J)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}}$$

The chance that the hazard will happen at any time interval 1, 2, 3, ..., k, ..., J is the Joint Probability which is the product of the above probabilities:

$$PL(\beta) = \frac{h_0(t)e^{\beta x_{(t_1)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}} \times \frac{h_0(t)e^{\beta x_{(t_2)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}} \times \frac{h_0(t)e^{\beta x_{(t_3)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}} \times \dots \times \frac{h_0(t)e^{\beta x_{(t_k)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}} \times \frac{h_0(t)e^{\beta x_{(t_J)}}}{\sum_j h_0(t)e^{\beta x_{(t_j)}}}$$

The above product is basically equation 4. The CPH model is estimated by finding the values of betas that maximize the joint probability PL. These betas will provide the best chance of predicting the correct time interval for the event to occur.

The CPH model will be used in the analysis of time-to-event data along with censoring and covariates. The dwelling time in this data varies and depends on the stop type (i.e., primary stop or secondary stop). As a result, two CPH models will be developed to predict the dwelling time with respect to the stop type. The process outlined in Figure 4-6 will be used to generate the dwell time to differentiate between censored data and event data. However, it is not easy to determine the length of the stop given no prior information on this in the literature. As a result, this model will rely on the average dwelling time with respect to the stop type. For the case of the primary stop model, the average of the dwelling time is calculated to be used in this model as the critical value τ which is 55 minutes. That is, if the dwell time is less than *55* minutes, the observation will be censored; otherwise if it is greater than or equal to *55* minutes, it will be labeled as an event. Similarly, the average of dwelling time is calculated for the secondary stop model and found to be 20 minutes. The calculated value will be used to differentiate between event and censored data following the same logic as in the case of the primary stop model.

A number of explanatory variables, as shown in Table 4-5, are devised and included in the CPH model. As for the primary stop model, we hypothesize that tours associated with higher number of stops will likely have longer dwelling time. Further, we expect some variability in the dwell time based on the type of industry serviced by the truck. We anticipate some variability for tours taking place in certain metropolitan areas. As for the secondary stop model, we expect that, as the travel time increases, the dwell time also increases, while other variables (e.g. metropolitan location, time of day, etc.) are more likely to decrease the dwell time since the modeled stops are secondary.

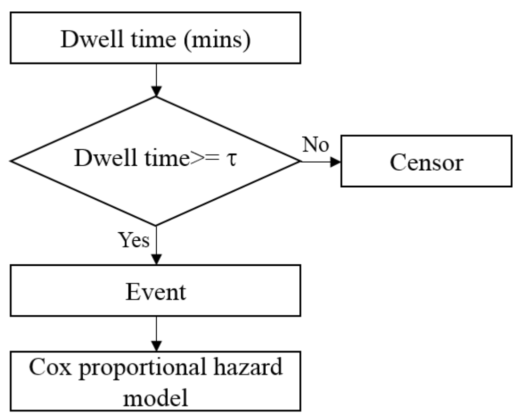


Figure 4-6: Classifying dwell time to event or censor

Table 4-5: Explanatory variables for the stop duration model

Variable Name	Description
Tour class	Class of the available tours:1= 1-stop tour; 2= 2-stop tour; 3=3-stop tour; and 4=4-stop tour
Metropolitan(i)	1 if stop is located in the same metropolitan *i* as the previous stop; 0 otherwise
Metropolitan(ij)	1 if stop is located in the metropolitan *j* which is not the same as previous stop's metropolitan *i* (i.e., $i \neq j$); 0 otherwise
Industry(n)	1 if stop pertains to industry type *n*; 0 otherwise
Time of Day	Start hour of the tour
AM Peak	1 if stop is visited in morning rush hour (6am - 8am); 0 otherwise
Morning	1 if stop is visited in morning (9am - 11am); 0 otherwise
Afternoon	1 if stop is visited in afternoon (12pm – 2pm); 0 otherwise
PM Peak	1 if stop is visited in afternoon rush hour (3pm – 5pm); 0 otherwise
Evening	1 if stop is visited in evening (6pm – 8pm); 0 otherwise
Night	1 if stop is visited in night (9pm – 5am); 0 otherwise

4.1.3.2 Non-Parametric Modeling

Typically, non-parametric techniques are used to visualize how the model behaves with respect to other significant variables. In this regard, the Kaplan-Meier model is considered as one of the most widely recognized non-parametric models that can be used to provide survival probability curves. What distinguishes this model is its capability of estimating the survival function in the presence of censoring. However, the outcome of the survival function curve is not smooth. The survival function is the probability that a truck will still be dwelling beyond time *t*,

$$\mathrm{S}(t) = \mathrm{P}(T > t), \qquad 0 < t < \infty \qquad (5)$$

The survival probability curve distinguishes between the model for all variables against the model of each variable. The graphical assessment studies the distributional characteristics of the dwell time in order to determine whether the developed model for all variables is sufficiently different when compared to the model with each variable only.

CHAPTER 5: RESULTS AND DISCUSSION

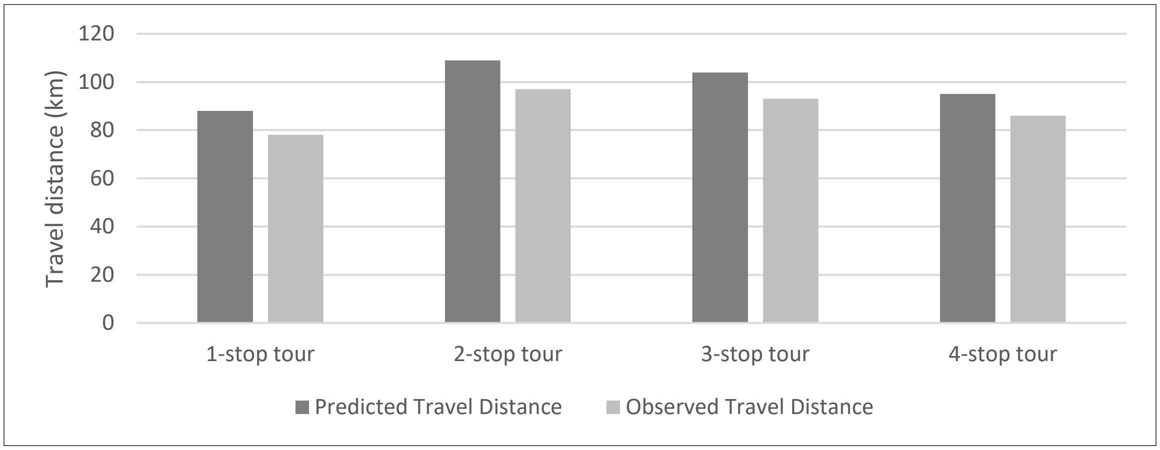
5.1 Travel Salesman Problem in GIS

The results obtained from running the Travel Salesman Problem (TSP) model in ArcGIS provide interesting insights about travel distance and travel time for the analyzed tours. The predicted travel distances for the tours are closely related to the observed ones obtained from the GPS data. However, the predicted travel time values were less reliable. Table 5-1 shows the Pearson correlation between the observed and predicted values for both travel distance and travel time.

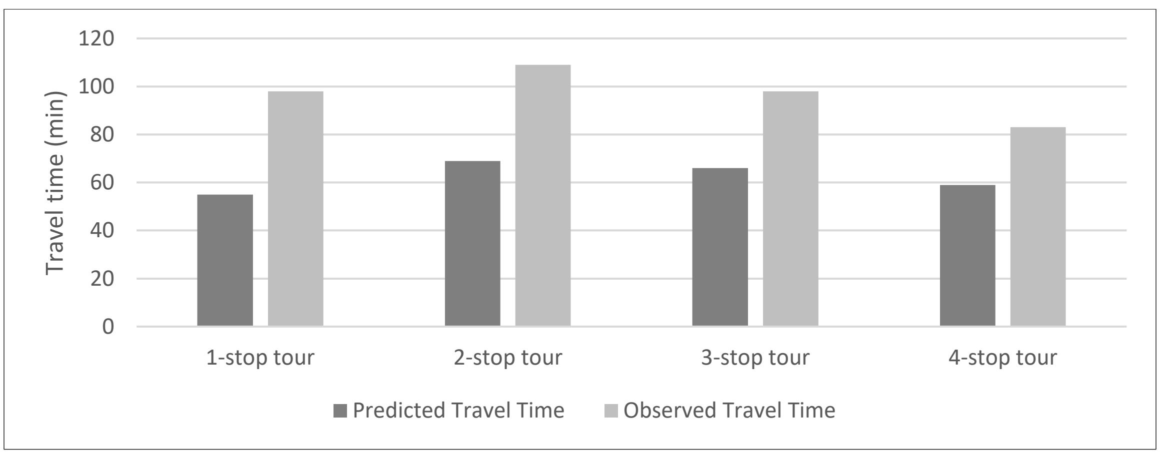
Table 5-1: A summary of correlation factor for travel time and travel distance

Variable	Correlation
Travel Distance	0.99
Travel Time	0.77

In general, travel distance for trucks is usually expected to follow the least resistance path on the road network. As such, the application of a TSP model while using length (km) of the road links as impedance should produce comparable results to those observed from the GPS pings. In that sense, the travel distance can be said to be highly predictable. By comparison, the use of free-flow travel time of the road links as impedance might not produce comparable results with the observed values. The observed travel time from the GPS pings is normally influenced by several factors that may include: 1) ongoing road construction, 2) presence of other vehicles or incidents on the road, and 3) the time waited at truck inspection stations. Figure 5-1 presents a comparison between the predicted and observed average road link impedance with respect to travel distance and travel time by type of tour. The results indicate that the predicted average travel distances is very close to the observed ones, while this is not the case for the predicted and observed average travel times. In the case of the latter, the predicted values are systematically lower than the observed values.



(a)



(b)

Figure 5-1: Comparison of predicted and observed (a) travel distance, and (b) travel time

5.2 Tour Service Area Model

The tour service area model is developed to predict the total minutes that will be covered by a given tour on the transportation network as described in the previous chapter. Here, a linear-regression model is estimated and the results are presented in Table 5-2, with all of the coefficients and their corresponding t-statistics. To avoid any confusion in reading the table, a positive coefficient suggests a larger service area —which means a longer tour— while a negative coefficient indicates a smaller service area —which means a shorter tour. According to the model, the size of the service area (minutes) for the modeled tours can be explained by the number of stops made in the tour, the municipality where the tour originated from within the Greater Toronto and Hamilton Area (GTHA), the time of the day when the tour started and the type of industry associated with the tour. Also, the specified interaction terms in the model reveal several observed heterogeneities with respect to the nature of the tours.

Based on the estimated coefficients, tours that have more stops tend to have larger service areas than the ones that have fewer stops. This is not surprising since a truck making more stops in a given tour is more likely to cover a larger geographic area on the transportation network. Relative to tours starting from the different municipalities within the GTHA, tours starting from Hamilton and Halton have smaller service areas, other things being equal. This means tour originating from Hamilton and Halton are short and most likely to be done within/around the GTHA. Although the effect is more pronounced in the case of the Halton tours in general especially for the tours starting during the evening period. Interestingly, the size of the tours from Halton that are associated with the manufacturing sector are not as small as the other tours originating from this municipality as depicted by the positive coefficient of the interaction term (*Halton* × *Manufacturing*).

While the majority of the tours starting during the AM peak period are associated with smaller service areas, the opposite is observed for the tours originating from Toronto during that period. Similarly, the size of the Toronto-based tours that are associated with the construction industry tend to be larger than other tours, other things being equal. By comparison, the size of the service areas tends to be smaller for the Toronto-based tours that start during the PM peak, evening and night periods. The same could be said about the tours associated with the service industry and which originate from this municipality.

According to the model, the size of the service areas for the Peel-based tours starting in the AM peak period tends to be the smallest when compared to all other tours starting during that period, other things being equal. However, the opposite effect is observed for the tours starting from this municipality during the evening period as discerned from the coefficient of the term (*Peel* × *Evening*). Here, tours start at night are more likely to go longer distance unlike the ones start in the morning and go for shorter distance. This makes sense because traffic at night is not as busy as it is in the morning. Therefore, carriers are very unlikely to dispatch trucks in the morning for long tours but instead send trucks at night. Similar to the AM peak case, tours starting during the morning period tend to be smaller in terms of their service area and this is even more pronounced in the case of the York-based tours as depicted by the coefficient of the term (*York* × *Morning*). Further, York-based tours that are associated with the manufacturing industry also tend to have a smaller service areas. This could suggest that the York tours are more localized in nature.

Durham-based tours starting in the evening period tend to have larger service areas, other things being equal. The same could be said about the Durham-based tours that are associated with the retail trade industry. By comparison, the opposite can be seen for the Durham-based tours that are associated with the public administration sector. While all tours that are associated with the public administration sector tend to have smaller service areas, those originating from Durham tend to be the smallest as depicted by the coefficient of the term (*Durham* × *Public Administration*), other things being equal. It is worth noting that, tours associated with light duty industries (e.g., services, public administration, and etc.) are more likely to be short and stay within/around the GTHA. Tours associated with the transportation industry tend to have a smaller service areas. Likewise, tours associated with the retail trade industry and which start during the PM peak period tend to have smaller service areas. Finally, afternoon-based tours from all industries tend to have smaller service areas except for the tours associated with the finance industries. The latter tend to have larger service areas as illustrated by the positive coefficient of the term (*Finance* × *Afternoon*).

Table 5-2: Regression estimation results for tour service areas

Variable Name	Beta	t-stats
Constant	4.615	184.43
Tour class	0.465	30.51
Origin		
Hamilton	-0.143	-3.64
Halton	-0.256	-7.66
Industry		
Transportation	-0.057	-3.03
Public Administration	-0.173	-2.00
Time of day		
AM Peak	-0.143	-4.81
Morning	-0.091	-3.34
Afternoon	-0.093	-3.52
Origin Vs. Time of day		
Halton × Evening	-0.177	-2.10
Toronto × AM Peak	0.172	2.82
Toronto × PM Peak	-0.242	-2.77
Toronto × Evening	-0.231	-2.67
Toronto × Night	-0.157	-2.88
Peel × AM Peak	-0.074	-1.97
Peel × Evening	0.066	1.85
York × Morning	-0.196	-2.92
Durham × Night	0.305	6.24
Origin Vs. Industry		
Halton × Manufacturing	0.109	1.65
Toronto × Construction	0.396	1.74
Toronto × Services	-0.088	-1.69
York × Manufacturing	-0.149	-2.43
Durham × Retail Trade	0.174	2.07
Durham × Public Administration	-0.868	-1.70
Industry Vs. Time of day		
Retail Trade × PM Peak	-0.102	-1.71
Finance × Afternoon	0.303	2.45
No. of Observations	4,111	
R-square	0.252	

5.3 Modeling Tour Stop Destination Choices

A total of six MNL models are estimated to determine the destination location of the stops made by trucks engaged in generating tours. The estimated models included: one MNL model for the 1-stop tour (Table 5-3); two MNL models for the 2-stop tour, where each stop has a separate MNL model (Table 5-4); and three MNL models for the 3-stop tour, where each stop also has a separate MNL model (Table 5-5). The estimation is performed in the NLOGIT 5.0 software. It should be noted that the concept for developing the different MNL models is similar regardless of the rank of the stop. Here, the truck's choice of a stop from a finite and discrete set of available stops is modeled as a function of the stops characteristics including the type of industry and geographic location of these stops.

5.3.1 One-Stop Tour Destination Choice Model

Table 5-3 presents the results of the one-stop tour destination choice model with all of the coefficients and their corresponding t-statistics. Most of the estimated parameters are statistically significant although few of the interaction terms were marginally significant (i.e., under 90% statistical significance level).

According to the model, the probability of choosing a stop as a destination increases if the stop is classified as Primary, other things being equal. This is not surprising since 1-stop tours are mainly generated to deliver or pickup merchandize and such activity would take place when primary stops are visited. On the other hand, very few 1-stop tours will take place where the visited stop is a secondary stop, albeit such activities might occur from time to time (e.g., truck fueling). Further, the probability of selecting a primary stop increases for tours that have a cumulative time of less than 50 minutes, as illustrated by the coefficient of the term (*CT Less than* 50_{min} × *Primary Stop*).

Table 5-3: MNL estimation results for 1-stop Tour model

Variable Name	Beta	t-stats
Primary Stop	0.371	2.54
Primary Stop × *CT Less than* 50_{min}	0.739	3.38
Cumulative Time (CT)	0.003	4.52
CT × Morning	0.001	2.20
Travel Distance (TD)	0.007	6.32
TD × PM Peak	0.002	1.94
TD × Durham(j)	0.005	3.33
TD × Waterloo(j)	-0.005	-3.09
TD × Simcoe(j)	-0.003	-2.03
TD × Q2	0.003	3.88
Enclosed Angle (θ)	0.001	2.64
CT Between 50_{min} *and* 150_{min}	0.450	2.53
CT Between 50_{min} *and* 150_{min} × *York(i)*	0.306	2.33
TD Between 50_{km} *and* 100_{km}	0.626	6.22
TD Between 50_{km} *and* 100_{km} × *AM Peak*	0.582	3.73
TD Less than 50_{km} × *Morning*	-1.525	-2.94
Retail Trade	0.360	2.68
Peel(i) × Retail Trade	-0.338	-1.87
Oxford(j) × Retail Trade	-2.225	-2.20
Waterloo(j) × Construction	0.721	1.85
Simcoe(j) × Public Administration	1.888	1.66
Hamilton(i) × Morning	0.536	2.39
Durham(j)× AM Peak	-2.101	-3.46
No. of Observations	959	
$LL(0)$	-2746.72	
$LL(\beta)$	-2533.5	
ρ^2	0.077	

The probability of choosing a stop increases when the cumulative time between the stop and the tour origin (i.e., location where the truck started) increases. This might come across as a counter intuitive result given that cumulative time is usually considered a disutility. However, in the context of the modeled tours, the obtained result is not surprising given the long-haul nature of the tours and the fact that the choice set (i.e., alternative stops) is confined by the service area covered by the tour. That is, the size of the delineated service area (minutes) for the tour is strongly correlated to the actual travel time between the stop location and the origin of the truck. As such, the true destinations are more likely to be located at the edge of the service area boundary. When forming the choice set, alternative stops are chosen to fall within the delineated service area. However, since the true stop location is more likely to be located at the edge of the boundary of the service area, stops closer to the origin of the tour are less likely to be chosen while ones far away from it are more likely to be chosen. While the probability of selecting a stop increases with the cumulative time between the stop and the origin, this tends to be more pronounced for tours starting in the morning as depicted by the coefficient of the term (*Cumulative Time* × *Morning*). This is not surprising since tours start in the morning are more likely to spend longer time before returning to their establishments.

Similar to cumulative time, the probability for selecting a stop increases for stops that are far away from the tour origin as discerned by the coefficient of the (*Travel Distance*) parameter. The impact of travel distance is even more pronounced for tours starting during the PM peak period and for stops that are located in Durham. By comparison, the impact of travel distance on the choice probability is reduced for stops located in Waterloo and Simcoe, other things being equal. Further, the impact of travel distance on the choice probability of a stop increases for stops that are located in the second Cartesian quadrant relative to the origin location of the tour.

Tours associated with certain industries exhibit higher probability of being chosen. For example, tours associated with the retail trade industry tend to have a positive impact on the choice probability. However, the impact of the retail trade industry on the choice probability is reduced for stops affiliated with the origin of the tour and when that origin is Peel. This could be attributed to the fact that Peel is more industrial in nature and as such stops in Peel will be less likely to attract retail trade freight activities. Also, the impact of the retail trade industry is even more pronounced for stops located in Oxford, as depicted by the coefficient of the term (*Oxford(j)* × *Retail Trade*). This could be explained by the fact that Oxford has a strong agriculture industry presence, and as such will be attracting transportation or agriculture industries as opposed to retail trade industries. Furthermore, stops associated with certain industries have a higher chance of being selected when these stops are located in foreign locations relative to the home of the tour, as depicted by the coefficient of the term (*Waterloo(j)* × *Construction*). This tends to even be more pronounced for stops linked with public administration and which are located in Simcoe, as depicted by the coefficient of the term (*Simcoe(j)* × *Public Administration*).

As mentioned before, the modeled tours are long-haul, which decrease the probability of choosing stops that are located within the GTHA. This is supported by the negative coefficient of the term ($TD\ Less\ than\ 50_{km}$ × *Morning*). The parameter suggests that the probability of choosing stops from the GTHA will decrease especially in the morning. This is further supported by the results pertaining to the negative parameter of the term (*Durham(j)* × *AM Peak*). Stops located in Durham and visited during the AM Peak period have a negative impact on the choice probability, other things being equal. As established earlier, the probability of selecting a stop increases as the travel distance also increases. However, the impact is more visible for distances ranging between 50 km and 100 km especially for stops visited during the AM Peak period, as depicted by the coefficient of the term (*TD Between* 50_{km} *and* 100_{km} × *AM Peak*).

5.3.2 Two-Stop Tour Destination Choice Models

Table 5-4 presents the results of the two-stop tour destination choice model with all of the coefficients and their corresponding t-statistics. This model consists of two separate MNL models: the first MNL model represents the behavior associated with the choice of the first stop, while the second MNL model represents the behavior for selecting the second stop. Most of the estimated parameters are statistically significant, although a few of the interaction terms were marginally significant (i.e., under 90% statistical significance level). The two MNL models have a rho-squared (ρ^2) value of 0.248 and 0.214, respectively.

First Stop Destination Choice Model

The probability of choosing a first stop decreases when the travel distance between the stop and the tour origin (i.e., location where the truck started) increases. Unlike the travel distance, the probability of choosing a stop increases when the cumulative travel time between the tour origin and the stop increases. The impact of cumulative time is even more amplified for stops lasting less than 50 minutes. This is not surprising since the first stop is considered the start-stop of the tour; therefore, this stop will most likely be the closest to the establishment when more than one stop is involved. Further, the probability of selecting a primary stop increases for stops that have longer cumulative time, as illustrated by the coefficient of the term (*CT* × *Primary Stop*). Also, the impact of cumulative time on the

Table 5-4: MNL estimation results for 2-stop tour model

Variable Name	1st stop model		2nd stop model	
	Beta	t-stats	Beta	t-stats
Primary Stop	--	--	0.731	2.09
Cumulative Time (CT)	0.004	1.95	--	--
CT Less than 50 min	1.344	5.18	--	--
CT × Primary stop	0.012	2.10	--	--
CT × Q2	-0.045	-3.04	--	--
CT × Morning	0.017	4.65	--	--
CT × Simcoe(j)	0.013	2.95	--	--
CT × Middlesex(j)	-0.011	-1.49	--	--
Travel Distance (TD)	-0.027	-4.23	-0.021	-5.68
TD Between 100_{km} *and* 150_{km}	--	--	0.721	1.97
TD × Toronto(j)	--	--	0.019	1.90
Enclosed Angle (θ)	--	--	-0.003	-2.30
Halton(i) × Q1	-1.450	-1.94	--	--
Halton(i) × Q2	3.060	2.55	--	--
York(i) × Q1	--	--	0.894	2.22
York(i) × Services	--	--	-2.399	-2.18
Toronto(j) × Q1	--	--	2.256	2.32
Waterloo(j) × Q1	--	--	-1.308	-2.26
Waterloo(j) × Evening	--	--	2.025	2.22
Retail Trade × Q2	2.710	2.81	--	--
PM Peak × Q4	--	--	2.569	2.13
Construction × PM Peak	1.472	1.80	1.266	1.8 3
Durham(j)	--	--	-2.064	-2.92
Durham(j) × Retail Trade	--	--	1.770	2.16
AM Peak	1.901	5.55	-1.860	-2.27
Toronto(i) × AM Peak	--	--	3.612	2.52
Peel(i) × Morning	-0.892	-2.15	--	--
Peel(i) × Evening	1.001	2.31	--	--
Oxford(j) × PM Peak	1.951	1.70	--	--
Services × Afternoon	--	--	1.081	2.88
Halton(i) × Wholesale	--	--	1.981	1.85
Brant(j) × Manufacturing	0.958	1.63	--	--
Wellington(j) × Construction	2.288	1.89	--	--
Oxford(j) × Construction	1.337	1.90	--	--
No. of Observations	776			
$LL(0)$	-284.59		-284.59	
$LL(\beta)$	-214.02		-223.72	
ρ^2	0.248		0.214	

choice probability is increased for stops located in Simcoe. By comparison, the impact of the variable on the choice probability is reduced for stops located in Middlesex. While the probability of selecting a stop increases for stops that have longer cumulative times and stops that are visited during a specific time of day. This tends to be more pronounced for stops visited in the morning, as depicted by the coefficient of the term (*CT* × *Morning*).

In this model, the enclosed angle (theta) between stops locations had no impact on the choice probability. Nonetheless, the spatial orientation of the chosen stops was further analyzed according to the Cartesian system, which basically consists of four Cartesian quadrants. It is found that stops associated with the home origin of the tour had a lower probability of being selected when that origin is Halton and the stops fall in the first Cartesian quadrant. Further, the impact of Halton, as an origin, on the choice probability of a stop increases for stops that are located in the second Cartesian quadrant relative to the origin location of the stop. However, the impact of having stops fall in the second Cartesian quadrant decreases the choice probability of choosing stops that take longer travel time to complete their duties. Also, stops falling in the second Cartesian quadrant see an increase in their choice probability when these stops are affiliated with the retail trade industry.

Throughout the entire day, stops visited during the AM Peak had a higher tendency of being chosen. This is not surprising since tours with a higher number of stops are more likely to start in the early morning, which is done to complete the tour and start another tour. However, the choice probability decreases for stops visited during the Morning when these stops are associated with the home origin of the tour and the origin is Peel. Interestingly, the opposite is observed for these Peel stops when visited in the Evening. This suggests that 2-stop tours from Peel are more likely to occur at later hours in the day.

The probability of selecting stops increases for stops located in certain metropolitan areas and which are associated with certain industries, as depicted by the coefficient of the term (*Brant(j)* × *Manufacturing*). Similarly, the probability for selecting a stop increases for stops associated with the Construction industry and for stops located in Oxford, as depicted by the coefficient of the term (*Oxford(j)* × *Construction*). The impact of the Construction stops tend to be even more pronounced for stops located in Wellington, as depicted by the coefficient of the term (*Wellington(j)* × *Construction*). Also, stops visited during the PM Peak have a higher chance of being selected if they are associated with the Construction industry, as depicted by the coefficient of the term (*Construction* × *PM Peak*).

Second Stop Destination Choice Model

Unlike the first stop model, the probability of choosing the second stop in the 2-stop tour case increases if the stop is classified as Primary, all other things being equal. The probability for selecting a stop decreases for stops that are far away from the previous stop, as discerned by the coefficient of the (*Travel Distance*) variable. However, this is not the case if the stop is located in Durham, as discerned by the coefficient of the (*Durham(j)*) variable. That being said, the probability of selecting a stop increases for stops located in Durham if these stops are associated with the retail trade industry, as depicted by the coefficient of the term (*Durham(j)* × *Retail Trade*), all things being equal.

Further, the probability of selecting a stop increases for distances ranging between 100 km and 150 km, as portrayed by the coefficient of the ($TD\ Between\ 100_{km}\ and\ 150_{km}$) variable. The probability of selecting a second stop increases for stops located in Toronto when the distance between the stop and the previous stop is large, as discerned from the parameter of the term (*TD* ×*Toronto(j)*). Also, the coefficient of the term (*Toronto(j)* × *Q1*) suggests that stops located in the Toronto region had a higher probability of being selected if these stops fall in the first Cartesian quadrant relative to the previous stop location, though the impact of having stops fall in the first Cartesian quadrant decreases the choice probability for stops located in Waterloo. With respect to the latter region, the choice probability increases for stops in Waterloo when these stops are visited during the Evening, as depicted by the coefficient of the term (*Waterloo(j)* × *Evening*).

Unlike the first stop model, the probability of selecting a second stop decreases for stops visited during the AM Peak. This is not surprising since the chosen stop is the last stop in the tour and as such is less likely to correspond to an AM Peak period. However, the opposite is observed for stops located in the home origin of the tour and when this origin is Toronto. One possible explanation could be that the tours associated with these stops are local (e.g. pickup/deliver done within the same metropolitan). Similar to the previous stop, stops visited during the PM Peak are more likely to be selected if they are associated with the construction industry, as depicted by the coefficient of the term (*Construction* × *PM Peak*). Also, stops visited during the Afternoon are less likely to be chosen if they are affiliated with the services industry, as depicted by the coefficient of the term (*Services* × *Afternoon*).

The probability of selecting the second stop decreases as the enclosed angle between the stop and the first stop increases. Also, the probability of choosing a stop increases for stops located in York when these stops fall in the first Cartesian quadrant relative to the location of the previous stop. However, stops located in York have a lower probability of being chosen if they are associated with service industries. By comparison, stops located in the home origin of the tour and which are associated with the wholesale industry have a higher chance of being selected when the origin is Halton, as depicted by the coefficient of the term (*Halton(i)* × *Wholesale*).

5.3.3 Three-Stop Tour Destination Choice Models

Table 5-5 presents the results of the three-stop tour destination choice models with all of the coefficients and their corresponding t-statistics. The presented results are for three separate MNL models: the first MNL model represents the behavior associated with the choice of the first stop, the second MNL model represents the behavior for selecting the second stop, and the third MNL model represents the behavior for selecting the third stop. Most of the estimated parameters are statistically significant, although a few of the interaction terms were marginally significant (i.e., under 90% statistical significance level). The three MNL models have a rho-squared (ρ^2) values of 0.174, 0.132 and 0.157, respectively.

First Stop Destination Choice Model

The probability of choosing a stop decreases when the travel distance between the stop and the tour origin (i.e., location where the truck started) increases. This result is sensible since the first stop in a 3-leg tour is more likely closer to the origin of the tour. The impact of travel distance becomes more pronounced in the case of stops visited in the PM Peak period, as depicted by the coefficient of the term (*TD* × *PM Peak*). Further, the probability of selecting a first stop increases for stops that are at a distance ranging between 150 km and 200 km from the origin, as discerned by the coefficient of the (*TD Between* 150_{km} *and* 200_{km}). Such parameter captured the long-haul nature of the modeled tours.

The probability of selecting a stop increases for stops associated with certain home origins and when visited during a specific time period. For instance, the positive coefficients of the terms (*Halton(i)* × *Morning)* and (*Hamilton(i)* × *Morning*) suggest that stops affiliated with Halton and Hamilton have a higher chance of being selected when visited during the morning period. By comparison, the opposite is observed for stops located in the Peel region, as illustrated by the coefficient of the term (*Peel(i)* × *Morning)*. As the results suggest, stops visited during the morning have a higher probability of being selected if they are associated with the transportation industry, as depicted by the coefficient of the terms (*Transportation* × *Morning)*. Also, stops located in Toronto and which are linked to the transportation industry have a higher chance of being selected by trucks originating from Toronto In the same context, stops located in Durham and visited during the PM Peak have a higher probability of being chosen by trucks originating from Durham. As for the attraction regions, stops located in Wellington have a lower chance of being chosen. However, this is not the case for Wellington-based stops that are visited during the Afternoon and the PM Peak periods, as depicted by the coefficient of the terms (*Wellington(j)* × *Afternoon) and (Wellington(j)* × *PM Peak)*, other things being equal. Similarly, stops visited during the PM Peak have a higher chance of being selected if they are located in Waterloo, as depicted by the coefficient of the term (*Waterloo(j)* × *PM Peak*).

The model was able to capture the direction of stops generated from the GTHA relative to their origins and the type of industry that they are associated with. For example, stops falling in the first Cartesian quadrant see an increase in their choice probability if they belong to the manufacturing industry, as depicted by the coefficient of the term (*Manufacturing* × *Q1*). Similarly, the probability for selecting a stop increases for stops associated with the wholesale industry in the case of stops falling in the second Cartesian quadrant, as depicted by the coefficient of the term (*Wholesale* × *Q2*). Also, stops falling in the third Cartesian quadrant have a higher probability of being chosen if they are associated with the Transportation industry, as depicted by the coefficient of the term (*Transportation* × *Q3*).

Second Stop Destination Choice Model

Unlike the first stop case, the probability for selecting the second stop increases for stops that are far away from the first stop location, as discerned by the positive coefficient of the (*Travel Distance*) parameter. The impact of travel distance is more pronounced for stops associated with the Services industry and for stops that are located in Middlesex. Stops located in their Toronto tour-home region have a higher chance of being selected if they are at a distance ranging between 200 km and 250 km from the previous stop, as discerned by the positive coefficient of the (*Toronto(i)* × *TD Between* 200_{km} *and* 250_{km}*)*. Unlike the first stop, the choice of the second stop decreases during the AM Peak period. However, the impact of the AM Peak is reversed if the second stop is associated with the wholesale industry, as depicted by the coefficient of the term (*Wholesale* × *AM- Peak*). Further, the probability of selecting a second stop tends to increase for stops visited during the Morning period, as discerned by the coefficient of the (*Morning*).

Stops located in Hamilton and which are associated with the manufacturing industry have a higher chance of being selected by trucks originating from Hamilton, as depicted by the coefficient of the term (*Hamilton(i)* × *Manufacturing*). The same could be said about these Hamilton stops if they are visited during the Afternoon period, as depicted by the coefficient of the term (*Hamilton(i)* × *Afternoon*). A similar trend is also observed in Halton as portrayed by the coefficient of the term (*Halton(i)* × *Afternoon*). Further, the probability of Halton-based trucks to choose stops from Halton decreases if these stops fall in the first Cartesian quadrant relative to the first stop, as depicted by the coefficient of the term (*Halton(i)* × *Q1*). By comparison, the positive coefficient of the term (*Hasting(j)* × *Q1*) suggests that the choice probability for stops increases if these stops are located in Hasting and fall in the first Cartesian quadrant, other things being equal.

Third Stop Destination Choice Model

The travel distance in this model has no impact on the choice probability on its own. However, the probability of selecting a final stop to be from York increases for trucks originating from York at a distance ranging between 100 km and 150 km from the previous stop, as depicted by the coefficient of the term (*York(i)* × *TD Between* 100_{km} *and* 150_{km}). Also, the probability of selecting a stop increases for stops located in Durham and which are at a further distance from the previous stop, as discerned by the coefficient of the term (*Durham(j)* × *TD*). In the same context, the probability of choosing a stop from Durham decreases if the stop falls in the first Cartesian quadrant relative to the previous stop, as illustrated by the coefficient of the term (*Durham(j)* × *Q1*). Furthermore, stops located in Wellington have a higher chance of being selected, other things being equal.

The probability of selecting a final stop decreases for stops visited during the AM Peak, Morning and Evening. However, the coefficient of the term (*Peel(i)* × *Morning)* suggests that the probability of choose a final stop from Peel by a truck originating from Peel will increase if the stop is visited during the morning. By comparison, the probability of choosing such a stop decreases if the stop is visited during the PM Peak. Furthermore, the probability for selecting a stop increases for stops associated with the Services industry and for stops located in Oxford. However, the choice probability decreases for Services stops located in the third quadrant Q3.

Table 5-5: MNL estimation results for 3-stop tours

Variable Name	1st stop model		2nd stop model		3rd stop model	
	Beta	t-stats	Beta	t-stats	Beta	t-stats
Travel Distance (TD)	-0.003	-2.41	0.007	6.12	--	--
TD × PM Peak	-0.008	-1.69	--	--	--	--
TD × Services	--	--	0.002	2.21	--	--
TD × Middlesex(j)	--	--	0.004	2.29	--	--
TD × Durham(j)	--	--	--	--	0.133	3.08
TD Between 150_{km} *and* 200_{km}	0.437	1.91	--	--	--	--
TD Over 400_{km}	--	--	-1.381	-1.66	--	--
TD Between 100_{km} *and* 150_{km} *× York(i)*	--	--	--	--	1.325	1.59
TD Between 200_{km} *and* 250_{km} *× Toronto(i)*		--	2.100	1.93	--	--
Wellington(j)	-2.064	-2.02	--	--	0.875	2.89
Services × Oxford(j)	--	--	--	--	1.320	2.21
Toronto(i)	--	--	-1.372	-1.80	--	--
Manufacturing × Hamilton(i)	--	--	1.352	1.43	--	--
Manufacturing × Q1	0.772	2.32	--	--	--	--
Wholesale × Q2	1.279	2.22	--	--	--	--
Services × Q3	--	--	--	--	-1.038	-1.95
Transportation × Toronto(i)	2.070	1.79	--	--	--	--
Transportation × Q3	0.546	1.78	--	--	--	--
Transportation × Night	--	--	0.470	1.54	--	--
Retail Trade × Q4	--	--	1.860	1.57	--	--
Halton(i) × Q1	--	--	-1.550	-1.88	--	--
Toronto(i) × Q3	--	--	--	--	2.102	3.04
Hasting(j) × Q1	--	--	0.775	1.92	--	--
Durham(j) × Q1	--	--	--	--	-1.275	-1.75

Table 5-5 Continued

Variable Name	1st stop model		2nd stop model		3rd stop model	
	Beta	t-stats	Beta	t-stats	Beta	t-stats
Oxford(j) × Q2	--	--	--	--	1.005	1.44
AM Peak	1.720	6.30	-0.937	-1.96	-2.275	-4.23
Wholesale × AM Peak	--	--	1.716	2.51	--	--
Morning	--	--	0.875	3.64	-2.800	-2.67
Peel(i) × Morning	-1.614	-2.02	--	--	2.105	1.92
Halton(i) × Morning	0.875	1.59	--	--	--	--
Hamilton(i) × Morning	1.610	2.01	--	--	--	--
Transportation × Morning	1.259	2.19	--	--	--	--
Public Administration × Morning	--	--	--	--	1.251	1.48
Afternoon	-1.222	-1.95	--	--	--	--
Wellington(j) × Afternoon	3.829	2.44	--	--	--	--
Hamilton(i) × Afternoon	--	--	1.550	1.64	--	--
Halton(i) × Afternoon	--	--	1.431	2.62	--	--
PM Peak	1.481	1.97	--	--	--	--
Q2 × PM Peak	1.394	2.03	--	--	--	--
Durham(i) × PM Peak	1.465	2.13	--	--	--	--
Peel(i) × PM Peak	--	--	--	--	-1.545	-2.53
Waterloo(j) × PM Peak	1.820	1.99	--	--	--	--
Wellington(j) × PM Peak	2.937	1.92	--	--	--	--
Evening	1.471	5.08	--	--	-2.537	-3.51
No. of Observations			128			
LL(0)	-377		-365.22		-374.05	
LL(β)	-311.45		-316.84		-315.23	
ρ^2	0.174		0.132		0.157	

5.4 Modeling Tour Stop Duration

5.4.1 Cox Proportional Hazard Duration Model

The Cox Proportional Hazard (CPH) model has been estimated twice; the first CPH model studied the primary stop duration, while the second CPH model studied the secondary stop duration. The following two sections namely; Primary Stop Duration Model, and Secondary Stop Duration Model, will discuss and present the results with all of the coefficients and their corresponding t-statistics[1].

Primary Stop Duration Model

A CPH model is estimated to predict the duration of primary stops of trucks. The results are organized and summarized into six groups as shown in Table 5-6. For brevity, the following section will highlight only key parameters form each group. A positive coefficient suggests an increase in the stop duration while a negative coefficient indicates a decrease in the stop duration.

The tour class plays a vital role in governing the stop duration. The model suggests that, an increase in the number of stops for a given tour is positively correlated with the time spent for a visited stop during the tour. Although the effect of the increase in number of stops should be the negatively correlated with the stop duration, the majority of the modeled tours in this study were 1-stop tours. As such, the duration of the stop in these

[1] It is worth noting that besides the CPH models, three Kaplan-Meier curves were plotted to show the predicted survival probability for three different group variables that were used in the specification of the CPH models. These are: Tour class, Origins, and Destination of the stops. The generated curves, shown in Appendix A, illustrates the distinction between the model and the group variables.

tours would usually take longer than other type of tours (e.g., 4-stop tour). The parameter of travel time is significant and indicates longer travel time is positively correlated with the stop duration.

In terms of the geographic location of the stops, the model suggests that the duration decreases for stops located with their tour-home regions in the case of Peel, York, or Hamilton. The impact is most pronounced in the case of Hamilton. In addition, stops located at non-tour-home regions like Durham, Halton, Peterborough also have shorter truck stop durations. However, the opposite is observed for further destinations from the GTHA such as Chatham-Kent, Wellington and Lambton.

With respect to the starting time of the tour and the tour-home region of stops, the model indicates that the duration tends to decrease for stops from Halton during the PM Peak. By comparison, the duration tends to increase the most for stops located in Toronto for tours generated during the Evening period. With respect to the starting time and the destination (i.e., none-tour-home region) of stops, the model suggests shorter durations for stops from Ottawa Wellington, Lambton and Perth for tours generated during the morning period. By comparison, longest durations are more likely to occur for stops from Halton and Toronto for tours generated during the Evening period.

With respect to the industry associated with the tours and the destination of the stops, the results suggest that the longest duration will occur for stops located in Hastings and tied to the public administration industry and for stops located in Lambton and tied to the Finance Industry. By comparison, the shortest durations are more likely to occur for stops located in Hamilton and associated with wholesale Industry, in Waterloo and associated with the public administration industry.

Table 5-6: Cox Proportional Hazard estimation results for primary stop duration

Variable Name	Beta	t-stats	Hazard Ratio
Tour Class	0.196	5.49	1.22
Travel Time	0.001	1.85	1.00
Origin			
Peel	-0.226	-3.90	0.80
York	-0.346	-3.98	0.71
Hamilton	-0.910	-6.89	0.40
Destination			
Durham	-0.125	-1.77	0.88
Halton	-0.340	-2.41	0.71
Peterborough	-0.439	-3.37	0.64
Chatham-Kent	0.598	3.62	1.82
Middlesex	0.247	2.85	1.28
Wellington	0.595	4.40	1.81
Lambton	0.603	3.31	1.83
Oxford	0.252	2.84	1.29
Time of day Vs. Origin			
AM Peak × Toronto	0.783	2.19	2.19
AM Peak × Hamilton	0.766	3.28	2.15
Morning × Hamilton	0.357	1.96	1.43
PM Peak × Halton	-0.516	-2.08	0.60
Evening × Toronto	0.943	2.45	2.57
Night × York	0.476	2.08	1.61
Time of day Vs. Destination			
AM Peak × Middlesex	-0.393	-1.95	0.68
AM Peak × Peterborough	0.751	2.10	2.12
Morning × Ottawa	-1.364	-1.92	0.26
Morning × Wellington	-0.659	-2.32	0.52
Morning × Waterloo	0.269	2.35	1.31
Morning × Lambton	-0.711	-1.95	0.49
Morning × Perth	-0.548	-2.37	0.58
Evening × Greater Sudbury	0.678	1.78	1.97
Evening × Toronto	1.947	3.35	7.01
Evening × Halton	2.967	2.93	19.43
Night × Essex	0.639	2.81	1.89

Table 5-6 Continued

Variable Name	Beta	t-stats	Hazard Ratio
Time of Day Vs. Industry			
Afternoon × Construction	-0.556	-2.54	0.57
Afternoon × Services	0.256	2.61	1.29
Evening × Manufacturing	-0.446	-1.92	0.64
Industry Vs. Destination			
Construction × Oxford	-0.455	-1.81	0.63
Construction × Perth	1.327	1.81	3.77
Construction × Grey	1.858	1.85	6.41
Construction × Frontenac	1.649	1.65	5.20
Wholesale × Ottawa	1.973	1.97	7.19
Wholesale × Hamilton	-1.030	-2.53	0.36
Wholesale × Chatham-Kent	1.202	2.00	3.33
Wholesale × Essex	0.458	1.62	1.58
Retail × Essex	-0.716	-2.51	0.49
Finance × Essex	1.629	2.81	5.10
Finance × Lambton	2.945	2.79	19.00
Services × Waterloo	-0.246	-1.95	0.78
Public Admin × Middlesex	1.547	3.04	4.70
Public Admin × Hastings	2.226	3.13	9.26
Public Admin × Waterloo	-0.822	-2.15	0.44
No. of Observations		4,406	
Concordance		0.605	
logrank test		367.3	
Degrees of freedom		48	
ρ^2		0.066	

Secondary Stop Duration Model

A CPH model is estimated to predict the duration of truck stops at secondary stop locations. The results are summarized in Table 5-7. The developed model achieved a rho-squared (ρ^2) value of 0.129.

As the name suggests, secondary stops are usually short in natural, unlike the primary stops. Therefore, the expectations of the results from a secondary stop model are more likely to be negative. Like the primary stop duration model, the parameter associated with travel time is still significant and indicates that longer travel time is positively correlated with the stop duration. This makes sense because truck drivers are more likely to have a break whether for refueling or taking a rest due after spending long time on the road. The impact of travel distance is only visible on tours that are generated during the AM Peak. This impact is positively correlated with the stop duration, as depicted by the coefficient of the term (*Travel Distance × AM Peak*).

Among the six origins comprising the GTHA, the model suggests that stops located in their home-tour regions of Halton or Toronto tend to have a negative impact on the stop duration. However, the shortest duration is associated with the Toronto stops. In addition, stops that are destined to certain none-home-tour regions (namely, Leeds, Peel, Toronto, Hastings, Waterloo, Essex, or Parry Sound) are more likely to have shorter stop durations. Again, stops that located in Toronto have the shortest stop duration.

Table 5-7: Cox Proportional Hazard estimation results for secondary stops

Variable Name	Beta	t-stats	Hazard Ratio
Travel Time	0.001	2.06	1.00
Travel Distance × AM Peak	0.002	1.85	1.00
Origin			
Halton	-0.603	-2.17	0.55
Toronto	-1.047	-4.51	0.35
Destination			
Leeds	-0.544	-1.54	0.58
Peel	-2.102	-2.79	0.12
Toronto	-1.667	-4.33	0.19
Hastings	-0.727	-3.46	0.48
Waterloo	-0.814	-2.73	0.44
Essex	-0.740	-2.29	0.48
Parry Sound	-0.727	-1.57	0.48
Industry			
Manufacturing	-0.432	-2.54	0.65
Transportation	-0.208	-1.67	0.81
No. of Observations		548	
Concordance		0.633	
Logrank test		68.8	
Degree of freedom		13	
ρ^2		0.129	

CHAPTER 6: CONCLUSION

6.1 Summary of Empirical Results

The broad purpose of this thesis was to analyze and model two aspects of commercial vehicle movements: 1) truck stop destination, and 2) truck stop duration at the micro-level. Also, these models were designed to be implemented as a part of a micro-simulation tour-based model. The work is based on truck tours that were created from a large GPS dataset that depicts the movement of Canadian trucks during the month of March 2016. The information was further analyzed to identify the tours originitaing from the Greater Toronto and Hamilton Area (GTHA) in Ontario to consider in the modeling exercise. To date, the majority of freight demand models are either too aggregate in nature or still too immature or under developed. The current study adopts the proposition that the analysis of truck tours within a micro-simulation framework can provide a more realistic picture of the true process governing the movement of trucks. To this end, econometric models were applied to respectively investigate the factors that affect stop destination and stop duration for a given tour. Therefore, identifying the factors that explain truck stop destination choice behavior and stop duration will help devise more effective travel demand models. Furthermore, the availability of detailed micro-data in this research will facilitate the development of models that are behavioral and policy sensitive to assist in informing future transportation plans. The work conducted in this thesis has not been done in the past; therefore, the study’s results offer a novel contribution to the transportation engineering literature. The results also contribute to the development of an operational integrated micro-simulation model of freight movement.

6.1.1 Tour Stop Destination Models

Location choice models are used to explain the destination of truck tour stops at the micro-level. Three classes of tours were considered in these models: One-Stop Tour, Two-Stop Tour, and Three-Stop Tour. A total of six Multinomial Logit (MNL) model were developed for the three classes of tours: one MNL model for One-Stop Tour, two MNL models for Two-Stop Tour, and three MNL models for Three-Stop Tour. The purpose of these models was to test the influence of various characteristics of stops on the location choice sets. The models made use of the following stop's characteristics: stop purpose, travel time and travel distance between the target stop and a previous stop, type of industry associated with the stop, the origin and destination geography of the stop, stop direction with respect to the previous stop, and time of day when the stop was visited.

The results offered promising insights about the truck destination choice when interaction terms are introduced. Table 6-1 provides a summary of the McFadden's Rho-square of the estimated MNL models, while Table 6-2 highlights the common variables among the six MNL models.

Table 6-1: McFadden's Rho-squared Values for the MNL Models

Tour class	McFadden's Rho-squared		
	1^{st} stop	2^{nd} stop	3^{rd} stop
1-stop tour	0.077	--	--
2-stop tour	0.248	0.214	--
3-stop tour	0.174	0.132	0.157

Table 6-2: Common Variables of the MNL Models

k-stop tour model	**1-stop tour**	**2-stop tour**		**3-stop tour**		
Variable Name	1^{st} stop model	1^{st} stop model	2^{nd} stop model	1^{st} stop model	2^{nd} stop model	3^{rd} stop model
Primary Stop	↑ +ve	--	↑ +ve	--	--	--
Travel Distance	↑ +ve	↓ -ve	↓ -ve	↓ -ve	↑ +ve	--
Travel Distance × *PM Peak*	↑ +ve	--	--	↓ -ve*	--	--
Cumulative Time	↑ +ve	↑ +ve	--	--	--	--
Cumulative Time × *Morning*	↑ +ve	↑ +ve	--	--	--	--
Enclosed Angle	↑ +ve	--	↓ -ve	--	--	--
AM Peak	--	↑ +ve	↓ -ve	↑ +ve	↓ -ve	↓ -ve
Peel(i) × *Morning*	--	↓ -ve	--	↓ -ve	--	↑ +ve*
Hamilton(i) × *Morning*	↑ +ve	--	--	↑ +ve	--	--

*Less than 95% significant

Overall, the results of the MNL models provided interesting insights about the choice behavior of the stop destinations chosen by trucks as part of the tours. To avoid any confusion in reading the information in Table 6-2, a positive coefficient suggests an increase in the probability of choosing a stop with respect to a certain parameter, while a negative coefficient indicates a decrease in the probability of choosing a stop with respect to a certain parameter. In general, the first stop is relatively consistent among the three classes of tour (i.e. 1-stop, 2-stop, and 3-stop tours). However, this consistency is gradually disappearing when it comes to the second stop for the case of 2-stop tours and 3-stop tours.

The primary stop dummy variable increases the probability of choosing a stop as a destination in the case of 1^{st} stop model and 2^{nd} stop model for the 1-stop tour and 2-stop tour, respectively. As for the travel distance, stops that are far away from the tour origin increases the probability of choosing a stop in the case of 1^{st} stop for the 1-stop tour, while it decreases the probability of choosing a stop in the case of 1^{st} stop for the 2-stop tour.

Also, the probability of selecting a stop decreases for stops that are far away from the previous stop as captured in the 2^{nd} stop model for the 2-stop tour. However, the probability of selecting a stop increases for stops that are far away from the previous stop as captured in the 2^{nd} stop model for the 3-stop tour. Furthermore, the probability of choosing a stop increases when the cumulative time between the stop and the tour origin increases in the case of 1^{st} stop model for the 1-stop tour and 2-stop tour. However, the cumulative time has no effect when a stop is visited in the morning period.

A wider enclosed angle (theta) increases the probability of choosing a stop relative to the origin in the case of 1^{st} stop model for the 1-stop tour, while it decreases the probability of choosing a stop relative to the previous stop in the case of 2^{nd} stop model for the 2-stop tour. However, theta has no impact on the choice set of the 3-stop tour. The AM Peak dummy variable was found to have a high positive significant impact on the 2-stop and 3-stop tours. Here, the probability of choosing stops that are visited during the AM Peak increases for the case of 1^{st} stop of the 2-stop tour and 3-stop tour. Further, the probability of choosing stops that are visited during the AM Peak decreases in the remaining stop models of the 2-stop tour and 3-stop tour. In the case of the 1^{st} stop of the 1-stop tour and 3-stop tour, the probability of selecting a stop located in Hamilton by a truck originating from Hamilton increases when the stop is visited during the Morning. On the other hand, in the case of 1^{st} stop of the 2-stop tour and 3-stop tour, the probability of selecting a stop located in Peel by a truck originating from Peel decreases when such a stop is visited during the Morning.

6.1.2 Tour Stop Duration Models

Different survival models were developed to study the factors affecting the duration of truck stopped event when tours take place. More specifically, the models included: Cox proportional hazard (CPH) model, and Kaplan-Meier model. Here, two CPH models were estimated to study the duration of stopped events at primary and secondary stop locations, respectively. The CPH models will be used as part of the micro-simulation tour-based model. On the other hand, a Kaplan-Meier model was developed for the primary stops to explore how the survival probability curve of the Cox proportional hazard model compares with non-parametric survival curves. Unlike the CPH models, the Kaplan-Meier model will not be part of the tour-base model as it was developed for illustration purposes. Overall, the results of the CPH models provided interesting insights about the tour stop duration by considering the influence of various characteristics of stops. The models made use of the following stop's characteristics: stop purpose, travel time and travel distance to the stop, type of industry associated with the stop, the origin and destination location of the stop, stop direction, and time of day when the stop was visited. Promising insights about the truck stop duration were obtained when interaction terms were introduced in the model. As for, **some of the key findings from the primary stop duration model include:**

- An increase in the number of stops – up to four stops– for a given tour is positively correlated with the time spent for a visited stop during the tour, which translates in a longer stop duration.
- An increase in the travel time is positively correlated with the stop duration, which translates in a longer stop duration.

- In general, the stop duration decreases for stops located within their tour-home regions.
- Stops located at none-tour-home regions but still close to the tour-home-regions tend to have shorter stop durations. However, stops located in none-tour-home regions that are further away from the tour-home-regions tend to have longer stop durations.
- Time of day and stop industry variables in the model had no impact on stop duration when specified on their own. However, interacting these variables with some factors improved the model.

Some of the key findings in the secondary stop duration model include:

- Longer travel time is positively correlated with the stop duration, which translates in a longer stop duration.
- Stops associated with longer travel distance and which are visited during the AM Peak tend to have longer stop duration.

6.2 Contributions and Policy Implications

The analysis conducted in this research offers an innovative effort to form the basis for developing a more comprehensive understanding of freight movement processes. The current thesis makes two key contributions: 1) it advances the current state of knowledge on freight demand modeling, and 2) it applies advance geo-spatial methods and statistical techniques to model the commercial vehicles movements. The developed models will allow planners to predict destinations and stops duration of commercial vehicles and in turn create schedules that can reduce stop duration and increase the efficiency of each tour, thereby saving time and money. Given the fact that, the geo-spatial characteristics of a large sample of truck tours from the Greater Toronto Hamilton Area (GTHA) is highlighted for the first time, the statistical analysis forms the basis for a novel tour-based microsimulation framework that will be built with data derived from passive truck GPS data.

From a transport policy presprective, modeling and understanding commercial vehicles movements is essential for both urban and regional transportation planning since commercial vehicles influence traffic and level-of-service on the transportation network. The results from this research can assist planners and decision-makers as it can help them predict destinations and stops duration of commercial vehicles in the province of Ontario in future freight transportation plans at the micro-level. However, failing to plan for freight activities in transportation plans could hamper the performance of transportation networks, resulting in distress in freight movement and economic performance. In short, the research conducted in this thesis addresses some of the drawbacks in existing freight demand forecasting models and offers a platform for performing better predictions using data derived from truck GPS data.

6.3 Limitations and Recommendations for Future Research

Finally, as mentioned earlier, the innovative models presented here use truck GPS data to address an important gap in current research in terms of the factors that could help researchers better understand the destination choice of truck stops and the duration of truck stopped events at such stops. Limitations of this empirical analysis can be attributed to the passive nature of the truck GPS dataset used in this research. This is particularly true because the records forming the GPS data were not originally intended as an input for transportation models. Accordingly, analyzing these records was very time consuming.

Also, the analysis would have been enriched if further information about the types of commodities carried by the modeled trucks was available. The size of the original dataset was massive as it covered the movement of trucks across all North America. Even the records pertaining to the province of Ontario were too large to be included in the models developed in this thesis. Consequently, the analysis had to be confined to the tours generated from the Greater Toronto and Hamilton Area (GTHA). Although the developed models can help improve the predictive abilities of freight demand models, the analysis did not explore the influence of accessibility and employment on the destination choice for stops. Further, the tours were filtered to include only tours associated with a 13-hours of driving time. Finally, the models did not include any information about the size of the analyzed trucks.

Future research that can help address the limitations listed above should focus on incorporating accessibility and employment variables in the model specification. Predicting and incorporating the size of the truck and the type of commodities shipped would be also an area of future research. Expanding the study area to cover a larger geography would also be beneficial. Also, considering non return to home tours is important sine these tours have not be modeled in the past. Another aspect to consider in future research is to study tours that cross the Canada-US international border. Future research could also utilize the advance discrete modeling techniques such as Mixed logit (MXL) model to improve upon the MNL approach used in this thesis. Finally, since the work conducted in this thesis contributes to the development of an agent-based micro-simulation truck movement model, future work should focus on implementing the developed destination and duration truck tour stop models to perform predictions and examine a variety of scenarios to promote policies for minimizing traffic congestion caused by commerical vehicles.

REFERENCES

Abraham, J.E., Hunt, J.D. (1999)' Semi-automated calibration of the MEPLAN model of Sacramento', *In: 6th Computers in Urban Planning and Urban Management Conference*, Venice, Italy.

Anon (2013). *City logistics research: a transatlantic perspective; summary of the first EU-U.S. transportation research symposium*. Conference proceedings / Transportation Research Board 50. Washington, DC: Transportation Research Board.

Chow, J. Y. J., Yang, C. H. and Regan, A. C. (2010) 'learned and the road ahead', pp. 1011–1030. doi: 10.1007/s11116-010-9281-1.

Disserta, R. A. (2011) 'The adequacy of Freight modelling methods to study policy / regulations ' implementation impacts in a City Logistics context . Application to the city of Lisbon , Portugal.

Doustmohammadi, E. *et al.* (2016a) 'Comparison of Freight Demand Forecasting Models', *International Journal of Traffic and Transportation Engineering*, 5(1), pp. 19–26. doi: 10.5923/j.ijtte.20160501.03.

Doustmohammadi, E. *et al.* (2016b) 'Modeling Freight Truck Trips in Birmingham Using Tour-Based Approach', *Journal of Transportation Technologies*, 06(05), pp. 436–448. doi: 10.4236/jtts.2016.65035.

Ferguson, M. *et al.* (2012) 'Transferability and enhancement of a microsimulation model for estimating urban commercial vehicle movements', *Journal of Transport Geography*. Elsevier Ltd, 24, pp. 358–369. doi: 10.1016/j.jtrangeo.2012.04.013.

Fischer, M. J. *et al.* (2005) 'Innovative Framework for Modeling Freight Transportation in Los Angeles County , California', (1906), pp. 105–112.

Gingerich, K. (2017) 'Studying Regional and Cross Border Freight Movement Activities with Truck GPS Big Data', *ProQuest Dissertations and Theses*, p. 251. Available at: https://search.proquest.com/docview/1987989688?accountid=27233%0Ahttp://technion-primo.hosted.exlibrisgroup.com/openurl/972TEC/972TEC_SP? [Last Accessed: March 17, 2020]

Gingerich, K., Maoh, H. and Anderson, B. (2015) 'Modelling the Determinants of Truck Tours Within Canadian Markets', (Figure 1), pp. 1–15.

Gingerich, K., Maoh, H. and Anderson, W. (2014) 'Classifying the purpose of stopped truck events: An application of entropy to GPS data', *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd, 64, pp. 17–27. doi: 10.1016/j.trc.2016.01.002.

Greaves, S. P. and Figliozzi, M. A. (2008) ‘Commercial vehicle tour data collection using passive GPS technology: issues and potential applications’, *TRB 87th Annual Meeting Compendium of Papers*, Washington D.C., pp. 1–18. Available at: https://ses.library.usyd.edu.au/handle/2123/19575.

Gliebe, J., Cohen, O. and Hunt, J. (2007) ‘Dynamic Choice Model of Urban Commercial Activity Patterns of Vehicles and People’, *Transportation Research Record: Journal of the Transportation Research Board*, 2003, pp. 17–26. doi: 10.3141/2003-03.

Huang, T. and Kockelman, K. (2007) ‘The Introduction of Dynamic Features in a Random-Utility-Based Multiregional Input-Output Model of Trade , Production , and Location Choice The Introduction of Dynamic Features in a Random-Utility-Based Multiregional Input-Output Model of Trade , Producti’, (January). doi: 10.5399/osu/jtrf.47.1.1061.

Hunt, J., Stefan, K., and Brownlee, A. (2006) ‘Establishment-Based Survey of Urban Commercial Vehicle Movements in Alberta, Canada: Survey Design, Implementation, and Results’, *Transportation Research Record: Journal of Transportation Research Board,* 1957 (1): 75-83.

Hunt, J. D. and Stefan, K. J. (2007) ‘Tour-based microsimulation of urban commercial movements’, *Transportation Research Part B: Methodological,* 41(9), pp. 981–1013. doi: 10.1016/j.trb.2007.04.009.

Jansuwan, S., Ryu, S. and Chen, A. (2017) ‘A two-stage approach for estimating a statewide truck trip table’, *Transportation Research Part A: Policy and Practice*. Elsevier Ltd, 102, pp. 274–292. doi: 10.1016/j.tra.2016.09.013.

Joubert, J. W. and Axhausen, K. W. (2009) ‘Inferring commercial vehicle activities in Gauteng , South Africa’, *Journal of Transport Geography*. Elsevier Ltd, 19(1), pp. 115–124. doi: 10.1016/j.jtrangeo.2009.11.005.

Joubert, J. W., Fourie, P. J. and Axhausen, K. W. (2010) ‘Traffic Simulation of Private Cars and Commercial Vehicles’, pp. 24–32. doi: 10.3141/2168-04.

Katchova, A. (2013). Survival Analysis. Available From: https://sites.google.com/site/econometricsacademy/econometrics-models/survival-analysis [Last Accessed: March 17, 2020]

Kim, H. and Park, D. (2017) ‘Empirical comparison of tour- and trip-based truck travel demand models’, *KSCE Journal of Civil Engineering*, 21(7), pp. 2868–2878. doi: 10.1007/s12205-017-0868-3.

Kuppam, A. R. *et al.* (2014) ‘Development of Tour-Based Truck Travel Demand Model Using Truck GPS Data’, *Transportation Research Board, 93rd Annual meeting,* (602), p. 27.

Maoh, H., Kanaroglou, P. and Woudsma, C. (2008) 'Simulation Model for Assessing the Impact of Climate Change on Transportation and the Economy in Canada', pp. 84–92. doi: 10.3141/2067-10.

Mladenovic, M. N. and Trifunovic, A. (2014) 'The Shortcomings of the Conventional Four Step Travel Demand Forecasting Process', *Journal of Road and Traffic Engineering*, 60(1), pp. 5–12.

Moore, A. M. (2017) '*Compendium of Tour-Based Freight Modeling Literature', Available at:* https://info.ornl.gov/sites/publications/Files/Pub104090.pdf [Last Accessed: April 17, 2020].

Oakes, D. (1977) 'The asymptotic information in censored survival data', *Biometrika*, 64(3), pp. 441–448. doi: 10.1093/biomet/64.3.441.

Outwater, M. *et al.* (2013) 'Tour based and supply chain modeling for freight: Integrated model demonstration in Chicago', *Transportation Letters*, 5(2), pp. 55–66. doi: 10.1179/1942786713Z.0000000009.

Samimi, A., Mohammadian, A. and Kawamura, K. (2010) 'A behavioral freight movement microsimulation model: method and data', *Transportation Letters*, 2(1), pp. 53–62. doi: 10.3328/TL.2010.02.01.53-62.

Sharman, B. W., Roorda, M. J. and Habib, K. M. N. (2012) 'Comparison of parametric and nonparametric hazard models for stop durations on urban tours with commercial vehicles', *Transportation Research Record*, (2269), pp. 117–126. doi: 10.3141/2269-14.

Stefan, K., McMillian, D., and Hunt, J. (2005) 'Urban Commercial Vehicle Movement Model for Calgary, Alberta, Canada'. *Transportation Research Record: Journal of Transportation Research Board*, 1921 (1): 1-10.

Tavasszy, L. A. (2006) 'Freight Transportation Modelling : international experiences', (January 2007).

Wang, Q. and Holguin-Versas, J. (2008) 'Tour-Based Entropy Maximization Formulations of Urban Commercial Vehicle Movements', *Association for European Transport and Contributors*, pp. 1–5.

You, S. I. and Ritchie, S. G. (2017) 'A GPS Data Processing Framework for Analysis of Drayage Truck Tours', *KSCE Journal of Civil Engineering*, 22(4), pp. 1454–1465. doi: 10.1007/s12205-017-0160-6.

Yang, C., Regan, A. C. and Son, Y. T. (2010) 'Another View of Freight Forecasting Modeling Trends', 14(2010), pp. 237–242. doi: 10.1007/s12205-010-0237-y.

National Academies of Sciences, Engineering, and Medicine. (2008). *NCHRP REPORT 606: Forecasting Statewide Freight Toolkit*. Washington, DC: The National Academies Press.

Wheatley-Price P, Hutton B, Clemons M. (2012) 'The Mayan Doomsday's effect on survival outcomes in clinical trials'. CMAJ. 2012 Dec 11; 184(18): 2021–2022. doi: 10.1503/cmaj.121616.

APPENDICES

Appendix A: Kaplan-Meier non-parametric survival probability

Figures 1-A through 2-A presents the Kaplan-Meier (KM) non-parametric survival probabilities for each group of the variables used to specify the CPH model. As shown in Figure A-1, a total of four curves are plotted for the modeled tour classes. The three curves labeled as (1), (2) and (3) represents the KM survival probabilities for the 1-stop tour, 2-stop tour and 3-stop tour, respectively. The curve labeled as (M) pertains to the developed model. According to the figure, the probability of a stopped event lasting for 75 minutes for stops associated with 2- and 3-stop tours is around 0.5. By comparison, the same probability is associated with a stop event lasting for 82.5 minutes for stops associated with 1-stop tours.

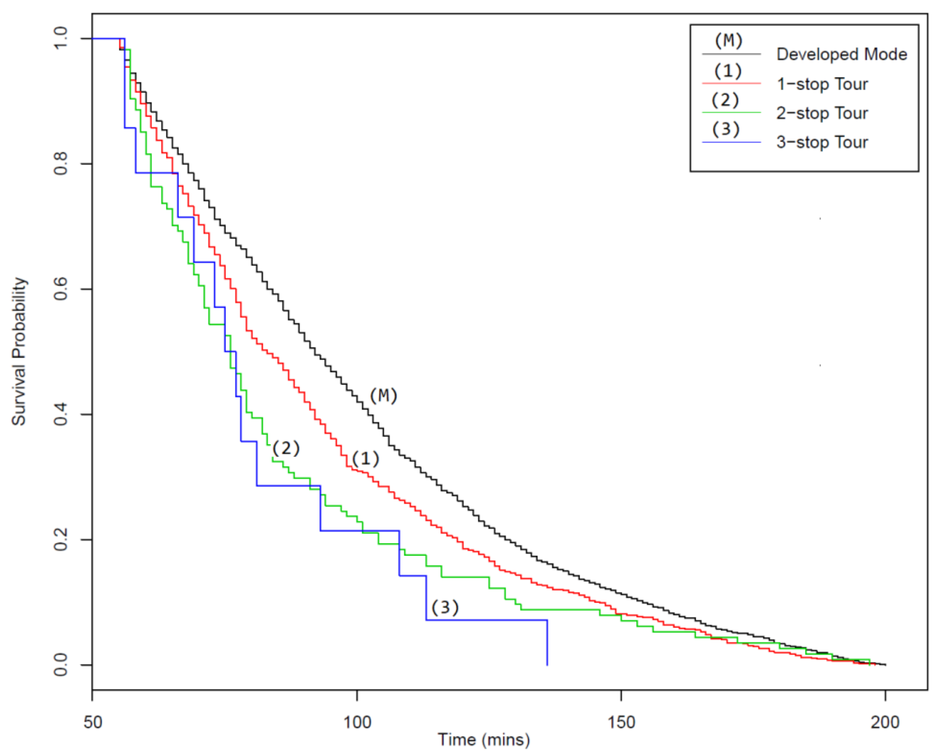


Figure A-1: Developed model vs. k-stop tour model

Figure A-2 provides a comparison between the developed model (M) and the significant origins of the stops: (1) Hamilton, (2) York and (3) Peel. The KM curves pertaining to York and Peel are quite similar to each other and to the developed model. However, the variation in the survival probability curves become more apparent at around 100 minutes and ends at 170 minutes. The probability of a stopped event lasting for around 100 minutes is approximately 0.5 for stops located in York. The time of a stopped event under the same probability for Peel is around 90 minutes. By comparison, the probability of a stopped event lasting for 100 minutes for stops in Hamilton is approximately 0.6.

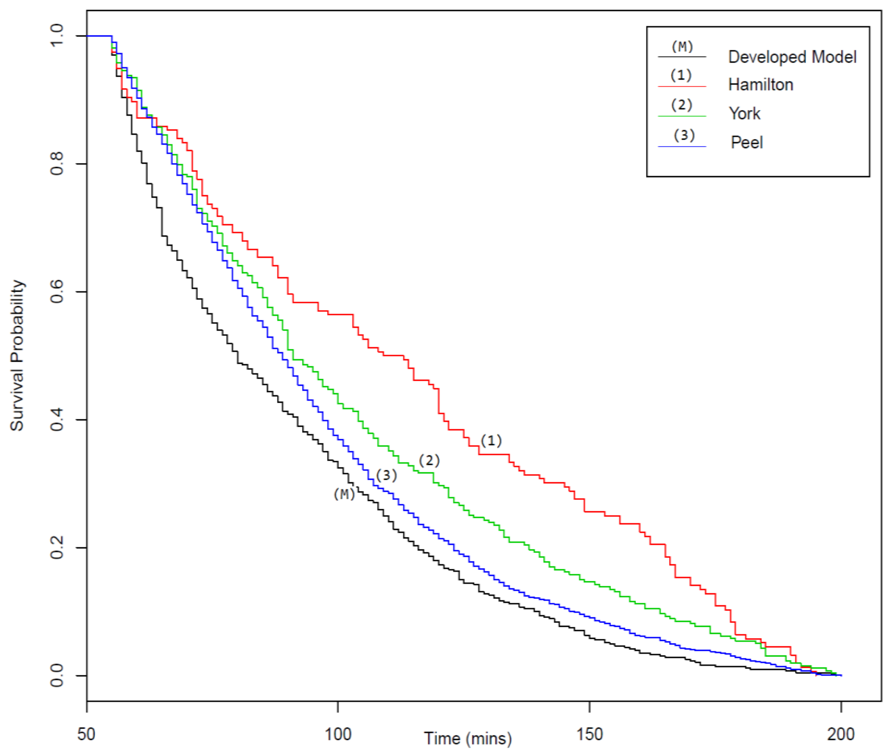


Figure A-2: Developed model vs. Origin model

Figure A-3 presents the KM survival curves based on the destination regions. As can be seen, there is a wider range with respect to the time of a stopped event to end at a destination stop. For instance, the probability for a stopped event to last for 75 minutes is around 0.5 for stops in Lambton. The same time is associated with a much lower probability of 0.4 for stops in Chatham-Kent. On the other hand, the probability of stopped event to last for 75 minutes is remarkably higher for stops in Durham and Peterborough.

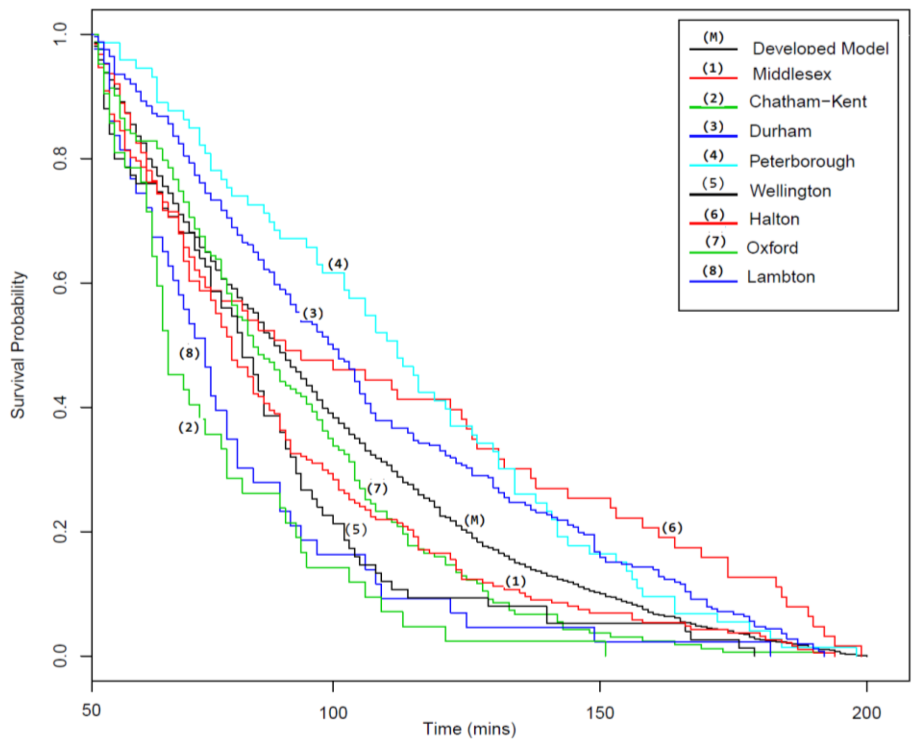


Figure A-3: Developed model vs. Destination

VITA AUCTORIS

NAME:	Ahmed Alshurafa
PLACE OF BIRTH:	Gaza, State of Palestine
YEAR OF BIRTH:	1993
EDUCATION:	University of Windsor, B.Eng., Windsor, ON, 2018
	University of Windsor, M.A.Sc.,Windsor, ON, 2020