Development of a Mode Choice Model to understand the potential impact of LRT on Mode Shares in the Region of Waterloo

by

Gursimran K. Saini

A thesis
presented to the University of Waterloo
in the fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Planning

Waterloo, Ontario, Canada, 2019

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that this thesis may be made electronically available to the public

Abstract

A new light rail transit system (LRT), ION, began operations in the Region of Waterloo in the June of 2019, and the second phase is yet to begin construction. The main thrust of this growth management project for the region was achieving sustainability goals by promoting denser development and boosting transit ridership. The LRT is integrated within the existing transit system, and this study intends to understand its impact on transit mode shares.

To understand the potential impact of introduction of a new transit system on mode shares, an analytical modelling approach is required. This research conducts spatial analysis in ARCMap to describe the current commuting patterns in the region of Waterloo, highlighting the spatial distribution of modal shares, top trip origins and destinations and trip distribution patterns by different modes. Furthermore, many nested logit and multinomial logit models were estimated to relate various socio economic, spatial and trip attributes to mode choice behaviour. The nested logit models did not prove to be a good fit for the available data, and the best multinomial logit model was finally used to understand the potential impacts of LRT.

The final model estimation projects 0.09% increase in commuter transit ridership for the estimated average decrease of 0.14% in travel time, which is a result of both, introduction of ION and realignment of transit routes. It is however, essential to note that ION is a driver of urban growth and development with potential to attract denser and mixed land use developments, which are both key to increasing transit ridership. This study is a start towards understanding the impacts of LRT and findings may prove to be a valuable resource to discerning spatial distribution of commuter trips in the region. Additionally, the model serves as flexible template, which can be employed to assess the impacts of LRT in the future and inform transit policy decisions.

Acknowledgement

I would like to begin by thanking my supervisor, Dr. Clarence Woudsma, for all the time and resources he provided me with during the completion of my graduate degree and thesis. I am forever grateful for his rich insights, patience, motivation and support which kept me going during this process. I am also thankful to him for giving me unique opportunities to work on projects which helped me hone my GIS and analytical skills.

My committee member, Dr. Mark Ferguson, for his guidance and technical input, especially in the modelling phase of this project. His comments and feedback continuously challenged my understanding of regression analysis and pushed me to explore the topic more deeply.

Scott MacFarlane, our GIS expert, who devoted valuable time to assist me in the data procurement process. His passion for GIS was infectious, and the lessons and skills I learnt from him during this process are indispensable.

My parents, who have immense faith in my abilities and gave me the courage to pursue my dreams in a new country. I can never thank them enough for the support, love and sacrifices they have made so I could channel all my energy in relentless pursuit of my dreams without worrying about anything else.

Dr. Kevin Curtis, who was the instructor for one of the first courses I took in this graduate program, for his passion to teach and invoke critical thinking in his students. He provided the feedback, support and encouragement which was essential for me during my first few months in this degree.

My friends, who were always available for brainstorming and insightful discussions through different stages of this research. They dealt with my excitement when things fell into place and despair when I was struggling, and for that I bow to their patience.

The school of planning and the administrative staff, who provided constant support with all non-academic aspects of this graduate program

Most importantly, I'd like to express my gratitude to God almighty, for blessing me with an amazing support system and ability to persevere.

Table of Contents

List of Tables:

List of Figures:

List of Maps:

List of Abbreviations:

ANN……………………………………………………………………………………………Artificial Neural Network

C-set...Choice set

CMA ....Census Metropolitan Area

CNL...Cross Nested Logit

CT………..Census Tract

CTV_D ....Car Truck or Van as a Driver

CTV_P ....Car Truck or Van as a Passenger

DA ....Dissemination Area

DT ....Downtown

GEV ....Generalized Extreme Value

GRT .... Grand River Transit

IIA .... Independence of Irrelevant Alternatives

GTFS………………………………………………………………………………….......General Transit Feed Specification

IV ....Instrumental Variable

LRT..Light Rail Transit

MLM.. Multinomial Logit Model

MSM..Motorcycle, Scooter, Moped

NLM...Nested Logit Model

OD... Origin Destination

RP…………………………………………………………………………………………………….....Revealed Preference

SP……………….. Stated Preference

Std Err..Standard Error

TAZ..Traffic Analysis Zone

TTS...Transportation Tomorrow Survey

UW..University of Waterloo

WLU………………………………………………………………………………………………Wilfred Laurier University

1 Introduction

The cities of the world are reinventing mobility to move towards a more sustainable system of urban travel. The transportation sector plays a huge role in global warming, contributing about 23% of total energy related Carbon dioxide emissions (Sims & Schaeffer, 2014). Climate Change concerns due to global warming, along with urban transportation issues like congestion (Urban Transportation Task Force, 2012), increased travel times and commute stress (Legrain et al., 2015) are fueling the increased interest in public transit and active transportation. Introduction of new modes of transportation require an understanding of the dynamics of daily travel choices of the masses to estimate travel demand. Mode choice modelling enables this understanding and informs policy decisions through forecasting and scenario planning. This exercise enables setting realistic mode share targets for the future and justifies investments in the projects which are proposed to achieve those targets.

The Region of Waterloo is a rapidly growing upper tier municipality with a population of 583,500 in 2016. The growth rate of the region at 5.5% exceeds both provincial (4.6%) and national (5%) averages (Region of Waterloo Community Profile, 2018). Furthermore, the region is projected to have a population of 742,000 by 2031, attracting nearly 10,000 new residents every year (Ministry of Municipal Affairs and Housing, 2017). In view of this growth, along with regional goals towards management of urban sprawl, efficient transportation, protection of farmland and environment and attracting new businesses, municipalities approved the Light Rail Transit (LRT) system called ION in 2011. The first phase of this project is a 19 km north south corridor connecting the downtowns of Kitchener and Waterloo, which opened to public in fall

2019, with an 18 km phase two extension into Cambridge which is still in the planning process. ION has been integrated into the existing transit system, by eliminating direct bus services along the LRT route. Thus, the LRT would have an impact not only around the transit station areas, but all transit travel in the region.

The introduction of LRT is expected to contribute in the shift of modal split, in favour of a greater share of public transportation in the region. The target for 2041 is to achieve 15% peak time transit shares, against the current 6.8% (Region of Waterloo, 2019). On a broad level, the purpose of this research is to understand the commuting mode choice behaviour of the residents of Waterloo region and to develop a mode choice model which will allow the examination of potential changes in commuting mode choice due to the LRT. This analysis will provide insights into ridership which can inform future public transit policy decisions.

1.1 Research Questions

The introduction of LRT in the existing transit system expands the mode choice alternatives for the region of Waterloo and adds to the sustainable transportation options available to the residents of the region. The motivation for this research is to understand the factors which will influence people to utilize LRT for their daily travel needs. This research employs quantitative methods like mathematical modelling and geospatial analysis to describe and choice behaviour.

This study aims to answer the overarching question – **"What is the potential impact of LRT on mode choice in the region of Waterloo?"**

In order to investigate this, the study has the following objectives:

1. Identify the factors which impact mode choice behaviour of commuters.
2. Describe the current commuting patterns in the region.

3. Develop a Model to establish a relationship between these factors and commuting patterns and choice behaviour.
4. Analyze how changes due to LRT will influence commuting mode shares.

1.2 Anticipated Research Contribution

It is said that the “best predictor of future behaviour is past behaviour.” This summarizes the underlying concept of modelling mode choice. The aim of modelling is to empirically understand how different factors impact the affinity towards a mode, and further, these relationships can be used to predict utility of new transit systems or improvements to existing infrastructure. These studies are especially useful as they point towards what exactly “matters” to the population in the region, which can then be exploited for efficient planning to achieve transit ridership goals.

This study is expected to provide analysis of the commuter travel patterns in the region. This will provide insight into the areas which receive and generate the largest commuter traffic volumes in region, the origins and destinations associated with them respectively, and the overall spread of different mode shares in the geography of the region. This has the potential to be a valuable resource for future transit policies, to identify areas for enhancing ridership by informing decisions around where infrastructure improvements should be made, which origin-destination pairs need better service and routes for transit system expansion.

The second major anticipated contribution of this exercise is to demonstrate the process of building a mode choice model, highlighting the factors which appear influential in the decision-making process around commuting mode choice for the residents of the region. Learning about these attributes will provide a template to test future scenarios and predict the outcomes of targeted policies and projects. Lastly, this study will reveal the change in mode shares that can be expected from the integration of LRT into the transit mix in Waterloo.

1.3 Thesis Structure

This study is organised into five chapters – Introduction, Literature Review, Methodology, Findings and Discussion and finally, Conclusion. The first chapter provides an overarching context for the research, background, goals and objectives.

The second chapter, Literature Review provides a theoretical framework this study. The existing relevant body of literature is discussed, and important findings are highlighted. It prepares a base by conceptual exploration of different modelling methods with respect to their pros and cons, and identification of attributes that can be linked to mode choice behaviour which informs the data collection process.

Following the literature review, Chapter 3 lays out the research approach and introduces the case study community – the Region of Waterloo. This chapter discusses the geographical unit of analysis, and further highlights the mathematics of regression analysis, data procurement and the process of estimating and simulating the model.

Chapter 4 details the results and findings from the spatial Analysis of commuter travel patterns in the region. The description is at two levels, generic for all modes, and secondly specific to active, auto and transit modes. Additionally, there is discussion on the findings of the regression analysis and the simulation of the Mode Choice Model.

Lastly, Chapter 5 briefly recaps the research goals, objectives and the results. The research contributions and limitations of the study are highlighted along with recommendations and opportunities or future research.

2 Literature Review

Commuting mode choice is dependent on a variety of economical, physical, social and environmental factors. Traditionally, mode choice models have been developed based on the general scale of trade-offs that individuals are willing to make between these factors (Sekhar, 2014). A combination of these factors drives an individual to exhibit mode choice behaviour which can be explained through utility theory focusing on economic profitability, and the theory of planned behaviour which focuses on the role of habits in choice behaviour (Beltman, 2014).

2.1 Theories Influencing Mode Choice:

Travel Utility is the value measurement made by an individual in travel decisions based on various factors like cost, comfort, travel time and safety (Y. Liu et al., 2019). Utility Theory is based on the assumption that an individual selects an alternative which maximizes his/her economic utility. This forms the basis of discrete choice models which parameterise utility functions in the form of independent variables (M. E. Ben-Akiva & Lerman, 1985). This theory works under an assumption that the value of each independent variable is known to the individual and that he/she can recognize the alternative which maximizes his/her utility (Beltman, 2014). However, although decisions are based on reasoning, an individual does not elaborately compute the value of each available alternative at the time of making a choice (Bruch & Feinberg, 2017). Furthermore, when utility maximization is considered synonymous to payoff maximization, it is implied that all individuals would make the same choice, which is not observed in real life situations (Hodgson, 2012).

An alternative view to expected utility theory suggests the role of habits in decision making(Gärling & Axhausen, 2003;Klöckner & Matthies, 2004;Marechal, 2018). This view is based on the observed repetitions in travel patterns, indicating that habitual travel choice discounts relevant information which rational decision-making accounts for, such as increased travel cost or travel time. Thus, larger the role of habit in the decision-making process, lesser is the role of deliberate information processing (Gärling & Axhausen, 2003). Furthermore, Lucas et al. (2011) suggests that vehicle ownership predetermines the choice of mode and intensity of its usage. Thus, initial investment made in the purchase of an automobile might deter an individual from walking or taking transit in favour of utilising the purchase, even if it is more expensive than alternative modes. This implies that an individual does not consider all the factors which impact his/her economic utility, instead opts for a mode out of personal situation or characteristics like attitude or habit. While both the theories partly explain mode choice behaviour, a collaborative view on reasons for mode choice variability is not clearly established, thus, the utility of each alternative mode cannot be determined. To overcome this barrier, discrete choice models use a random utility approach to predict mode choice behaviour, which assumes that utilities contain a deterministic component and a randomly distributed 'error term' (Markley, 2007). This implies that utility maximization and thus mode choice, can be represented as a probabilistic phenomenon which accounts for uncertainties in mode choice behaviour (Hess et al., 2018).

2.2 Mode Choice Modelling

Modelling is a key technique of understanding decision making processes and the quantitative and qualitative relationships which underlie these decisions (Hensher & Button, 2008). Utilizing these techniques to forecast and strategize transportation systems characterizes transportation modelling (Khan, 2007). Transportation Modelling is a four-step process which includes (1) Trip Generation, (2) Trip Distribution, (3) Mode Choice and (4) Route Choice (McNally, 2007). This study focuses on the third step of the process which attempts to understand the factors behind the choices which people make while selecting a mode and utilizes this understanding to predict their decision making behaviour (Koppelman & Bhat, 2006).Mode Choice modelling is an integral part of transportation modelling which reflects proportions of trips by alternative modes based on various performance variables and trip maker characteristics (McNally, 2007).

Discrete Choice Modelling

Discrete Choice Modelling is based on random utility maximization theory and is widely applied in transportation. There are three four types of discrete choice models, (1) Logit (2) Generalised Extreme Value (GEV) (3) Probit and (4) Soft Computing. The existing findings on these models have been summarized below:

2.2.1 Logit Models

Logit Models can be binary or multinomial. A binary model provides an individual with only two choice alternatives while a multinomial logit model (MLM) contains a larger set of choice alternatives (Khan, 2007). Eluru, Chakour, & El-Geneidy, (2012) used MLM to understand mode choice behaviour of university students in Montreal and concluded that students are more likely to choose transit in comparison to faculty and the major determinant of choice is time (in vehicle, waiting and walking). Furthermore, women were reported to be less sensitive to time compared to men, but these results are counterintuitive, and the author suggests further research in this regard. Moniruzzaman & Farber (2018) analysed factors which influence sustainable travel among students in GTA using MLM and found that availability of transit pass and bike ownership are significant factors. Although multinomial logit models are widely used, they work under the assumption that the random utility components of different available alternatives are independent and identically (IIA) distributed which means that unobserved components of utility of each alternative are not co related (Day, 2008). This implies that an individual's choice between two modes remains unaffected by availability of other options (Cheng & Long, 2007).

In a multinomial logit model, change in the probability of one alternative, equally draws from the probabilities of all the other available alternatives (Sekhar, 2014). This drawback is overcome by using a Nested Logit Model (NLM) which places similar alternatives together in different subsets, thus accommodating different degrees of interdependence between alternatives (Day, 2008). Abdel-Aty & Abdelwahab (2001) developed a triple nested logit model to understand mode choice behaviour in Florida and reported that travel time (access, transit waiting time, in vehicle), transit fare, car ownership and number of transfers were significant to the decision of mode choice. Furthermore, the researchers commented on the importance of a proper nesting structure, which accommodates existing alternatives and retains flexibility to add new ones. While NLM has advantages over MLM, it is noted that dividing similar alternatives into subsets is not an accurate representation of actual competitive structure among alternatives as there may be dependencies across subsets (Sekhar, 2014). Additionally, developing an appropriate structure for the nest becomes increasingly difficult as the number of alternatives increases.

2.2.2 Generalized Extreme Value (GEV) Model

Cross Nested Logit Model (CNM) is an improvement over the NLM, furthering increasing its flexibility, although at the cost of greater complexity (Bierlaire, 2006). It is the simplest form of GEV model that allows for different degrees of co-relations for unobserved utility across alternatives (Hess, 2005). This model acknowledges co-relation among alternatives that are in different subsets (Bastarianto et al., 2019) developed a tour-based mode choice model for commuters in Indonesia using MLM, NLM and CNM. The researchers reported that travel cost and time were the most significant attributes which contributed negatively to utilities, and the former is more significant than the latter. Additionally, car, bus and rail were preferred over motorcycles, and lower income commuters were more likely to ride a motorcycle. Lastly, female commuters were more likely to choose bus over rail. Although findings were generally consistent over all three models with comparable magnitude and signs, formal testing of the results reveals highest confidence in CNM, with some attributes showing greater significance in a cross nested approach than in the other two models. However, due to the improvement of log likelihood of NLM, when joint choices are grouped based on tour type, it was reported to be appropriate to evaluate a mode choice behaviour.

The GEV family has various other model structures, for example, Link Nested Logit Models, Paired Combinatorial Logit Models and Ordered Generalised Extreme Value Models, but the applicability of these models is limited due to their complex structure and the existing body of literature does not provide evidence of extensive use of these models in transportation planning.

2.2.3 Probit Models

Probit Models are used when the utilities of different alternatives in a choice set are co-related in a complex way. The major difference between the logit and probit models is the underlying assumption about error distribution (Sekhar, 2014) . The error term in a logit model follows a standard logistic distribution while probit uses a normal distribution. Ghareib (1996) evaluated the applicability of logit and probit models by investigating mode choice behaviour in different cities of Saudi Arabia. The research revealed that although probit models have a more reliable theoretical base, logit models are superior in terms of goodness of fit[1] and controllable calibration. These conclusions were supported by the findings of Dow & Endersby (2004), who reported that probit models should only be utilized when travel behaviour of population under study is complexly co-related. It was further reported that the main motivation of using probit over logit, the IIA problem, was rarely relevant in the results. Probit models are not very widely used in transportation planning as the results of this approach are comparable to logit models, but there is an increased complexity in calibration and loss of flexibility, which does not allow the model to be replicated between different space time sampling frames (Sekhar, 2014).

2.2.4 Soft Computing Models

Soft Computing models include Artificial Neural Network (ANN) and Fuzzy logic based models. These models include exploiting artificial intelligence for mode choice modelling (Sekhar, 2014) ANN is a data processing system which is inspired by the working of biological nervous systems. (Ramanuj & Varia, 2018). Hensher & Ton (2000) compared ANN and Nested logit structures to model commuter mode choice in Sydney and Melbourne and concluded that while ANN is a valuable tool to predict mode choice behaviour, it requires behaviour oriented datasets and the superiority of either of the two models could not be established. Xie, Lu, & Parkany (2003) developed mode choice models using ANN and Decision Tree (DT) methods and compared them

[1] How well the results fit a set of obervations

with a traditional multinomial approach, and reported that the models provide comparable results, but ANN and DT have slightly better performance.

Fuzzy logic approach has been explored by various studies (Kedia, Saw, & Katti, 2015; Pulugurta, Arun, & Errampalli, 2013; Edara, 2003) and the model results have been reported to be more accurate than traditional multinomial logit models and enable closer understanding of human behaviour. This approach is a modelling tool which enables evaluation of linguistic variables like high, low, often, many, rarely. Thus, it breaks free from the rigid approach of logit models where events either occur or they don't (Errampalli et al., 2013). This approach acknowledges that real life decisions are complex and modeling them requires accommodation of some degree of uncertainty (Kumar et al., 2013). Fuzzy logic approach is user friendly, less time consuming and does not require extensive coding knowledge (Edara, 2003). However, fuzzy logic models, are dependent on stated preference data and require respondents to think more deeply on their choices, instead of providing a binary 'yes/no' response. Additionally, variables like 'high', 'often' may be interpreted differently by different respondents. Hence, the researcher has less control over the experiment, and credibility of results depends on the seriousness of respondents.

Conclusively, Although Soft Computing Models have been reported to yield more accurate results, they have not been considered in this research due to data availability, technical and financial constraints.

2.3 Mode Switch Modelling

Mode choice modelling has been attempted by various modelling techniques, resulting in valuable outputs and lessons. However, this research intends to not only understand mode choice behaviour but determine potential of mode shift towards transit due to introduction of LRT. Taking into account the research objectives, this section explores the approach, results and lessons from mode switch modelling which have been documented in the literature.

North American Modal Shares are largely dominated by cars. Ontario transports 65.6% of its commuters through single driver cars, and an additional 12.3% through carpooling. (Ministry of Finance, Ontario, 2016). Although the use of private vehicles has reduced from 79.9% (2016) to 77.9% (1996) for commuting purposes, the share of private vehicles is still significant. While most transportation policies in the region outline reduction of GHG emissions by promoting public transit and active transportation, it is important to understand the dynamics of this choice. Furthermore, it is also interesting to note how commuter behaviour changes due to the introduction of a new transit mode in the existing mix of mode choice alternatives for an individual.

Forsey, Habib, Miller, & Shalaby (2013) analyzed the impacts of BRT-Lite system on commuting mode choice in Toronto. This study used revealed preference data collected after the system was opened to public use and compared it with pre-BRT data in a GEV model, to capture heterogeneity of the datasets. The variables used were In-vehicle time, walk time, wait time, cost, number of household vehicles, trip distance, gender (dummy) and Age (dummy). The study revealed that introduction of BRT line did have an impact on mode choice preference of commuters, and transit improvements have greater impact on mode share than traffic congestion. Furthermore, it was reported that in vehicle time was the least burdensome to the travellers. Additionally, the model revealed post-secondary school trips were less than the actual observed change in mode share. This was attributed to change propelled by branding, advertising and effective communication to students.

Ladhi, Ghodmare, & Sayankar (2018) used stated preference data for forecast and revealed preference data to understand present mode choice behaviour in New Delhi, India. A binomial logit model was developed to understand the personal and travel characteristics which impact mode choice. The variables considered in modelling were gender, vehicle ownership, ingress distance to Metro Rail, age, income and cost. The study revealed that 28.8% of users switched from personal motor vehicle (pmv) to metro while 57% shifted from buses. The former was attributed to excessive vehicular congestion, lesser time and cheaper cost while the former was a result of overcrowding in buses, lesser travel time in metro and lack of direct bus service Werner et al. (2016) evaluated the attractiveness of a light rail extension and discredited the theory most LRT riders would be existing bus riders, and revealed that LRT can, in fact, attract additional ridership.

Yang et al (2013) investigated the behaviour impacts due to introduction of Metro Service in Xi'aN, China using a binomial logit model employing stated and revealed preference data. The study indicated that mode shift to metro is more attractive to auto drivers in the suburban regions, and the preference to metro is higher among female auto and taxi users than males. Furthermore, it was highlighted that shift from auto mode might decreased to 8% from the estimated 19% due to incomplete transfers and inadequate modal joints. The attributes considered in the study were gender (dummy), occupation, income, car availability, trip purpose (dummy) and cost.

Ashalatha, Manju, & Zacharia (2013) analyzed mode choice behaviour of commuters in Thiruvananthapuram, India using a multinomial logit model using revealed preference data. The study revealed that with age, preference to car increases; and increase in time/distance and cost/distance causes people to shift from public transit (bus) to car and two wheelers. The attributes considered were age, gender, income, vehicle ownership, distance and cost.

Idris, Habib, & Shalaby (2014) used multinomial logit model to study commuter's preferences and mode switch behaviour toward public transit in Toronto, Canada. In addition to the commonly considered variables like cost (transit and parking), time (in vehicle, waiting, access and egress), vehicle ownership etc, transit planning attributes were considered, for example, transit technology (BRT, LRT, Subway), crowding level and schedule delay. The study revealed that people are more likely to shift towards rail-based modes instead of rubber-tired modes like a bus transit system.

Both multinomial logit models and binomial logit models find applicability in the analysis of mode switch behaviour. An overview of literature reveals that binomial logit models are used to determine the probability of shift of passengers one mode to another while multinomial and nested logit models are more useful to understand the factors determining choice behaviour, which can then be employed to analyze the probability of mode switch based on how these attributes will be altered by the introduction of a new mode of transportation.

2.4 Determinants of Mode Choice

The factors which determine mode choice have been explored by various studies (Chee & Fernandez, 2013; Creemers et al., 2012; Buehler, 2011; Tyrinopoulos & Antoniou, 2013). Ortúzar S. & Willumsen (2001) categorized these determinants into three groups, (1) trip related variables

(2) mode related variables and (3) Variables related to the characteristics of the trip maker. Any additional variables such as weather, form a fourth category of external factors. (Beltman, 2014). All these variables are considered and then eliminated in the modelling stage of the study depending on their statistical significance based on empirical evidence and purpose of the study. The various determinants and their impact on mode choice are discussed below:

2.4.1 Trip related variables

Trip related variables include travel time and distance, trip purpose and trip timing. Travel time is shown to have a significant impact on mode choice (Creemers et al., 2012), and degree of reliability on travel time boosts public transport ridership (Van Loon, Rietveld, & Brons, 2011) Furthermore, separating access, egress and in vehicle time contributes to a better understanding of choice behaviour (Hensher & Rose, 2007). Generally, reduction in travel time and number of transfers trends to boost transit ridership, and longer access and waiting times are a deterrent to choosing transit for travel. However, (Idris, Habib, & Shalaby, 2014) reported that travel time and cost had lower importance in mode switch behaviour than other factors like transit technology, crowding and schedule delay.

Corpuz (2007) studied the factors which impact mode choice and reported that car use is the highest among commuters while public transit is highest for educational trips followed by commuter trips. This establishes a link between trip purpose and mode choice. However, this was attributed to other factors such as availability of free public transit passes and lower car ownership rate among students. Similarly, high commuter share in public transportation was credited to high serviceability. Thus, the mode choice depends on the other characteristics of the trip and traveller rather than its purpose per se.

The time of the day at which the trip takes place is a combination of various factors such as the purpose of trip, schedule flexibility and travel time (Day, 2008). Corpuz (2007) reported that private vehicles remain the most used mode throughout the day while transit usage picks up during the mornings and late afternoons when the number of trips also increases. However, the increase in transit usage relative to car is not substantial, and the observed difference can be attributed to the speed, cost and service frequency advantages which transit enjoys during peak period, reducing the attractiveness of the car.

2.4.2 Mode Related Variables

Mode related variables may be quantitative, for example, travel cost and transit frequency , or qualitative, like comfort, security and reliability (Ortúzar S. & Willumsen, 2001). Travel cost includes direct costs, such as purchase of transfer, ticket or gas and indirect costs such as parking (Beltman, 2014).

Travel cost reduction encourages choice of the cheaper mode, (Bastarianto et al., 2019), however, Ganji et al, (2013) reported that decreasing travel time is a more effective way of boosting transit ridership than cost reduction. Furthermore, individuals who pay higher parking costs are more likely to switch mode rather than those who enjoy free or underpriced parking, in which case reducing parking availability propels the switch to transit (Alavi, 2016). Bai, Li, & Sun (2017) concluded that public transit ridership seems to witness a greater hike due to reduction of travel cost by 10%, rather than a similar increase in car cost. Additionally, The transit alternative which offers higher frequency of service tends to be more attractive to travellers (M. E. Ben-Akiva & Lerman, 1985).

Mode based characteristics such as security, reliability and comfort have been shown to influence mode choice of an individual (Ben-Akiva & Morikawa, 2002; Hu, Zhao, & Wang, 2015) These factors can be described as an individual’s perceptions towards a mode and offer explanation for variations in mode choice behaviour between individuals exhibiting similar characteristics, which is otherwise represented by alternate specific constant in a model (Bahamonde-Birke et al, 2017). However, Al-Ahmadi (2017) concluded that the cost of obtaining mode perception data and the difficulty of quantification reduces the usefulness of these parameters in choice modelling.

2.4.3 Traveller Related Variables

Various trip maker characteristics such as age, gender, car ownership and income influence the mode choice behaviour and are recurring factors in most mode choice models (Hensher & Rose, 2007; Ashalatha, Manju, & Zacharia, 2013; Ding & Zhang, 2016; Hasnine, Lin, Weiss, & Habib, 2017). These factors may also have an inter relationship, for example, an individual with higher income, is more likely to own a car, and hence travel using it (L. Cheng, Chen, Wei, Wu, & Hou, 2014) Furthermore, personal attitudes play a role in the decision-making process as people with a negative attitude towards public transit are more likely to use a car (Popuri et al, 2011).

An individual’s age influences the factors which he/she considers while making mode choice decision. For example, quality of service such as transit frequency is less important to older people (65 and above), than bus stop density (Su et al., 2009). Furthermore, the purpose of travel for older people is largely consists of leisure and shopping trips. Additionally, the younger old (65

– 74) are more likely to use a car (Schmöcker et al., 2008), and trips which requires vehicle or route transfer discourage the use of transit (Nitta, 1998).

Yang et al (2013) examined gender-based differences in mode choice and observed that men prefer travelling by car while women prefer bicycles. In context of commuting, however, Habib (2014) reports that women are more car oriented because it offers more safety and comfort, and those with children are more likely to drive alone. However, Acker & Ng (2018) reported that women travel shorter distances and prefer public transit to car. It is further suggested that this may be a result of gendered division of work in households, where women are subject to caregiving or household responsibilities along with employment, resulting in choice of housing or job location which minimizes travel time. The contradiction in results in the existing body of literature suggests that generalizing the role of gender in mode choice is difficult as it may be a result of many unknown factors.

Car ownership has been reported to be an important factor in mode choice decision of an individual (Sartori & Robledo, 2013). Individual preferences like inclination towards environmental protection (Lo et al., 2013) and awareness of environment concerns discourages car ownership (Tao et al., 2019). Furthermore, Tao et al. (2019) through a comparative assessment between a Guangzhou and Brisbane concluded that vehicle ownership in Guangzhou is associated with higher income while in Brisbane vehicle ownership is common in low income households. The impact of incomes on mode choice is observed in both cities, but more so in Guangzhou. Furthermore, car owners have been reported to have a negative attitude towards transit usage in both cities.

2.4.4 External Variables

Various studies have studied the influence of external factors like weather (Böcker, Dijst, & Prillwitz, 2013; Saneinejad, Roorda, & Kennedy, 2012; Liu, Susilo, & Karlström, 2017) and landuse (Maat, van Wee, & Stead, 2005; Sun, Ermagun, & Dan, 2017; Zhang, 2004) on mode choice behaviour.

Precipitation, Wind and temperature are the weather parameters that noticeably influence modal choice. (Beltman, 2014) Precipitation discourages cycling, both in commuter and recreational cyclists (Brandenburg et al., 2004). Furthermore, precipitation of more than 1mm an hour was observed to cause modal shift from bicycle to car (Sabir et al., 2008). It was further concluded through their research in Netherlands, that in comparison to normal temperatures (0-10°C), the chances of selecting a bicycle for travel decrease by 5.5% if the temperature is low, i.e. between -8 to 0°C, and contribution of cars in modal share increases to up to 52% in the same temperature range. The same research also concluded a reverse in trend for temperatures above 10° and up to 25°C, but after the threshold of 25° is reached, cycling witnessed a sharp decline. Wind too, has a strong impact, decreasing cyclist numbers two times than the number of pedestrians on windy days (Saneinejad et al., 2012).

Stover & McCormack (2012) studied the impacts of weather on transit ridership and reported seasonal variation in the influence of weather parameters in modal choice. Precipitation was observed to reduce transit ridership in all seasons, cold temperatures during the winter and high winds during winter, spring and autumn. Similar conclusions were made by (Guo, Wilson, & Rahbee (2007), with warmer temperatures being reported to increase ridership.

Urban form is an important factor which influences modal choice as with increase in distance between origins and destinations, the rate of bicycling decreases, due to additional time and effort required for the trip (Heinen et al., 2010). Generally, individual choice towards active transportation tends to rise with mixed land uses, which results in decrease in distances between work, home and recreation facilities; seamless street connectivity and employment density at the origin and destination (Rodríguez & Joo, 2004). Shorter trip distances achieved by restricting sprawl and promoting compact mixed-use developments in northern and westerns Europe are important factors responsible for high cycling rates in the region (Pucher et al., 2010). Although, Rodríguez & Joo (2004) suggested limited relationship between residential densities and bicycle use, higher densities have an inverse relationship with car use, and thus promote cycling (Heinen et al., 2010).

The tables below summarise variables which have been considered in various relevant studies. The first table provides the variables and lists the studies, while the second table highlights the details of the associated studies. Note that 'general' study focus indicates that the intent of the study is to understand mode choice behaviour without a specific focus.

Table 2.1: Variables In Mode Choice Modelling

Variable	**Description/ Coding**	**Studies/ Study ID**
Time	**Transit:** Access Time (min) Waiting Time (min) In vehicle Time (min) Egress Time (min) OR Mean travel Time by transit (min) Initial waiting time for transit users (min) **Auto:**	1,2,3,4,5,6,8,9, 10,11,12,13

	Mean In Vehicle Time (min)	
Cost	**Transit:** Transit Fare ($) **Auto** Parking Cost ($) Running Cost ($)	1,2,3,4,5,6,8,9, 10, 11,13
Transit Pass	1 – Yes / 2 - No	10,
Age	Coded in Ranges. Ex: 15 – 24 25 - 34 35 – 44 45 – 54 55 – 64	1,2,3,5,6,10,11,12
Employment	1 - Full Time 2 – Part Time 3 - Casual 4 - Not at all in last 6 months	1,2,3,6,9,13
Comfort	1 – Very Comfortable with seat 2 – No seat but freedom of movement 3 – No seat, also crowded	9
Household Size	Number as Reported	1,2,7,10
Car Ownership	1 – Yes 0 – No Or Number as Reported	2,3,4,5,6,8,9,11,12,13
Income	(Household or Individual) Coded in Ranges Ex: <$ 5000 $5000 - $10,000 $10,000 - $15,000 $15,000 - $20,000 $20,000 - $30,000 $30,000 - $40,000 ……. … >100,000	1,2,5,6,8,9,12,13
Gender	1 - Male 0- Female	1,2,3,5,6,7,9, 10, 11,12,13
Marital Status	1 – Single 2 – Married 3 – Divorced 4 – Widowed	2,7
Living Situation	1 – Alone 2 – Roomates 3 – Partner	7

	4- Family	
Bike Ownership	1 – Yes / 2 – No	7
No. of Transfers	Number as reported (0, 1,2,3+)	3,4
Children in HH/ Retirees in HH/ Employed in HH	Number as reported	1,8,10
Transit Techonology	1 – LRT 2 – BRT 3 – SUBWAY	2
Mode	1 Car (As driver, as passenger) 2. Bus 3. Train 4. Others	1,2,3,4,5,6,7,8,9,10,11,12,13
Dwelling Type	1 – House 2 – Townhouse 3 – Appartment	2
Driver's License	1 – Yes / 2 – No	2,3
Purpose of Travel		8,13
Distance Travelled	Distance (in km)	5,6,10,11
Employment Density	Employees/ km square	-
Population Density	People/ km square	8
Landuse Mix	Calculated b/w 0 (no mixture) – 1 (Equal Mixture) using Cervero's (1998) Entropy Formula	7
Density of Pedestrian Networks	km/ km square	7
Transit Frequency	Number of Days per week as Reported (How often commute to work place is made)	7
DERIVED	Time (Access, In vehicle, Egress) Cost Car Availability Income Gender Age HH Structure Density Number of Transfers	

Table 2.2: Context of Reviewed Studies

Id. No.	Author & Year	Country	Model Type	Study Focus
1	(Hensher & Rose, 2007)	Australia	NLM	General
2	(Idris et al., 2014)	Canada	MLM	Mode Switch & Transit
3	(Eluru et al., 2012)	Canada	MLM	General
4	(Abdel-Aty & Abdelwahab, 2001)	USA	NLM	General
5	(Ashalatha et al., 2013)	India	MLM	General
6	(Y. Liu et al., 2019)	China	NLM	General
7	(Moniruzzaman & Farber, 2018)	Canada	MLM	Sustainable Transit
8	(Buehler, 2011)	Germany & USA	MLM	Determinants – USA vs Germany
9	(Ding & Zhang, 2016)	China	MLM	General
10	(Hasnine et al., 2017)	Canada	MLM, NLM, CNL,	Urban travel and mode choice
11	(Forsey et al., 2013)	Canada	GEV	New Transit and Mode Choice
12	(Ladhi, Ghodmare, & Sayankar,2018)	India	BLM	New Transit and Mode Shift
13	(Wang et al., 2013)	China	BLM	New Transit and Mode shift

2.5 Data in Mode Choice Models – Aggregate v/s Disaggregate

Mode Choice Modelling can be undertaken using an aggregate or disaggregate approach. The former uses data that represents the behaviour of a group instead of a single individual, employed in trend analysis and direct demand models (Sekhar, 2014). This approach has been criticized in literature due to exclusion of the behavioural aspect of individuals and non-generalizability, leading to the use of disaggregate choice models (Barff et al., 1982).Disaggregate models use data collected at an individual level, which is then utilized to describe combined behaviour of the group thus allowing greater interpersonal variability (Koppelman & Bhat, 2006).

Despite the support in literature for disaggregate mode choice models, aggregate models have advantages, as they are quicker to estimate because the data can be sourced from secondary sources. Furthermore, aggregate choice models find validation from many researchers due to weaker impacts of behavioural attributes and chances of inadequate model calibration due to lack of enough observations in case of disaggregate data (Ortuzar, 1982). Conclusively, the time constraint and the nature of available data push this study towards an aggregate approach which nevertheless has potential to produce robust results.

3 METHODOLOGY

This thesis analyses how different factors impact mode choice behaviour in the region of Waterloo with a focus on the potential changes due to the introduction of the ION LRT. The overall goal of the research is to contribute towards transit planning in the region by enabling a better understanding of influences on transit mode choice for commuting. This chapter elaborates the research approach and methodology adopted in this study to model the commuting travel choices in the region.

This chapter is organized into four major sections. The first section introduces the study community, which is the Region of Waterloo, highlighting its employment and transportation profile. The second section elaborates on the approach and mathematics of the model. The third section explains the use of data, its sources and assumptions. Lastly, the fourth section briefly highlights the steps that were undertaken in the calibration and estimation of the model.

3.1 Introduction to the Study Area

The Region of Waterloo consists of seven municipalities – Wellesley, Woolwich, Waterloo, Kitchener, Cambridge, Wilmot and North Dumfries, with a total population of 601,220 (Region of Waterloo, 2019). Growing at the rate of 5.5% per year, it is expected to add 185,000 people over the next 15 years. Located about 115 kms from Toronto, the region is home to three post-secondary educational institutions, an expanding tech centre and has a GDP per capita of $51,536, higher than both, provincial and Canadian averages.

The region has a diverse economic base with manufacturing sector providing highest employment (17.8%). Trade (15%), Educational Services (10%), and Technical Services (9%)

account for more than half of the employment base in the region in 2016 (Region of Waterloo, 2017). The region hosts tech giants like Google, Shopify, Open Text and Blackberry which are attractive employers for local talent in computer science and engineering among other fields. These factors support the above average growth of the region and predicted population growth.

The region's mode shares is largely dominated by cars, amounting to about 87.7% of all commutes. Furthermore, 80.1% of these trips are undertaken by a single driver, and the remaining 6.8% include trips in car as a passenger. Additionally, only 6% of total trips in the region are undertaken via transit and the active transportation share is slightly less at 5.5%. (Statistics Canada, 2017). The domination of car in the existing mode split, along with expected population growth prompted the planning and development of a light rail transit (LRT) system in the region. The Region intends to achieve more sustainable mode shares by 2031, targeting increase in transit usage to constitute 15% and active transportation (walking/cycling) to constitute 12% of the total trips. (Region of Waterloo, 2010). In addition to promoting transit and active transportation usage, the region intends to limit urban sprawl, manage congestion, preserve region's environmental resources, enhance walkability and increase the construction and maintenance of new roads.

The LRT was approved in 2011, planned to complete in two phases, the 19 km north south corridor connecting the downtowns of Kitchener and Waterloo in the first phase, and 18 extension into Cambridge in the second phase. The advantages of this project include (1) congestion reduction through strategic route planning, (2) facilitates compact development, hence preserving agricultural land and ground water resources (3) Enhances walkability by inducing compact development (4) eliminates the need to build 500 kms of new road lanes over the next 20 years, reducing construction cost by 40% (5) Reduces sprawl by attracting new development in the existing built up area. (Region of Waterloo, 2014b)

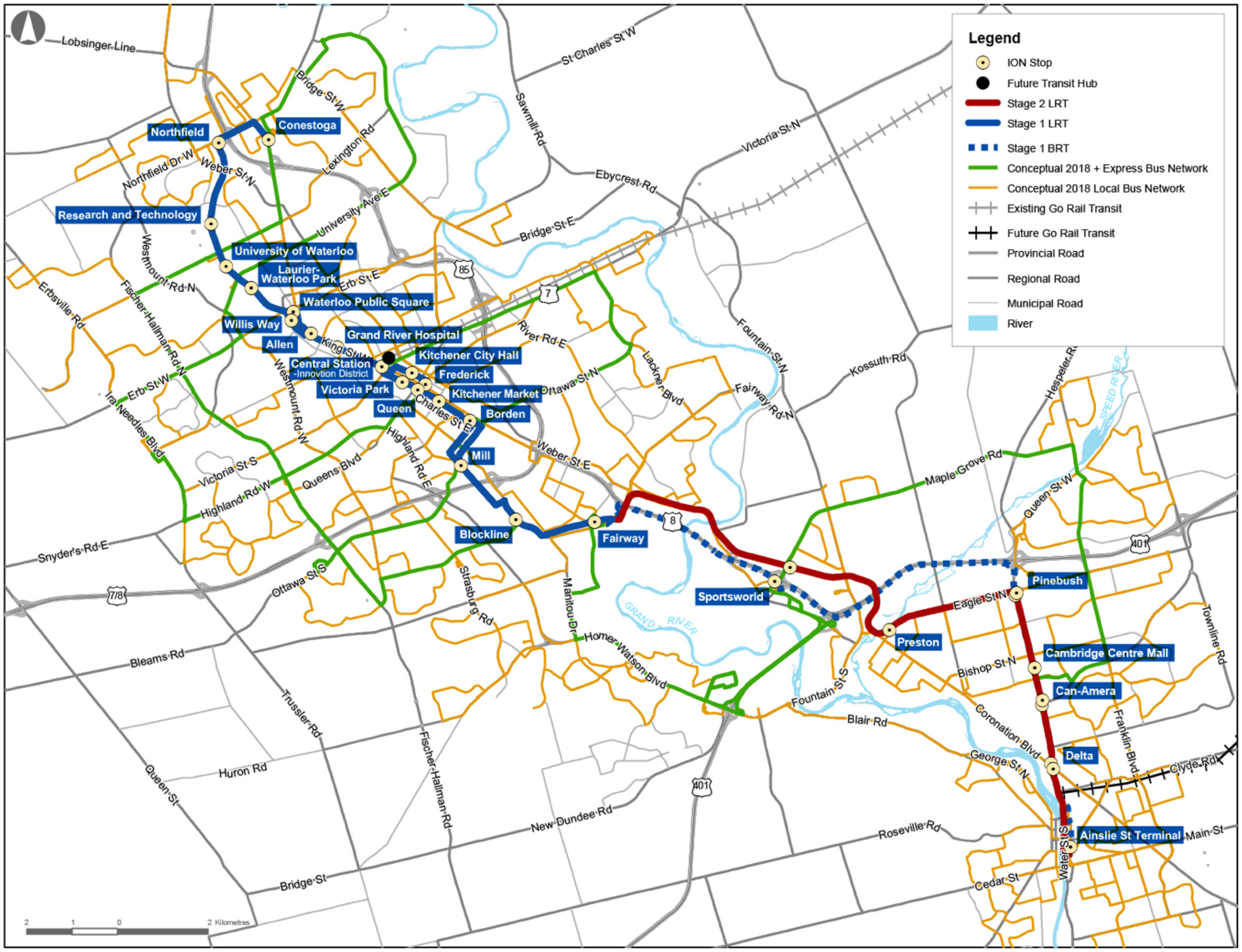


Figure 3.1 Planned LRT System

3.2 Modelling Approach

Four types of mode choice models - Logit, GEV, Probit and Soft Computing, and their variations were explored in the literature review. The utilization of these models in understanding mode choice behaviour is well documented in the literature along with their advantages and disadvantages. Multinomial logit models are critiqued due to the underlying assumption that the relative probability of choice between two alternatives remains unaffected by availability and characteristics of other modes. This issue is resolved by using a nested logit approach, which groups similar alternatives together, allowing for interdependence between alternatives. Nested logit approach is widely used, is simpler to formulate and estimate than Probit or Soft Computing models, and has been reported as being fairly accurate, delivering similar results when compared to more complex GEV modelling frameworks like the Cross-Nested logit model. However, the existing body of literature also cautions about the difficulty in nesting alternatives, which results in models with relatively low goodness of fit measures. This exercise compares both the Multinomial logit and the nested logit models to arrive at the model specification which best explains the mode choice behaviour in the region. Conclusively, the Literature suggests that Nested Multinomial Logit Modelling approach is most suitable to develop a mode choice model for the Region of Waterloo, which will be tested ahead.

3.3 Model Formula

Nested logit models group together similar alternatives, which allows for dependence between choice behaviour for similar modes within one subgroup but still maintains independence between choice behaviour across different groups. (Abdel-Aty & Abdelwahab, 2001). This section uses a generic nested structure (Figure 3.2) for mode choice to explain the mechanism of model calibration and estimation.

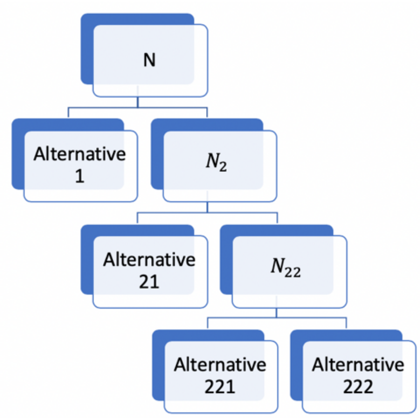


Figure 3.2 Sample Nest for NLM

The efficient methodology for the estimation of Nested logit models was explained by (Bierlaire, 1995), and has been summarized below:

In the nest depicted in figure 3.2, the probability (P) that an individual x chooses an alternative $n \epsilon N_i$ can be described as:

$$\mathrm{P}(n) = P\left(\frac{n}{N_i}\right) \times \mathrm{P}(N_i) \qquad \text{--- (1)}$$

Where, $\mathrm{P}(N_i)$ is the probability that x chooses N_i while $P\left(\frac{n}{N_i}\right)$ is the probability that an individual chooses the alternative n knowing he has chosen N_i.

a. Calculating Utilities

As discussed in the literature review, a nested modelling approach works on the principle of random utility maximization. The utility (U) of mode choice alternative i is the sum of deterministic component (V) and Gumbel-distributed random component (Ɛ). The equation for utility where an individual x chooses an alternative i can be represented as:

$$U_i^x = V_i^x + \mathcal{E}_i^x = \sum_j \beta_j \, c_{ij}^x + \mathcal{E}_i^x \qquad \text{--- (2)}$$

The Systematic Utility (V) is the function of observed socio-economic attributes (c) and an alternative specific constant associated with that attribute (β) for the j^{th} characteristic considered by x for the alternative i. For example, if the attributes used in the model are travel cost (TC), Access Time (AT) and In-vehicle Time (IT), the Utility for mode alternative i will be:

$$V_C = \beta_1 \times TC_i + \beta_2 \times AT_i + \beta_3 \times IT_i \qquad \text{--- (3)}$$

Where,

β_k - Coefficient which Determines the significance of the associated attribute in the utility of the mode

Thus, the expected maximum utility for subset N_i where θ_i is a positive coefficient to be estimated can be defined as:

$V_{N_i} = \theta_i \, log \sum_{j\epsilon N_i} e^{V_j}$ --- (4)

b. Probability Equations

The logit formulation for $P\left(\frac{n}{N_i}\right)$ and $P(N_i)$ in (Equation 1) can then be described as:

$$P(N_i) = \frac{e^{V_{N_i}}}{\sum_{j=1}^{p} e^{V_{N_j}}}$$

$$P\left(\frac{n}{N_i}\right) = \frac{e^{V_n}}{\sum_{j\epsilon N_i} e^{V_j}} \quad \text{--- (5)}$$

For Example, In figure 3.2, if the deterministic utility for a mode choice alternative i is represented as $V_i(i = 1, 21 \; or \; 221, 222)$, the expected maximum utilities will be:

$$V_{N_{22}} = \theta_2 \log(e^{V_{221}} + e^{V_{222}})$$

$$V_{N_2} = \theta_1 \log(e^{V_{21}} + e^{V_{22}}) \quad \text{...(6)}$$

Now the Probability of alternative 222 can be calculated as:

$P(222) = P\left(\frac{222}{N_{22}}\right) \times P\left(\frac{N_{22}}{N_2}\right) \times P(N_2)$, where individual probabilities can be represented as:

$$P(N_2) = \frac{e^{V_{N_2}}}{e^{V_1} + e^{V_{N_2}}}$$

$$P\left(\frac{N_{22}}{N_2}\right) = \frac{e^{V_{N_{22}}}}{e^{V_{21}} + e^{V_{N_{22}}}}$$

$$P\left(\frac{222}{N_{22}}\right) = \frac{e^{V_{222}}}{e^{V_{221}} + e^{V_{N_{222}}}} \quad \text{---(7)}$$

c. Model Estimation

Nested Logit models can be estimated simultaneously, using the Full Information Maximum Likelihood (FIML) method, or sequentially, using the bottom up approach where the lower levels of the nest in the model are calculated first, which are then entered as values to calculate the upper levels. Sequential estimation is, however, less efficient, and the superiority of the FIML has been established by various studies. (Hensher, 1986; Brownstone & Small, 1989; Bierlaire, 1995). FIML is widely accepted for estimation of nested models as it uses all available information to estimate the entire model in a single phase, instead of calculating the upper and lower levels separately, and constraints the common parameters across all alternatives to be equal. (Train, 2003)

The maximum likelihood method enables computation of joint probability of the whole sample as all the characteristics are defined in the data set, and the Probability that an individual x, chooses an alternative n_x is dependent only on the coefficients β (Equation 2) and θ (Equation 6). For each set of (β, θ), the likelihood function ($\mathcal{L}$) for joint probability for sample size X is:

$$\mathcal{L}(\beta, \theta) = \coprod_{x=1}^{X} P(\beta, \theta) \qquad \text{… (8)}$$

In maximum likelihood method, the intention is to find set of coefficients (β, θ) which maximizes the above equation. This can be represented as:

$$max_{(\beta,\theta)}\mathcal{L}(\beta, \theta) = max_{(\beta,\theta)} \coprod_{x=1}^{X} P_x(\beta, \theta) \text{ where:} \qquad \text{… (9)}$$

$P_x(\beta, \theta)$ is the probability given by the model that n_x is chosen.

If $\mathcal{L}'(\beta, \theta) = \log \mathcal{L}(\beta, \theta)$, where $\mathcal{L}'$ is the log likelihood function, the equivalent mathematical equation can be represented as:

$$max_{(\beta,\theta)}\mathcal{L}'(\beta, \theta) = max_{(\beta,\theta)} \sum_{x=1}^{X} \log P_x(\beta, \theta) \text{ … (xx)} \qquad \text{…(10)}$$

d. Model Outputs

Running the model with all the datasets is expected to give (1) Significant Parameters, their coefficients and corresponding standard errors and t-statistics values (2) Log likelihood values at zero i.e. the equal probability model and at convergence and finally (3) Rho-squared indicators (ρ^2). (Anwar, 2013)

Log Likelihood of a model is not absolute, meaning it has no meaning independently and is used to compare different models to determine predictive power and accuracy. Standard Error (Std. Err.) is the error associated with the coefficients of the parameters used for calculating the utilities in the model. Standard Errors are used to determine z, P>|z| and form confidence intervals for parameters. Furthermore, z and P>|z| (also called p value, or 2 tailed test value) are used to determine significance of a variable in the logistic regression by comparing p value to an alpha, which is 0.05 at 95% confidence interval and 0.15 at 85% confidence interval. In this process of null-hypothesis testing, p values less than the chosen alpha are statistically significant. Significant p-values indicate that the parameter is an important factor in estimation of utilities for a mode and should be retained in the model. The magnitude and direction of significance is determined by the coefficient values. Positive coefficients imply direct while negative coefficients represent an inverse relation with the probability of choice. Additionally, nested logit models are usually calibrated at 95% confidence intervals, meaning that that there is 95% chance that the calculated interval will contain the true population mean.

Rho-squared indicators (ρ^2) are used to describe the goodness of fit measures of the model. ρ^2 values are based on relationship among log likelihoods at $\mathcal{L}'(0)$, which has no coefficients and hence results in equal probability of each mode alternative being chosen, $\mathcal{L}'(C)$ representing a constants only model, $\mathcal{L}'(\beta')$ which is the function for estimated model and finally $\mathcal{L}'(*)$ which is the Log likelihood for perfect prediction model. The value of ρ^2 lies between 0 and 1, where 0 implies that the estimated model is same as the referenced model, while 1 implies that the estimated model is a perfect model. Additionally, $\mathcal{L}'(*)$ always has a value of 0. ρ^2 can then be mathematically represented as:

$$\rho^2 = 1 - \frac{\mathcal{L}'(\beta')}{\mathcal{L}'(0)} \qquad \text{... (11)}$$

3.4 Data Description

3.4.1 3.1 Unit of Measurement

It is essential to establish the unit of measurement before elaborating on data sources, procurement and usage. Most disaggregate Mode choice models are developed using an individual respondent as the unit of measurement (Idris, Habib, & Shalaby, 2014; Liu, Chen, Wu, & Ye, 2019; Moniruzzaman & Farber, 2018), while aggregate models use Traffic Analysis Zones (TAZ),Census Tract (CT) or Dissemination Area (DA) (Foth, 2013). Disaggregate data presents a picture of individual characteristics and preferences while aggregate data represents the travel patterns resulting from them, thus, as this research intends to study the impact of introduction of a new mode (LRT) on transit ridership, the latter is a suitable approach. Additionally, aggregate data is readily available, easier to compute and quicker to collect.

In the context of data availability, the most significant data set for this study is the commuter travel flows which is available at the CT level. Since the origins and destinations of trips are based on census tracts, further aggregating or disaggregating this data has a tendency to result in misinterpretation and false conclusions, while other socio-economic characteristics can be obtained or calculated for this unit of measurement. Thus, for this study, Census Tract is selected as a suitable unit of measurement.

Census Tracts (CTs) are relatively small geographic areas which have a population between 2,500 and 8,000 persons, with a preferred average of 4000; but Central Business Districts, Downtowns, major commercial zones or industrial zones may have population outside of this range. These are 'created' in Census Agglomerations or Census Metropolitan areas which have a core population of 50,000 or more in the previous census and maintained in case of subsequent population decline. Lastly, CT boundaries follow easily recognizable physical features, streets, property lines or municipality limits, and may be split into two if population exceeds 8000.

3.4.2 Data Preparation

a. Mode of Travel

The most important source of data for this research is the Journey to Work, a special tabulation obtained from Census of Canada, made available for use in this thesis through the generosity of Professor Ahmed El-Geneidy of McGill University. The data represents the commuting flows between Census Metropolitan Areas (CMA) at the Census Tract (CT) level for the employed labor force above the age of 15 years, having a usual place of work, for 25% of the sample. The employed labour force in this context, includes individuals who had employment and are considered a part of the labour supply in the economy during the reference week of the census survey and does not include full time students. This data was used to create cross sectional matrices to represent the total number of trips between all census tracts and trips by different modes.

Each observation represents the trip origin and destination for aggregate commuter flows between census tracts. This data allows tabulation of the number of workers in each CT and the number of jobs in each CT. Each commute is further classified as per the average annual income of the respondent and their main mode of commuting. The "main mode of commuting" has different alternatives under sustainable and auto modes. Auto modes include Car, truck or Van as a driver (CTV_D), Car, truck or Van as a passenger (CTV_P) with 1,2,3 or more passengers and Motorcycles, Scooter and Moped (MSM). For the purpose of this study, CTV_P has not been differentiated by the number of passengers and has been considered as one mode – "Car, Truck or Van as a Passenger". The sustainable modes include active transportation – Bike (BI) and Walking (W), and transit in the form of Bus (BU). An additional category of Other Methods (OM) includes modes which do not fall into either of these categories such as skateboards, hoverboards or electric scooters. This category was excluded in model formulation as all these modes were grouped together, making travel time calculations impractical. Furthermore, only 0.75% of the total trips were made using these modes.

It is essential to note an underlying assumption of the data, that commute to work originates from the place of residence, which might not be true in case the respondent is on a business trip and reported the place of work or main mode of commuting based on that trip. Furthermore, respondents may have a secondary residence close to place of employment and might travel to their homes on weekends. Additionally, some flows are supressed if the number of respondents in the census are below a threshold (20). This data was the major source for calculating the probability of choice for each mode for travel between different origin destination pairs.

There is a total of 108 census tracts in the region of Waterloo, which were used to generate flowlines on ArcMap for possible trips between them. Thus, there were a total of 11,664 possible origin destination pairings. However, the 'Journey to Work' dataset had 7338 unique pairs; thus 4,326 records were removed from analysis. Furthermore, 28 datasets were removed as they had no reported trip data, and another 2139 observations were excluded as there were 0 trips between those origin – destination pairs. Lastly, 103 intra- OD pairs were removed due to the lack of travel time data. Effectively, the final dataset used to run the model had 5068 origin-destination pairs.

b. Travel Time

Travel time was calculated for each different mode by solving for an O-D cost matrix in the network analyst extension of ArcMap. All travel distances were calculated from the centroid of the origin census tract to the centroid of the destination census tract. For Auto as a driver mode mode, (CTV_D), the road network GIS dataset was utilized to calculate travel time based on posted speed limits on the roads. The vehicle was assumed to take the shortest distance from its origin to reach the destination. However, this approach did not account for congestion and signal delays that might be experienced by commuters during peak travel time. To account for these delays, the travel time on the shortest route was referenced from google maps for three times during the peak period (7:30 am, 8:30 am and 9:30 am). Such observations were made for 5 geographically spread-out destinations for all 108 census tracts, resulting in 540x3 observations. The results for different times during the peak period were averaged, 540 'real travel time' observations were obtained. On the basis of these, the calculated travel time by the shortest route was scaled up by 18% to enable depiction of actual travel time conditions during the peak hours. The same travel times were also used for motorcycles, scooters and moped (MSM) as they do not have any given advantages over cars during peak period. The travel time for CTV_P was obtained by increasing CTV_D by 5 minutes to account for travel logistics such as pick up and drop off. Analysis revealed that average travel time as CTV_P is 16 minutes, thus, 5 minutes is a reasonable assumption for travel logistics.

Active modes' travel times were calculated based on the assumption that these commuters take the same route as the auto vehicle users, as commuter pedestrian and cycling activity has not been mapped by the region. Additionally, pedestrian network too, is only available for the City of Cambridge, thus, this approach was followed to enable uniformity between origin-destination pairs. Furthermore, the times were calculated using an average speed of 30 km/hr (500m/min) for bicycles and 4.3 km/hr (1.2 m/s) for pedestrians based on Region of Waterloo's active transportation Masterplan. (2014)

Transit travel times were calculated based on GTFS (General Transit Feed Specification) data obtained from the region, using the Network Analyst Public Transit Data Model tool in ArcGIS Pro. The tool allows estimation of travel time by transit including the egress and access distances at any given time. The number of transfers, bus schedules, waiting time and routes are built into the tool, and it gives the most efficient method to travel by transit. For the purpose of this study, calculations were done for every 15 minutes from 7 am to 10 am, and then averaged to obtain transit travel time from the centroid of the origin census tract to that of the destination. It is important to note that a major drawback of this approach is that the transit travel time within the census tract could not be obtained and for those census tracts which do not have transit connectivity, the 'transit' travel time is largely the 'access' or walking time. This, however, has no impact on the question this study is trying to answer as the census tract pairs which have such issue, have predominantly auto based mode shares.

c. Travel Cost

Travel Cost emerged as an important variable in the review of previously developed mode choice models, but it was difficult to exactly calibrate this at an aggregate level, as commuting cost depends on many factors such as frequency of travel, bus pass ownership and vehicle age. Attempts were made to account for 'time cost' based on the hourly wage of the residents of the census tract which would be gauged from average household annual income, and then comparing it against the travel time. However, this approach was concluded to be misleading as annual income and travel time were included as independent variables and including 'time cost' was repetition of the same data, which had the potential to skew the overall results. Furthermore, LRT system has not levied new charges on transit users and would not be a major differential in the pre and post ION analysis. Travel time on the other hand seems to have witnessed a change in the post ION phase and was deemed to be more important to the study.

d. Network Density/ Intersection density

The correlation of built form with mode choice has been established in the literature review. Intersection density is an important aspect of built form that is associated with walkability and transit usage as high intersection density shortens access distances and hence provides more alternatives for transit users. (Ewing & Cervero, 2010) Intersection density was computed in ARCMap using the collect events tool, and all junctions where more than two different road segments joined were included, while cul-de-sacs were excluded from the overall analysis.

e. Average Vehicle Ownership

Vehicle Ownership data was obtained at the Traffic Analysis Zone level from the Transportation Tomorrow Survey (TTS). An overlay analysis was then done on ArcMap to estimate this variable at the census tract level by calculating weighted averages, depending on the area of intersection between the census tract and TTS zone.

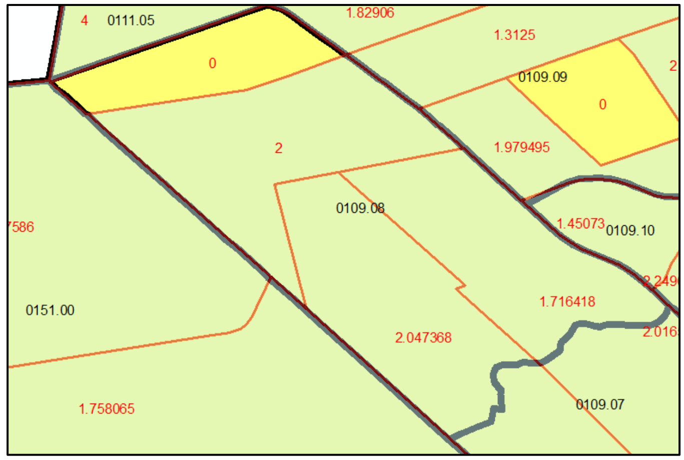


Figure 3.3 Overlay Analysis to calculate Vehicle Ownership

In the original data TTS dataset, many zones which intersected with census tracts had 0 vehicle ownership. Further investigation revealed that these were non-residential zones. It is thus, important to note that only residential area of the census tract was used to enumerate the vehicle ownership averages.

For example, in figure 3.3, CT 109.08 comprises of 4 TTS zones with average vehicular ownership 0,1.71, 2 and 2.04. Let area of TTS zone lying within the Census Tract be denoted by $A_{0,}\ A_{1.71,}\ A_{2,}\ A_{2.04,}$ respectively. The average vehicular ownership for the Census Tract was then calculated using the formula: $[\frac{0.71\ x\ A_{1.71}}{A_{1.71}+A_2+A_{2.04}}+\frac{2\ x\ A_2}{A_{1.71}+A_2+A_{2.04}}+\frac{2.04\ x\ A_{2.04}}{A_{1.71}+A_2+A_{2.04}}]$, instead of $[\frac{0.71\ x\ A_{1.71}}{A_0+A_{1.71}+A_2+A_{2.04}}+\frac{2\ x\ A_2}{A_0+A_{1.71}+A_2+A_{2.04}}+\frac{2.04\ x\ A_{2.04}}{A_0+A_{1.71}+A_2+A_{2.04}}]$.

f. Other Variables

Apart from the variables described above, which were majorly calibrated using GIS, many other variables such as living situation, educational qualification, marital status, gender, income, population density and employment density emerged as common variables in mode choice modelling. All these variables were obtained from Statistics Canada, which sources this data from the Census (2016). Since this study focuses on commuting behaviour, variable values for the population of working age (15 – 64 years), which has the potential to contribute in the labour force was included for modelling and analysis.

Living situation was included to accommodate household characteristics of families making mode choices. It was categorized into four groups – Married or common law partners with a child, Married or common law partners without a child, Single Parents and lastly individuals not living in census family. Since household size is not a census variable per se, these categories account for commuting mode choices as a function of their household characteristics, such as living with a partner, presence of a child, being a single parent and lastly living alone or in a non-census family environment.

Education qualification is of potential interest when modelling mode choice, as it is interesting to note if higher qualification increases environmental awareness and drives individuals towards transit. It was divided into six categories – no certificate or diploma (NOC), high school or equivalent certificate (HS), apprenticeship or trades diploma (APP), College or non-university certificate/diploma, University diploma below bachelor level and finally, University diploma/certificate/degree at bachelor level or above.

Marital Status and Gender were variables that were frequently included in studies to assess the influence of household structure and gender on mode choice. Gender was included as male and female ratios, while marital status was included for both genders. Lastly, median overall income values were used to examine the impact of income and affordability on mode choice behaviour as the hypothesis was higher income would lead to higher affordability, hence higher probability of vehicular ownership leading to greater share of commuter trips. Lastly, the population density, calibrated from census data, and employment density, calculated from 'Journey to Work' data were included to test how spatial distribution of workplaces and residences impact commuters' mode choice behaviour.

3.5 Model Specification and Estimation

The Literature review has established the superiority of Nested Logit Models for Choice or Mode Share Modelling. This estimation was carried out using NLOGIT 6 which allows up to four levels of nests and allows analysis of revealed (or proportional aggregate) data. This process included four steps – (1) Setting up data, (2) Specifying the Model (3) Simulating the Model and (4) Interpretation of Results. The fourth part of the process has been detailed in Chapter 4.

3.5.1 Setting Up the data

NLOGIT allows two methods of instruction – Dialog boxes and script-based command lines. This analysis was undertaken using Script based command lines as it offers greater flexibility. The Data has two types of variables – depend and Independent. The dependent variable fits into one field which in this case is the "choice" consisting of proportion of commuters travelling by the corresponding mode. The choice thus, had to be coded to fit into a single field, which required the number of observations per Origin-Destination pair to be equal to the number of modes that the model is being estimated for, which in this case was six (CTV_D, CTV_P, Bus, Walk, Cycle, MSM). The following code was used in excel to replicate the OD pairs 6 times:

```
Sub InsertRows()
Dim I As Long, J As Integer, Nb As Integer
For I = Range("A5200").End(xlUp).Row To 2 Step -1
Nb = 6
```

```
For J = 1 To Nb - 1
Rows(I + J).Insert xlDown
Rows(I).Copy
Rows(I + J).PasteSpecial '
Next
Next
Range("A1").Select
Application.CutCopyMode = False
End Sub
```

The Travel times and choices were then transposed using the "Transpose" Array formula in excel.

An example of the Resulting Input Data Can be Seen in Figure 3.4:

Modes	Cset	Choice	TT	VOwn	IntD	EMPD	Gen	MMCL	MNMCL	FMCL	FNMCL
1	6	0.5	2.6230118	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
2	6	0	7.6230118	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
3	6	0	18.58365	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
4	6	0.5	22.5774027	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
5	6	0	3.25114599	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
6	6	0	2.6230118	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
1	6	0	4.97594	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
2	6	1	9.97594	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
3	6	0	34.89973	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
4	6	0	45.9693794	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
5	6	0	6.61959064	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337
6	6	0	4.97594	1.7397713	14.1743914	291.36249	0.5111633	0.4436782	0.554023	0.4963855	0.5084337

Figure 3.4 Input Data Sample - Coding for Dependent Variable

The Modes were coded as – CTV_D (1), CTV_P (2), Bus (3), Walk (4), Bicycles (5) and finally MSM (6). Additionally, this model was estimated under the assumption of a fixed choice set, implying every mode is available to every commuter making the choice, which realistically might not be ideal, as individuals who do not own a bicycle, or car do not have those modes in his/her choice set. Similarly, commuters in areas not served by GRT, for example North Dumfries, do not have access to transit and thus it should not be part of the choice set. However, the aggregate nature of data made this distinction between OD pairs difficult and thus a Cset of 6 was used for all trips. Lastly, it is important to note that the Input data file was imported using CSV (Comma Delimited) format as excel files (.xls or .xlsx) are incompatible with NLOGIT.

3.5.2 Model Specification

Considering the significance of variables and the availability of data, this investigation focusses on select socio-economic and built environment variables using the application of the theory of utility-maximizing choice. Under this, individual's preferences determine potential mode choices under different scenarios. Model Specification is the process of testing statistical significance of various attributes included for analysis, and then selecting the ones that prove to be the best fit to answer the research question. This involves multiple model runs, gradually adding and testing for more variables with each run and eliminating non-significant attributes.

Different Specifications were tested both, nested logit and multinomial logit models to arrive at best specification. For the former, two nest structures were tested in the trials (Figure 3.5 and 3.6), however the three-step model was rejected due to very high IVs (expected between 0 and 1) within the sub-nest, high standard errors and insignificant Travel Time Coefficients. Additionally, MSM was excluded from analysis after the first few runs due to relatively small number of observations, which skewed the model results.

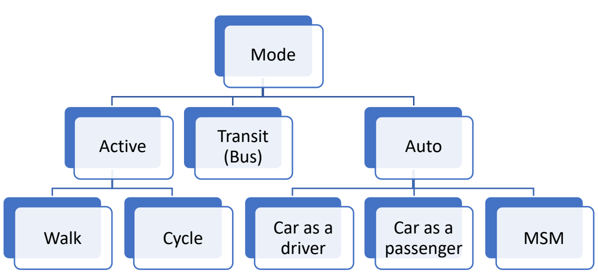


Figure 3.5: 2 Level Nest

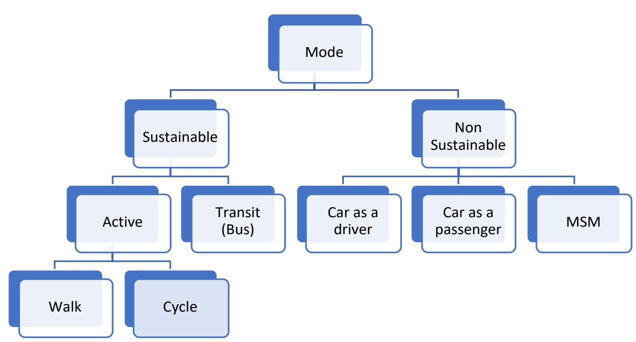


Figure 3.6:3 Level Nest

Different specifications were tested for the model until one which had significant coefficients for all variables was reached. The Script used to run the final model is provided below:

```
Nlogit
;Lhs=CHOICE

;Choices=CarD,CarP,Bus,Walk,Bike
;Tree= Private(CarD,CarP),PT(Bus),NMT(Walk,Bike)
;ivset:(Private,PT)
;start=logit
;Maxit=100

;Model:
U(CarD) =              + tt1*TT  /
U(CarP) = b_CarP + tt2*TT   /
U(Bus)  = c_Bus    + tt3*TT  + inc3*INCOME /
U(Walk) = d_Walk + tt4*TT  + vo4*VOWN + em4*EMPD /
U(Bike) = e_Bike $
```

Where;
TT – Travel Time
INCOME – Median Annual Household Income
VOWN – Average Vehicular Ownership per household
EMPD - Employment Density at the Origin Census Tract
b_CarP, c_Bus, d_Walk, e_Bike are Unobserved components of the Utility for respective modes and finally,
U(CarD), U(CarP), U(Bus), U(Walk) , U(Bike) are utilities associated with respective modes.

Below is the output obtained from this specification:

Table3.1: Results from Nested and corresponding Multinomial Logit Models

	Nested Logit		**Multinomial Logit**	
Parameters for dependent variables*	*Coefficients*	*p-value*	*Coefficients*	*p-value*
Car Truck or Van as a driver (CTV_D)*				
Travel Time	-0.02460	0.08	0.04105	0.00
Car Truck or Van as a Passenger (CTV_P)*				
Constant	-1.17406	0.00	-1.52364	0.00
Travel Time	-0.09	0.00	-0.02577	0.96
Transit*				
Constant	0.29820	0.18	0.42902	0.45
Travel Time	-0.01701	0.01	-0.01025	0.00
Income	-1.11704	0.01	-0.23902	0.00
Walk*				
Constant	0.22124	0.23	0.34748	0.51
Travel Time	-0.01610	0.00	-0.03142	0.00
Vehicle Ownership	-0.18839	0.07	-0.54714	0.08
Employment Density	0.00248	0.11	0.00939	0.61
Bicycles*				
Constant	-1.77806	0.00	-3.84499	0.00
IV Parameters				
Private	1.94			
Public Transportation	1.94			
Non-Motor Vehicles	4.04			
Log-Likelihood function	-2816.25071		-2856.04953	
*Dependent Variables				

Explanatory variables included in the above model (Travel Time, Income, Vehicular Ownership and Employment Density) are clearly significant, when tested at 85% confidence interval, i.e. p value is compared against alpha 0.15. Logistic regression coefficients for time for CarD, CarP, Bus, Walk, Cycle are negative, implying that a decrease in TT would increase the ridership by that mode. On the other hand, coefficient for employment density is positive, implying CTs with higher employment density tend to witness higher number of trips by walking. McFadden Pseudo R-Squared value is about 0.68, and the values leaning towards 1 represent models with higher accuracy in prediction. Additionally, the likelihood ratio of Chi Square (Significance level) is 0.00, below alpha (0.15), which rejects the null hypothesis that adding independent variables to the model has not significantly increased the ability to predict the decisions made. Therefore, it can be concluded that the model coefficients are statistically significant, implying that this model fits significantly better than an empty model or a constant only model.

This Nested logit model, however, failed to show relationship of mode choice with other socio-economic and Built form variables such as Marital Status, or presence of children in the household. Moreover, no variables seemed to be significant for the choice of bicycles, thus the utility for this mode was constant only. Additionally, the IV parameters for Nested logit model are above 1 (expected to be between 0 and 1), which implies that modes grouped together in the nest do not relate well to each other. Theoretically, IV <0 implies that an increase in the utility of an alternative in the nest, would decrease the expected maximum utility of that nest and hence decrease the probability of choosing that nest, while IV = 0 indicates that the alternatives grouped together do not relate to each other and the increase in utility would not affect the choice of that nest. Furthermore, IV>1 indicates that increase in utility of an alternative would not only increase the probability of selection of that mode, but all other modes in the nest. Lastly, IV parameters = 1 make the nested logit model equivalent to multinomial logit. For this reason, a multinomial model with the same specification was tested. The script used was:

```
Nlogit
;Lhs=CHOICE

;Choices=CarD,CarP,Bus,Walk,Bike
;start=logit

;Model:
U(CarD)  =            + tt1*TT  /
U(CarP)  = b_CarP + tt2*TT   /
U(Bus)    = c_Bus    + tt3*TT  + inc3*INCOME /
U(Walk) = d_Walk + tt4*TT  + vo4*VOWN + em4*EMPD /
U(Bike)  = e_Bike $
```

The output obtained from this specification is shown in table 3.1. The log likelihood values of both models remain comparable. Additionally, the p value, i.e. Prob (chi squared>value) continued to be significant beyond alpha = 0.05, meaning the multinomial model too, was statistically significant. Thus, as the various specifications of nested logit models failed to give IVs between 0 and 1, and the results from both models remained significant, it was concluded that despite general advantages of Nested Logit Models, the aggregate nature of data provides better and more confident results with the Multinomial Logit Model. Further to this, different specifications were tested for the multinomial logit model. The final MLM specification used for prediction is:

```
Nlogit
;Lhs=CHOICE
;Choices=CarD,CarP,Bus,Walk,Bike

;start=logit
;Maxit=100
;Model:
U(CarD) =               + tt1*TT + vo1*VOWN /
U(CarP) = b_CarP     + tt2*TT  /
U(Bus)  = c_Bus      + tt3*TT  + inc3*INCOME /
U(Walk) = d_Walk + tt4*TT  + em4*EMPD + po4*POPD + vo4*VOWN/
U(Bike) = e_Bike    + tt5*TT  $
```

Where;
TT – Travel Time
INCOME – Median Annual Household Income
VOWN – Average Vehicular Ownership per household
EMPD - Employment Density at the Origin Census Tract
b_CarP, c_Bus, d_Walk, e_Bike - Unobserved components of the Utility for respective modes
tt1, tt2, tt3, tt4, tt5, vo1, vo4, em4, po4, vo4 – Coefficients for respective variables for corresponding modes
and finally,
U(CarD), U(CarP), U(Bus), U(Walk), U(Bike) are utilities associated with respective modes.
The results obtained from this specification are in appendix figure 7

All but one variable coefficients (p values) in the above model output are statistically significant at alpha 0.15. Vehicle Ownership for Car, Truck or Van as a Driver is insignificant, but still retained as it is significantly different from Vehicle Ownership coefficient for walking. Furthermore, it provides explanation for difference in choice behaviour between the two modes. Prob (chi squared> value) continues to remain 0.00, which is below 0.05, hence implying the model variables are statistically significant. Although the previous multinomial logit model (figure 3.8) had greater log likelihood, the values and the standard errors remain comparable as well. Thus, due to the ability of incorporating more variables to explain choice behaviour, this model was selected for further analysis.

This model specification was further used to predict mode shares for 2018 and compared with the actual mode shares. (Figure 3.7) The model is deemed to be accurate if the sum of choice probabilities for the predicted mode share is the same as that of actual mode shares. The results in this case were fairly accurate, and the minor differences can be attributed to 41 bad observations in the sample.

	CTV_D	CTV_P	Bus	Walk	Cycle
Model	4228.57	310.44	328.91	110.39	48.69
Actual	4227.49	310.34	328.79	110.34	48.67

Figure 3.7 Comparison of Actual and Model Choice Probabilities

3.6 Simulating the Model

In 2019, Post LRT, the average transit travel time is calculated to decrease by 0.14% The Mode Shares for this change was simulated along with those for 10, 15 and 20% decrease in Transit Travel Times. The command used to execute the simulation is:

;Simulate; scenario:TT(bus) = [*]n where n = 1-Percentage/100; for example 20% decrease would be 1 – (20/100)

The results obtained from different scenarios are below:

Table3.2: Model Simulation Results

		0.14% Decrease	10% Decrease	15% Decrease	20% Decrease
Change in Mode Shares	CTV_D	-0.077	-0.832	-1.301	-1.810
	CTV_P	-0.006	-0.063	-0.098	-0.135
	Transit	+0.086	+0.926	+1.447	+2.010
	Walk	-0.002	-0.021	-0.032	-0.044
	Bicycle	-0.001	-0.010	-0.015	-0.021

The different simulation scenarios show an increase of 0.09 – 2% in the Transit Modal Share, with highest shift from Car Truck or Van as Drivers, followed closely by CTV_P. Overall, for the current estimated change in travel time, the shift towards transit is insignificant. The Interpretation and findings of the modelling exercise and simulation have been discussed in detail in the next chapter.

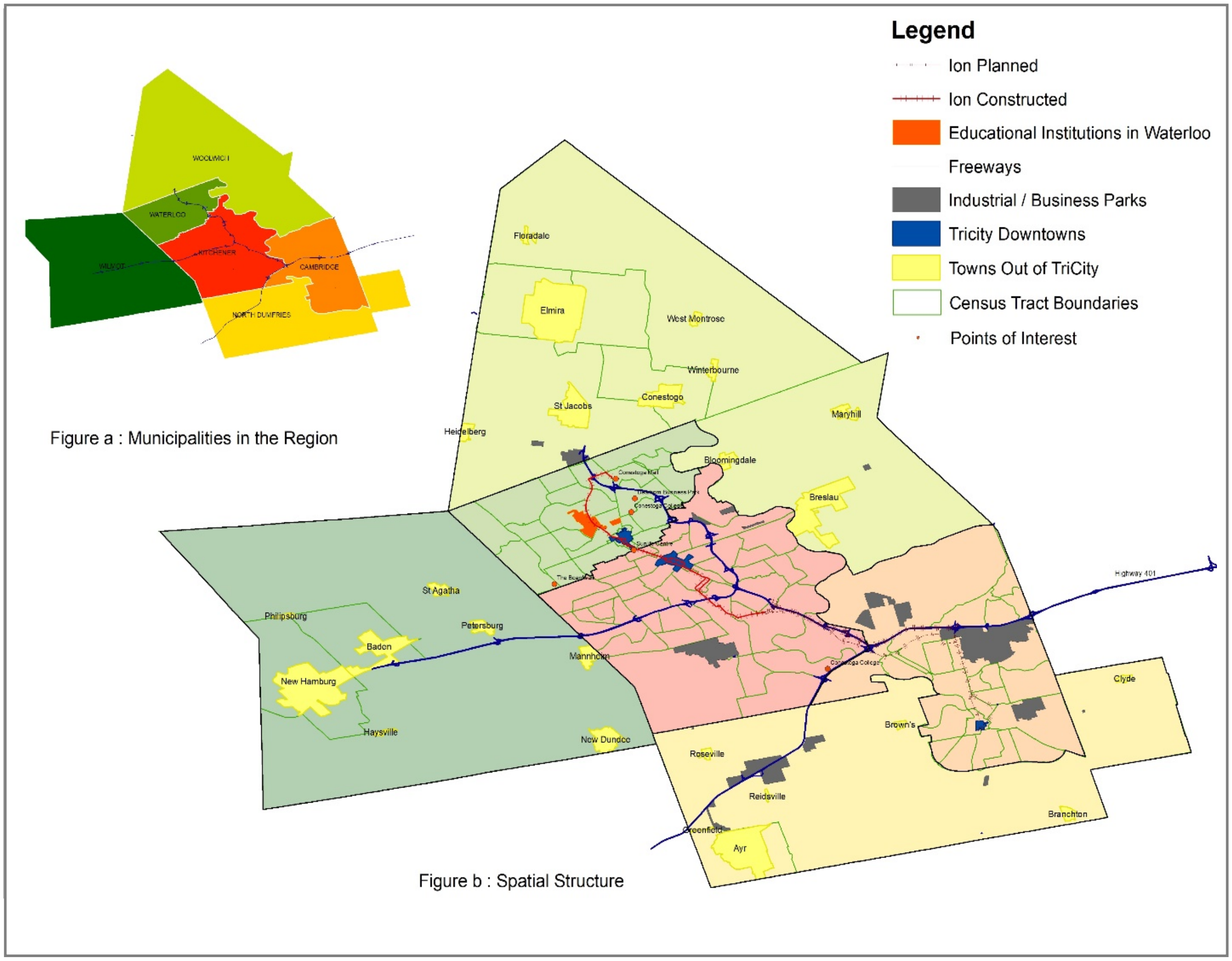
4 Findings and Discussion

The Region of Waterloo has seven municipalities, out of which one – Wellesley, does not have any designated Census Tracts. The other six – Woolwich, Wilmont, North Dumfries, Kitchener, Waterloo and Cambridge, are presented in Map 1 (Figure a). The ION route passes through Waterloo, Kitchener and Cambridge, hereafter referred to as the Tricity. The surrounding municipalities have urban populations concentrated in small towns such as Elmira, St Jacobs and Breslau in Woolwich, New Hamburg and Baden in Wilmont and Ayr in North Dumfries. (Map 1, Figure b).

There is an apparent variation in the urban fabric and landuses of these municipalities. Cambridge and Kitchener, for example, have designated industrial and business areas, which are expected to attract larger shares of commuter traffic, whereas Wilmont does not have such distinct Employment zones. Other industrial/business areas in the Region are along 401 in North Dumfries and on the edge of Woolwich. Waterloo has two major educational institutions – University of Waterloo (UW) and Wilfred Laurier University (WLU), with UW being the top public employer in the region with 5000+ employees.

The region is in intersected by Provincial Highway 401, which combined with Regional Highway 7/8 and 85 constitutes the highest hierarchy of road network in the region. The largest private employer in the Region is the Toyota Motor Manufacturing Plant, located in the Cambridge Industrial Area, near the intersection of 401 and Highway 8. Phase 1 of the ION which began operations on 21, 2019, begins from Conestoga Mall in Waterloo and connects UW, Uptown Waterloo, Downtown Kitchener and ends at Fairview Mall. The second phase of the project, which has currently not begun construction, will extend the network to Downtown

Cambridge.



Map 1

4.1 Current Travel Patterns

This section describes the current commuting patterns in the region of Waterloo with the aid of choropleth maps. The analysis provides an overview of the overall percentage of trip origins and destinations in region. The pattern is further detailed by examining the employment destinations associated with the top four commuter origin generating Census tracts, and then repeating the process for the top four employment destinations. The rest of the section is organized based on three modes – active transportation, including cycling and walking; transit, including buses and lastly auto, constituting of trips by Car Truck or Van as Drivers (CTV_D), Car Truck or Van as Passengers (CTV_D) and Motorcycle, Moped or Scooters (MSM).

Analysis for each mode includes examining the percentage share of that mode for all trips originating or destining in each CT. This provides an overall picture, showing how mode shares are distributed in different CTs, and the dominant mode of commuting from and to these areas. Flow analysis is then conducted for the four Census Tracts which have the highest number of trips originating and destining by the mode under study in the region. This helps in understanding the travel patterns between trip origins and destinations, facilitating the recognition of areas which generate traffic towards a destination. It is important to note that, the flow analysis has been conducted based on the ***number*** of trips a CT obtains and generates and not the modal shares, as it provides a better picture of how the traffic flows by different modes, for example, Downtown Kitchener as an employment destination does not have the largest commuter active transport modal share, despite receiving the maximum *number* of active transportation trips. Lastly, there is an attempt to explain these patterns based on spatial variables such as intersection density, Income and household vehicle ownership.

Overall, this analysis consists of 187,425 trips, out of which CTV_D occupies 78.5% share with 147,140 trips, followed by CTV_P at 7.2% (13,555 trips), Transit at 6.8% (12,835), Walking at 5.3% (9,850), Cycling at 1.5% (2,575), MSM at 0.1% (200 trips) and lastly, Other Methods at 0.7% (1,270) trips. Additionally, weighted averages for travel times indicate that average CTV_D trip is about 15 minutes, CTV_P is 16 minutes, transit is 49.29 minutes, walk a surprising, 43 minutes, while that for cycling is 10 minutes. The median travel time for walking, however, is about 30 minutes. (Figure 4.1). The unexpectedly high travel times for walking can be attributed to the assumption made in calculation, that pedestrians follow the same routes as other modes, while in reality pedestrian tend to make way through the shortest route, even if it undesignated.

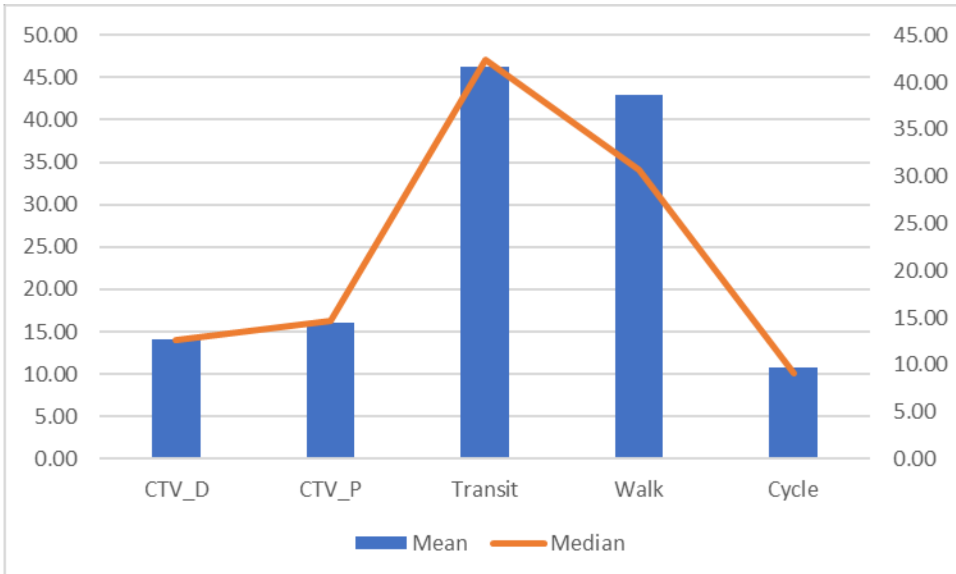


Figure 4.1: Mean and Median Travel Times (Minutes)

4.1.1 Overall Distribution of Trips

a. Trip Origins – Overall Distribution

Auto based modes are dominant among commuters in the region, with over 85% of all trips being undertaken by cars, while transit and active transportation occupy about 6% of modal share each. (Figure 4.2).

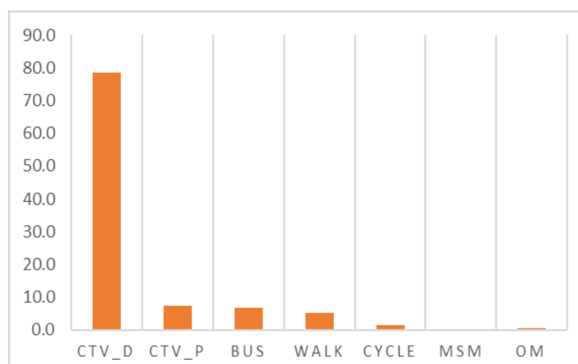
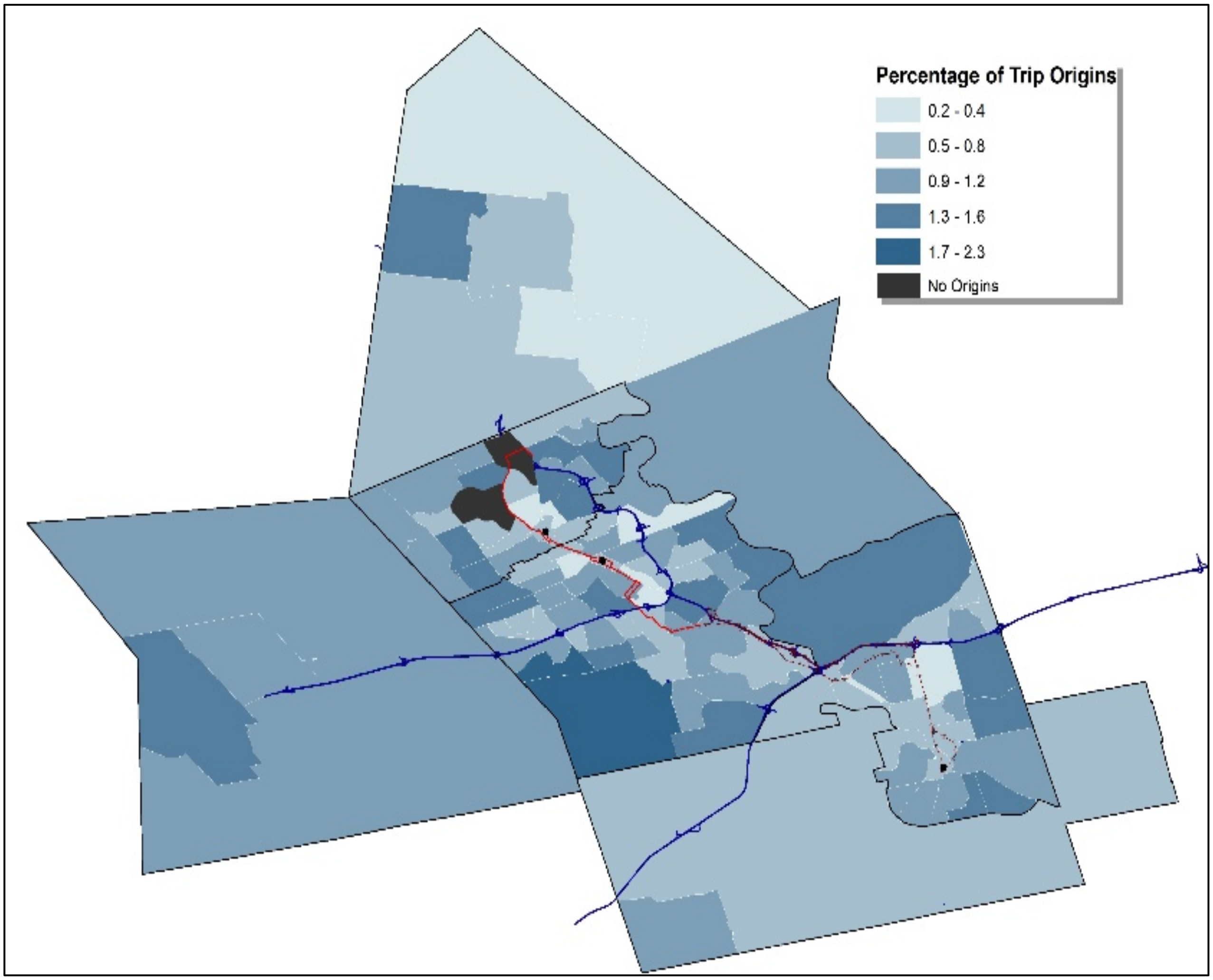


Figure 4.2: Modal Shares of Trip Origins

The overall distribution of trip origins is fairly dispersed across all CTs, with individual origin percentages of total commuting trips ranging from 0.2 to 2.4%. Census Tract 101.02, comprising primarily of the University of Waterloo, and 106.03, which is commercial in nature have no trip origins associated with them. Trip Origin proportions are comparatively higher in and around distinct employment centres, such as the Kitchener and Cambridge Industrial areas, implying commuters choose to live close to employment hotspots.

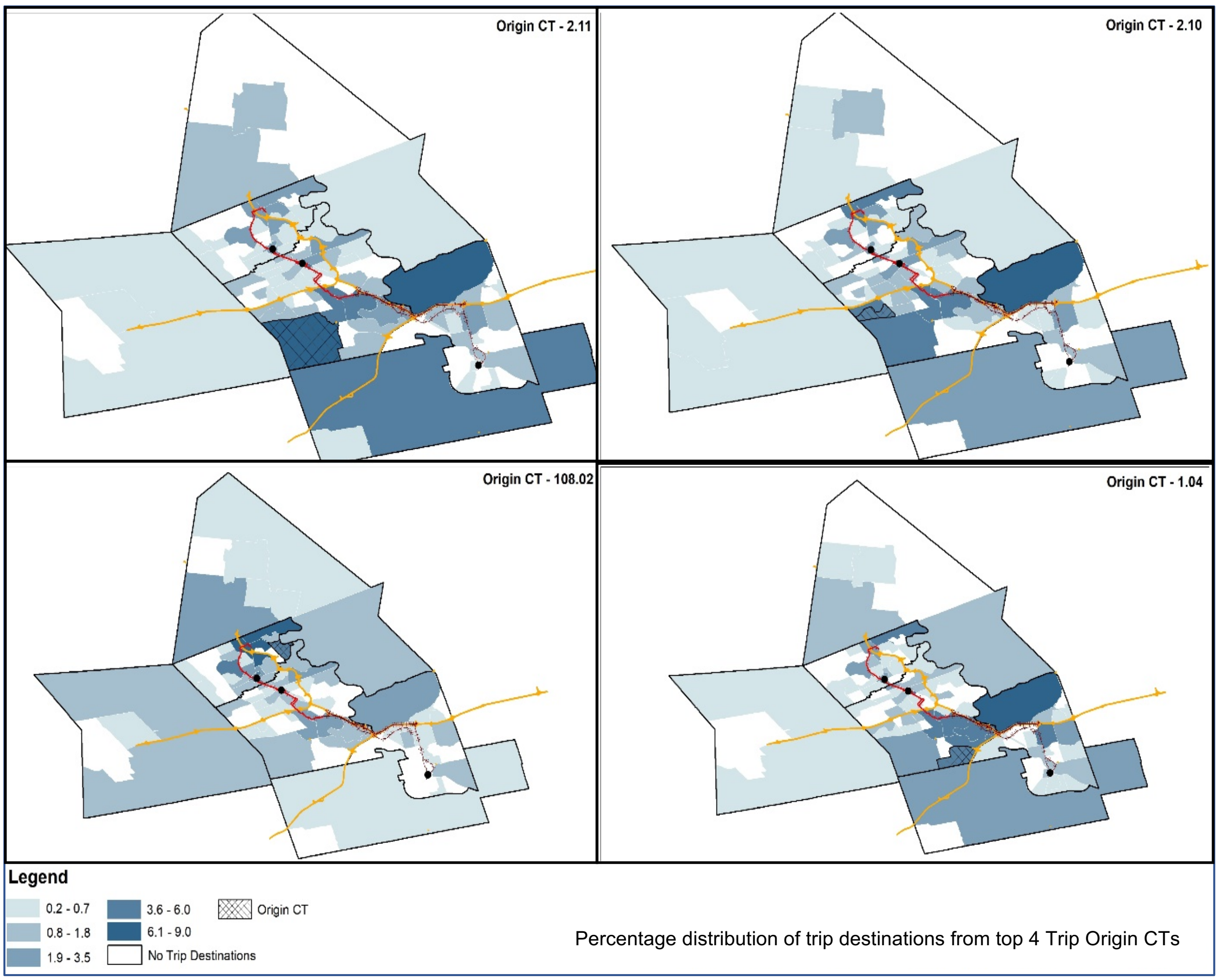


Map 2

b. Flow Analysis of the Top 4 Trip Origins

Flow analysis undertaken to study the destinations associated with the Census Tracts which are the top four trip generators in the region, reveals that although there are clearly some areas which attract more commuters, such as the Industrial areas in Cambridge, Kitchener and around the 401, overall the trip destinations are dispersed throughout the region. The Top 4 Trip Origin

CTs do not witness trips towards New Hamburg, eastern Woolwich and other CTs in the region which are primarily residential. These areas do not have specialized employment opportunities, which explains this pattern.



Map 3

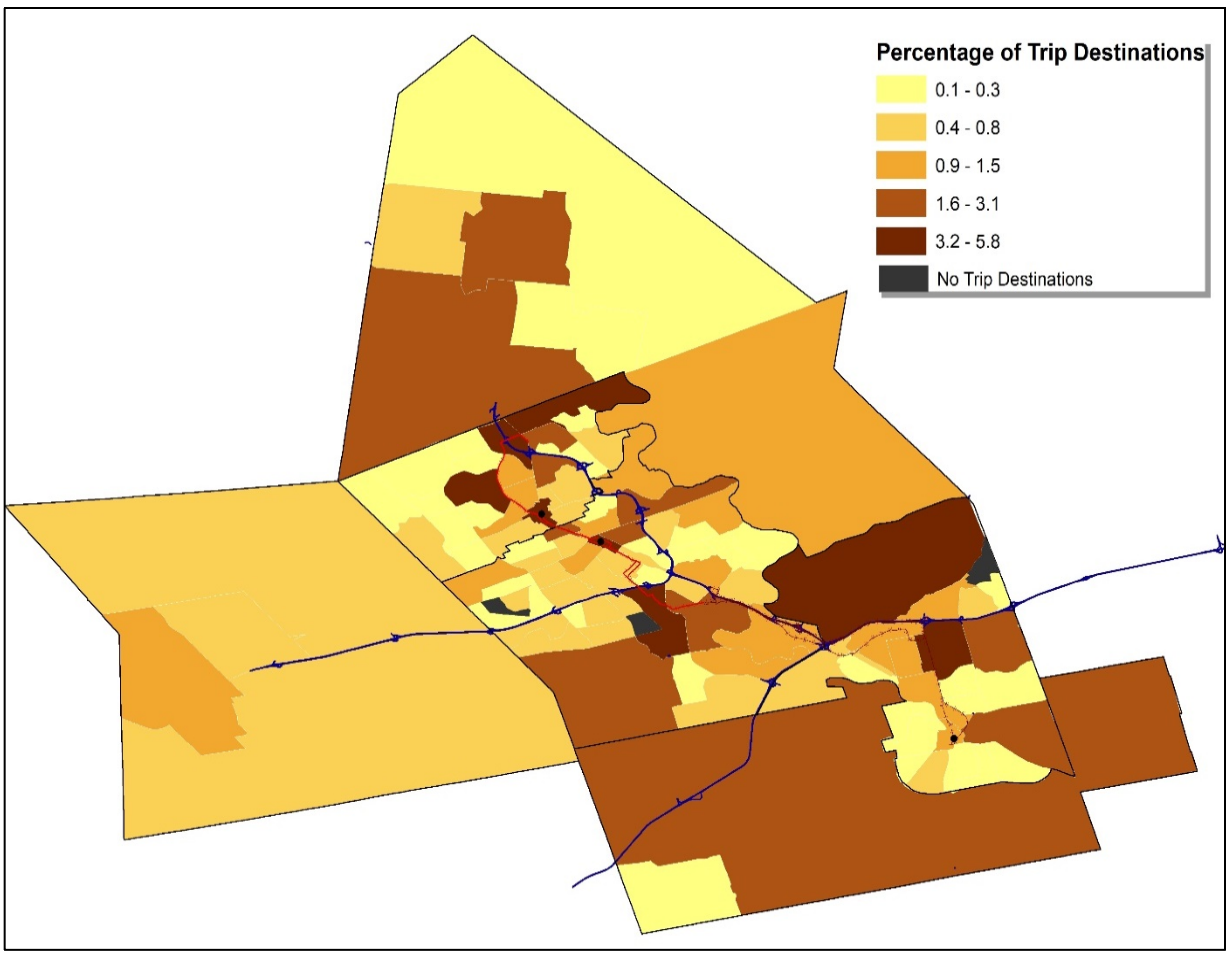
Three of the top four trip origin CTs are in Kitchener, and have destinations toward

Three of the top four trip origin CTs are in Kitchener, and have destinations towards industrial areas in Cambridge, Kitchener, along 401 and a fair percentage towards the downtowns and UW (3.5 – 6%). Additionally, CT 108.02 in Waterloo, the third most commuter trip generating

CT in the region, has trips destined towards UW, Dearborn Business Park and the commercial areas near Regional Highway 85. This is the expected pattern due to higher employment and business activity in these zones. Lastly, this distribution pattern shows that the top 4 origin CTs do not have accessibility to destinations through the ION.

c. Trip Destinations – Overall Distribution

Trip Destinations, overall, are relatively more concentrated, with Kitchener Downtown, Uptown Waterloo, University of Waterloo and Cambridge Industrial Areas receiving the dominant number of trips, and industrial/ business areas of Woolwich and North Dumfries following on a close second. It aligns with the expected commuter travel patterns as these are high employment



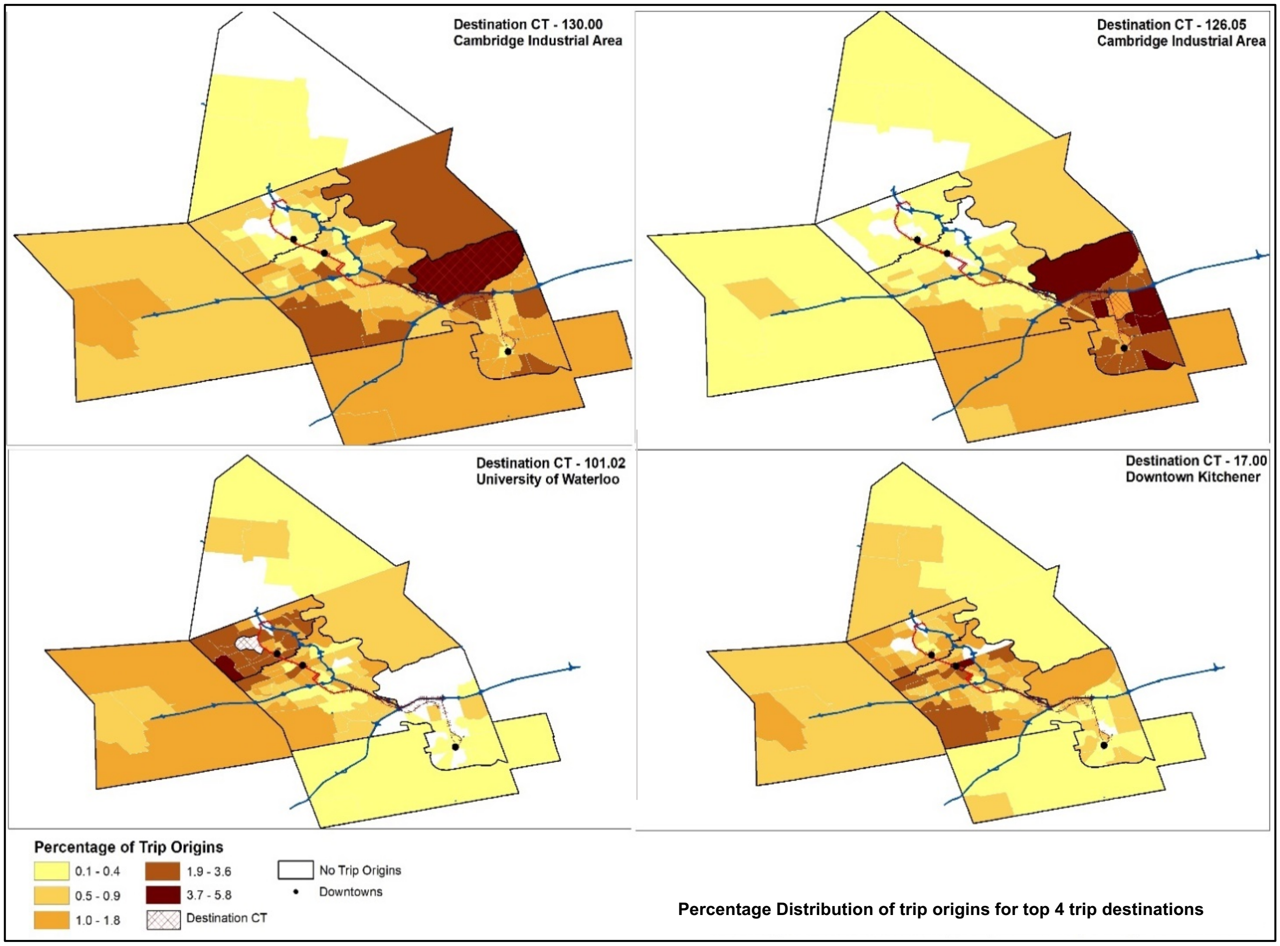
Map 4

zones. The primarily residential areas have high trip origins and low trip destinations. It is also interesting to note that while the census tracts abutting the LRT route fall under moderate range of trip origins, the trip destinations are concentrated around and connected by Ion.

The three census tracts which do not have any trips destinations associated with them primarily consist of green areas and some surrounding residential land, for instance, Census Tract 131.05 in Cambridge, is vacant green land, 2.04 in Kitchener consists of Country Hills Park, and 8.05 in Waterloo comprises of Westheights Park.

d. Flow Analysis of the Top 4 Trip Destinations

The top 4 trip Destinations in order are the industrial areas of Cambridge, University of Waterloo and Downtown Kitchener, which all lie on the ION (functioning or planned) route. These locations collectively are destinations for about 20% of trips which originate in the region. CT 130.00, where the Toyota Plant is located in Cambridge, receives 5.77% of commuters, out of which the highest percentage (4.7%) originate in the same CT. It also receives the second highest share from Breslau and the residential areas around the Kitchener Industrial areas. However, these locations do not have access to the Toyota Plant through the LRT. Trip Origins to UW, are concentrated in the census tracts surrounding the University and the pattern is similar in downtown Kitchener. However, the percentage of trips originating from these CTs maxes out at 5.8%. This suggests that even though ION may not capture these trips, it connects the moderate trip origin zones to these destinations and has the potential of increasing transit ridership from these zones.



Map 5

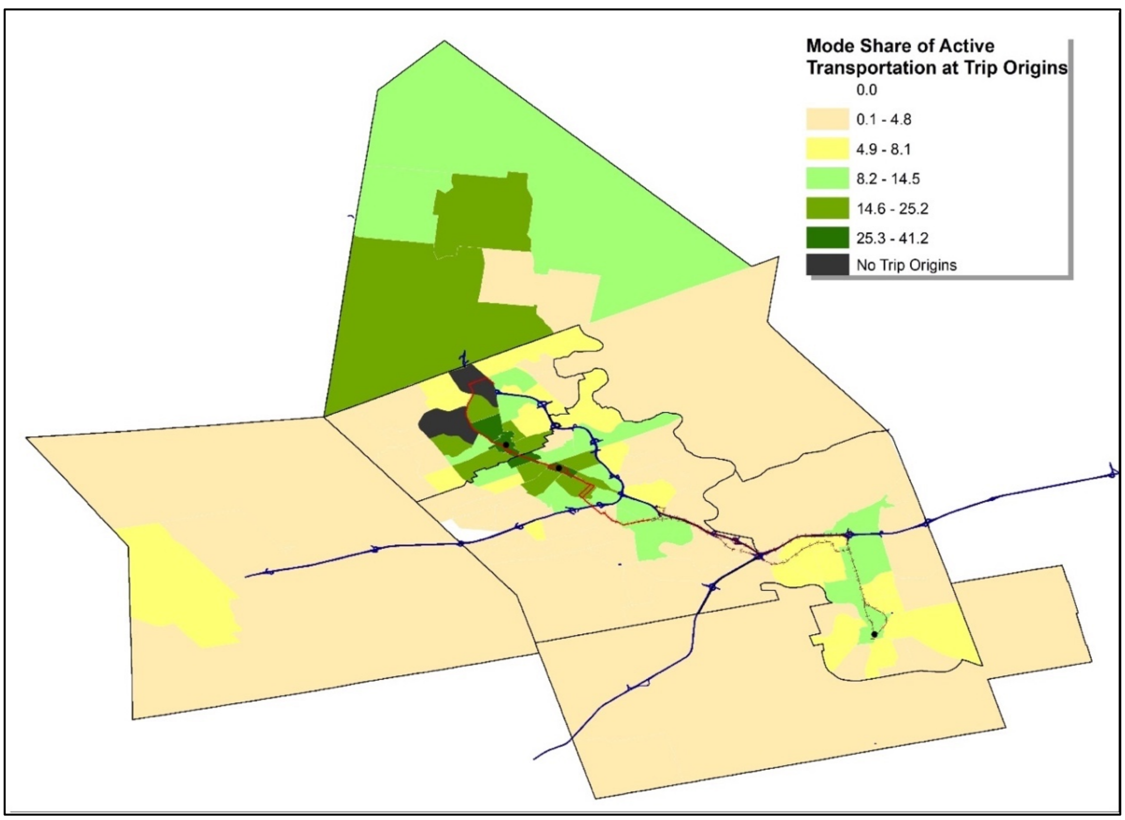
4.2 Distribution Analysis by Mode

The following section analyses the modal share of active transportation (Walk, Cycle), Driving (CTV_D, CTV_P, MSM) and Transit (Bus) over the census tracts in the region. These patterns have been assessed based on the characteristics associated with high concentrations of these modes which emerged in the literature review.

4.2.1 Active Transportation

a. Trip Origin

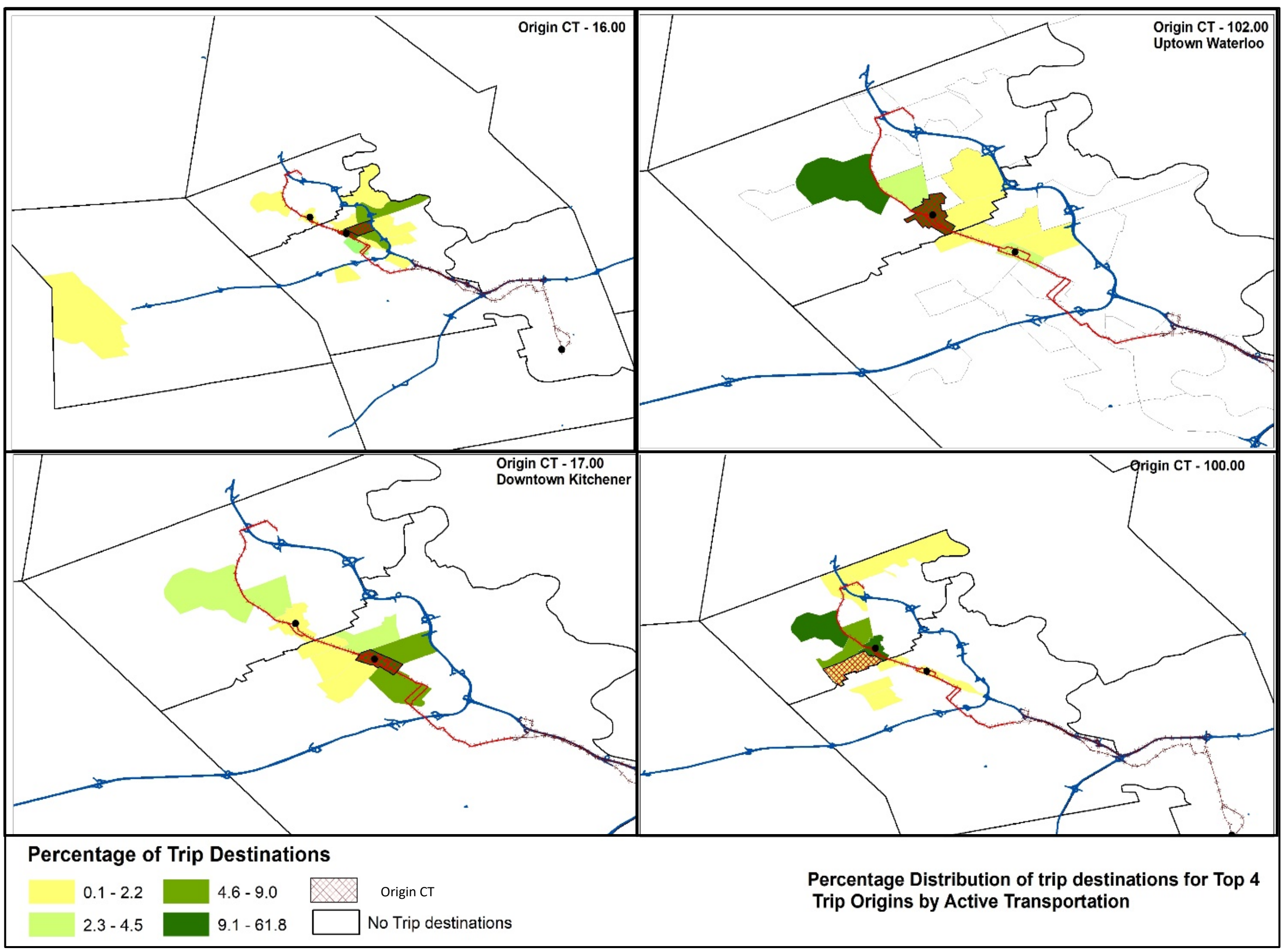
Map 6 represents the percentage share of active transportation commuting trips originating in a CT. Trip Origins by Active Transportation are concentrated around the downtown areas of Kitchener, Waterloo and the ION route. Woolwich surprisingly has higher active transportation activity in Elmira and St Jacobs townships, while the other suburban areas have low active transportation activity; Cambridge Industrial area despite attracting a large number of trips, has a low share (8.2 – 14.5%) of trip origins by foot or on cycle. This suggests commuters live in and around the industrial areas do not resort to active transportation for travelling to work. Downtown Cambridge too, has lower active transportation trip origins compared to DT Kitchener and Waterloo.



Map 6

b. Flow Analysis of the Top 4 Trip Origins by Active Transportation

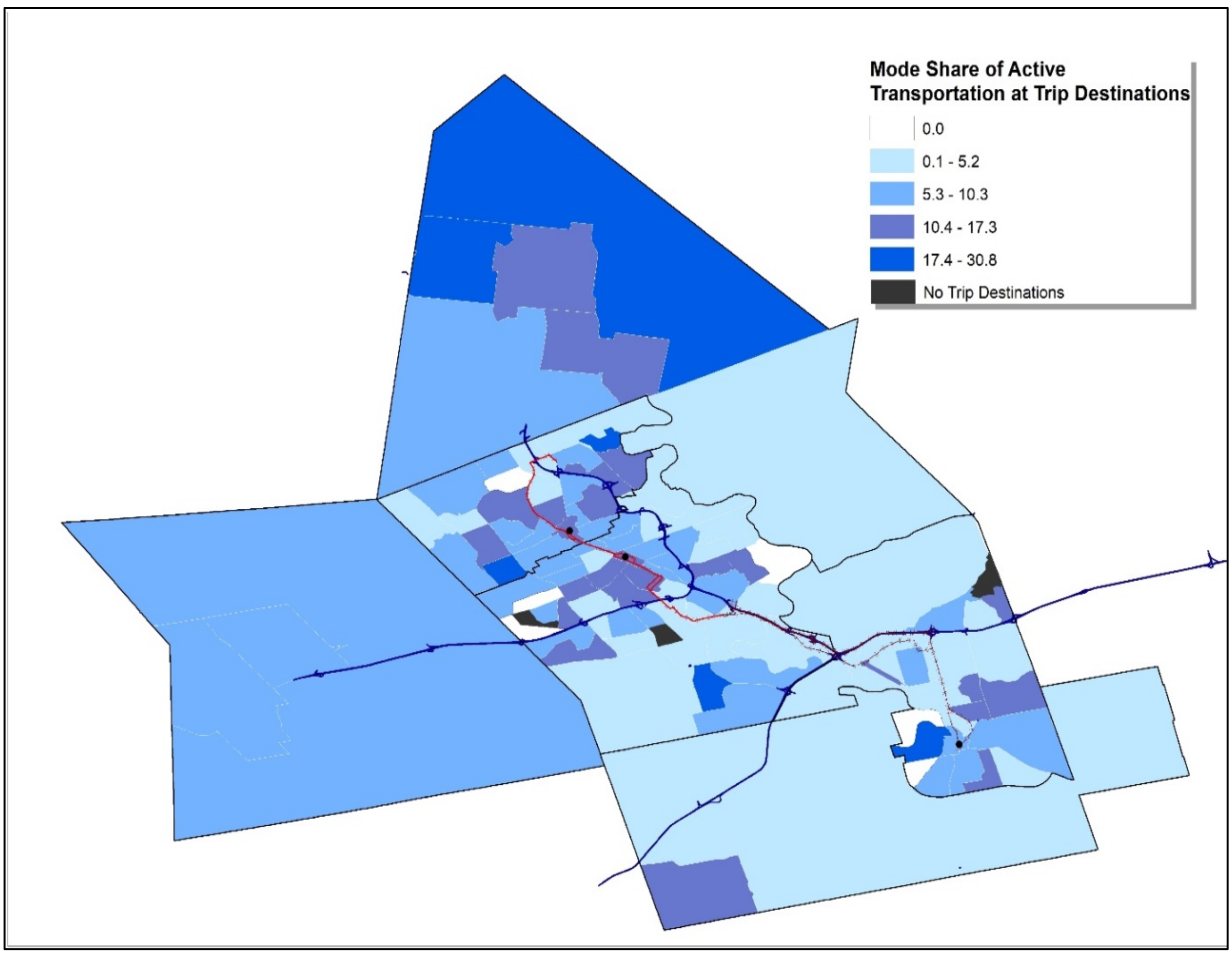
Flow Analysis of the top 4 Active Transportation Trip Origins seconds the observations made in the active transportation mode share analysis. The active transport trip origins are located on the LRT route, and the destinations too largely remain along that stretch. There is thus a potential of shift of active transit trips to LRT. It is surprizing to see active transport activity between New Hamburg and CT 16.00, which is next to Downtown Kitchener and has public service employment such Service Ontario and Passport Canada offices, but only 10 trips were reported, which were all through bicycles.



Map 7

c. Trip Destinations – Active Transportation

Active transportation mode shares at Destinations are highest in areas which are not distinct employment zones, but residential with supporting services. The destinations on the ION route, for example, the downtown areas of Kitchener, Waterloo, UW and Cambridge Industrial area near DT, fall in the second highest range. A clear relationship does not emerge between origins and destinations as all areas which had higher number of origins do not necessarily have higher number of destinations.



Map 8

Furthermore, walking dominates over cycling in all census Tracts, and cycling activity is concentrated either in and around the tricity cores, around the Downtowns or in smaller surrounding townships. (Appendix, Figure 1).

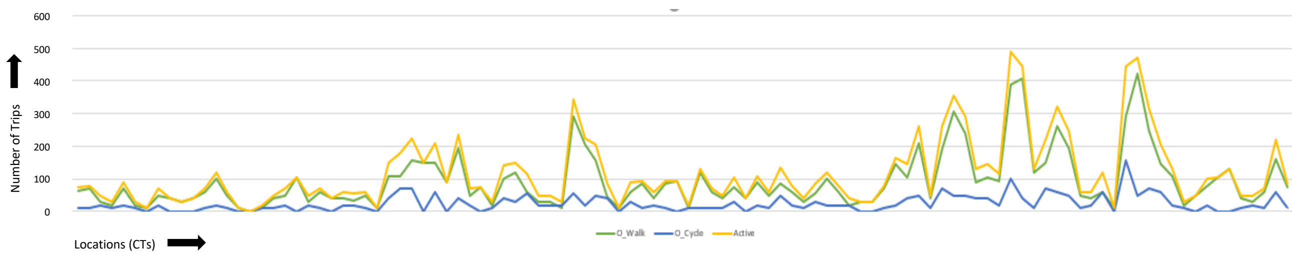
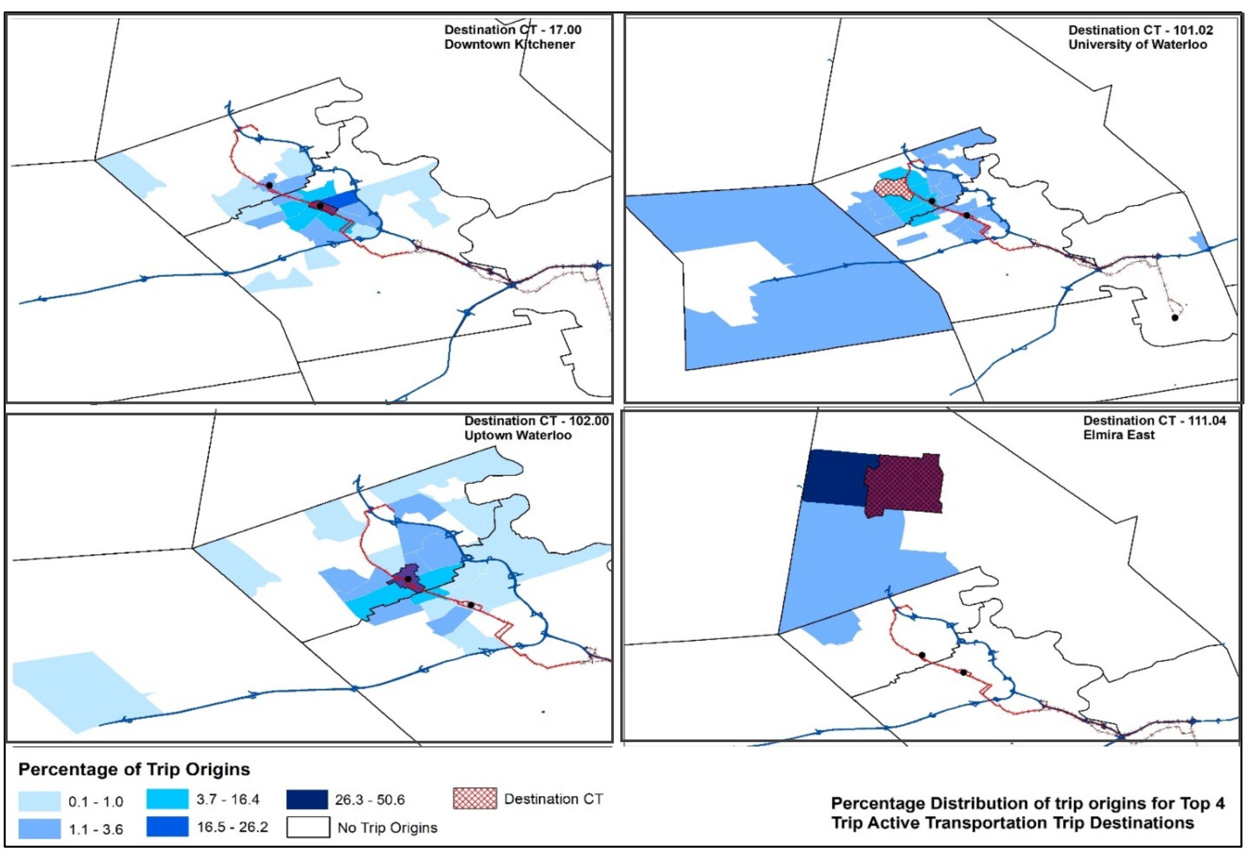


Figure 4.3 Number of Walking and Cycling trips in Active Transportation Activity

d. Flow Analysis of the Top 4 Trip Destinations by Active Transportation

The top four active transport destinations in order are DT Kitchener, UW, Uptown Waterloo and Elmira East. University of Waterloo has trip origins from Wilmont's smaller townships like



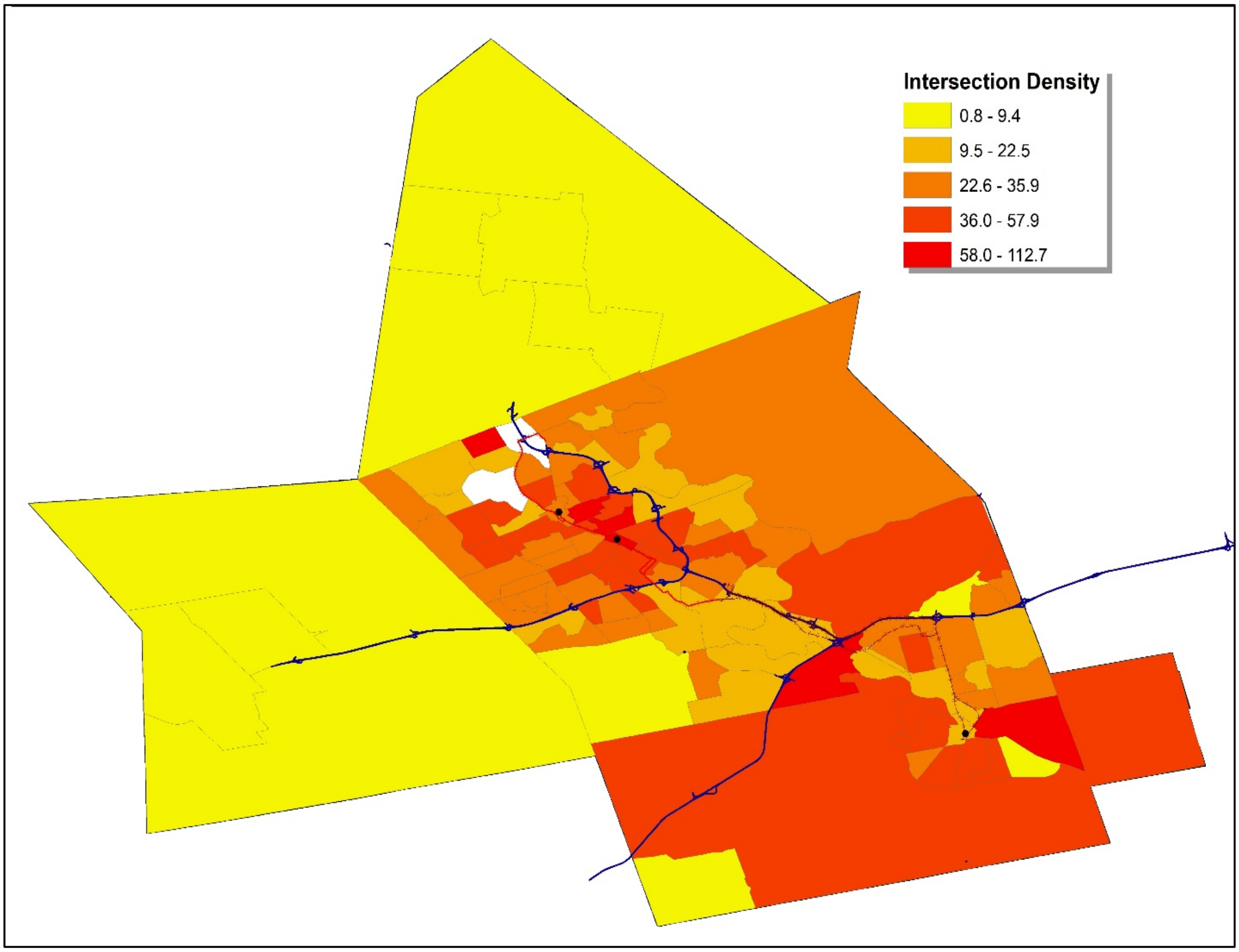
Map 9

Mannheim and Petersburg, which is surprizing, considering the distance from the destination. The trip origins for DT Kitchener, University of Waterloo and Uptown are dispersed, with higher percentages concentrated at shorter distances from the destinations.

93 percent of Woolwich's active transportation Activity is concentrated in Elmira (93%), with 43% trips within the destination Census Tract, and 50% in the abutting CT. About 2% of the trips originate from St Jacob's and Heidelberg. This pattern is a result of low intersection density overall, (Map 10), while higher expected intersection density within Elmira township.

e. Active Transportation and Intersection Density

Active Transportation patterns are consistent with measures of walkability explored in this study, as the CTs with higher trip origins and destinations also have higher intersection density. (Map 10). It is however interesting to note that that Woolwich has relatively high active transportation modal shares despite low Intersection density. This can be attributed to the fact that intersection densitities are calculated for the complete geographical area in census tract, while the cycling activity might be concentrated in the small townships which fall in these CTs.



Map 10

4.2.2 Transit

a. Trip Origins: Transit Occupies 6.8% of the total commuter travel activity. Trip Origins by transit are, like active transportation, concentrated in the central corridor in the Tricity. There is negligible commuter transit usage in the surrounding municipalities of Woolwich, North Dumfries and Wilmont. Analysis of Grand River Transit (GRT) routes and bus stops (Table 4.1) reveals that a probable reason for this pattern is the low level of service in these

Table 4.1: Number of Bus Stops in Municipalities

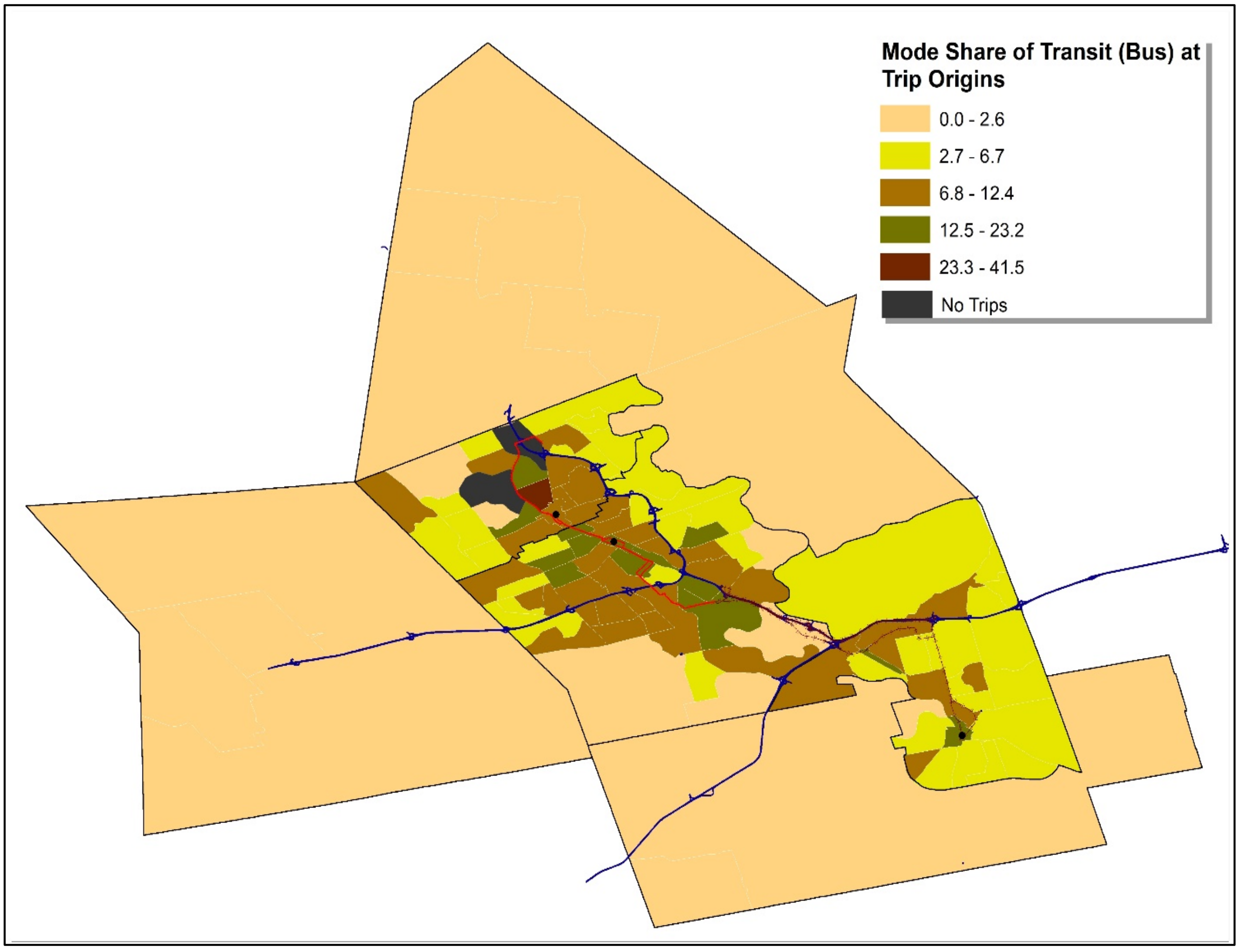
Municipality	Number of Bus Stops
Cambridge	737
Kitchener	1255
Waterloo	666
Wilmot	40
Woolwich	54

Source: RoW Open Data

municipalities compared to the Tricity. (Appendix, Figure 2). The highest trip Origins by transit are around the University of Waterloo and Wilfred Laurier University. The census tracts around the ION route on either side fall in the high to moderate range of commuter transit usage. This can be attributed to high frequency and multiple route options being available in the central corridor. Thus, there is a potential of smooth transition of bus riders to ION in an integrated network as the route already has a high transit ridership base.

b. Flow Analysis of the Top 4 Trip Origins by Transit

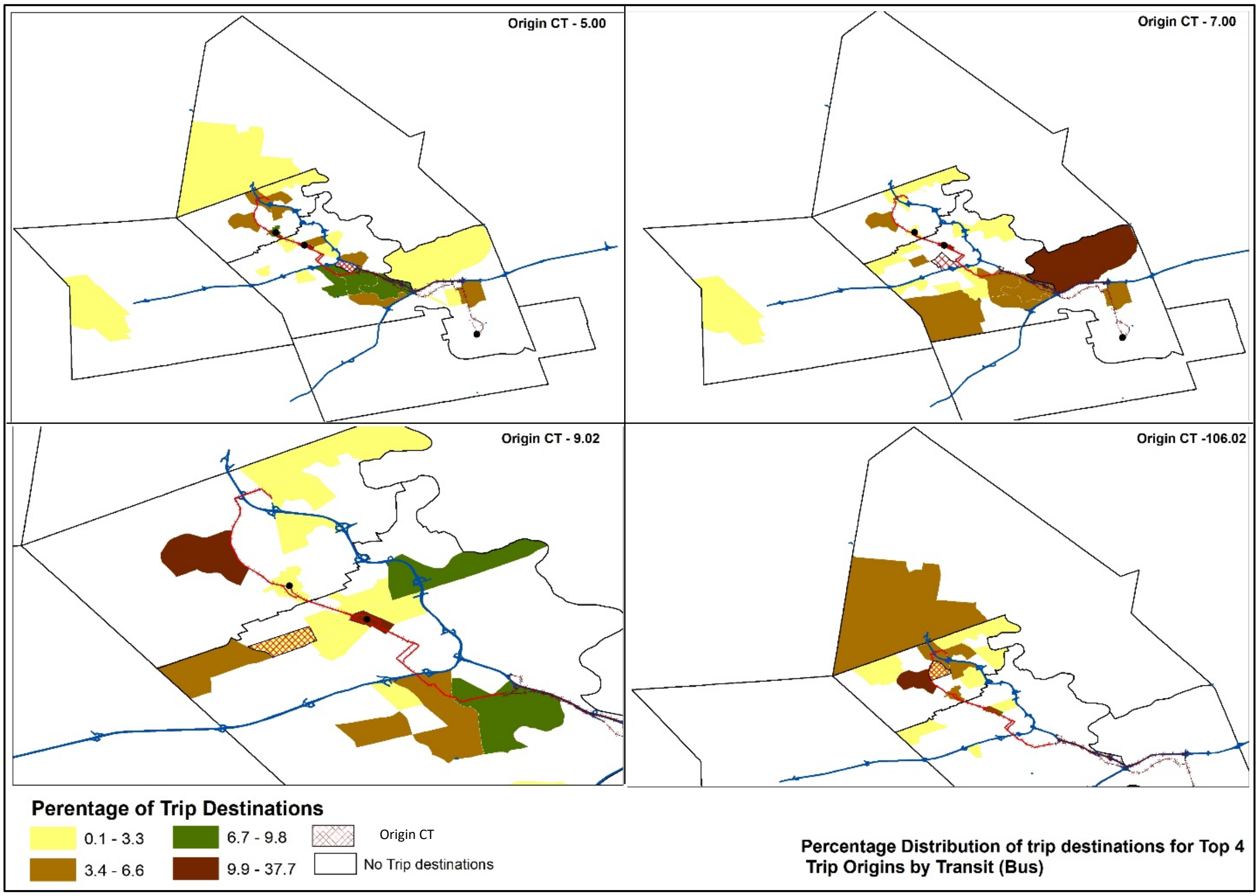
Map 11 presents the distribution of trip destinations from the top four transit trip origin census tracts. The Origins and Destinations lie either directly on the LRT route, or in high bus transit service area, indicating that commuters who currently use bus can easily transition to the LRT. It is interesting to note that the Toyota Plant in Cambridge industrial area has no trip destinations by transit from CT 9.02 and 106.02, despite being one of the biggest employers in the region. This suggests the need for better transit connectivity between these areas. The University of Waterloo remains one of the top destinations from all the four trip origins.



Map 11

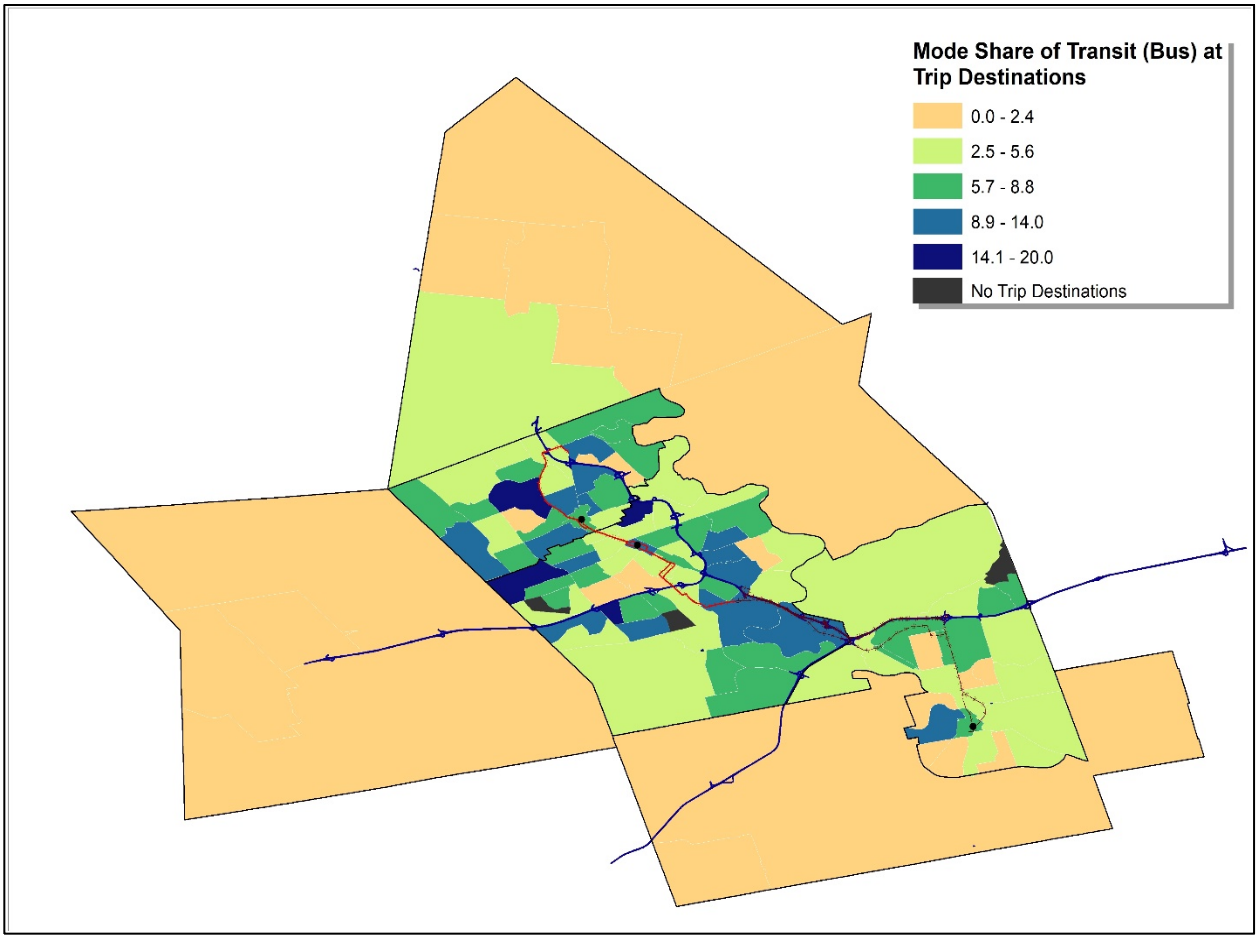
c. Trip Destinations:

Transit destinations too, like origins, are higher in the tricity which has higher frequency and better transit network. University of Waterloo has about 20% transit destination mode share, one of the highest in the region. However, not all census tracts around the ION corridor have high transit modal shares at destinations, despite a high level of service. It would be interesting



Map 12

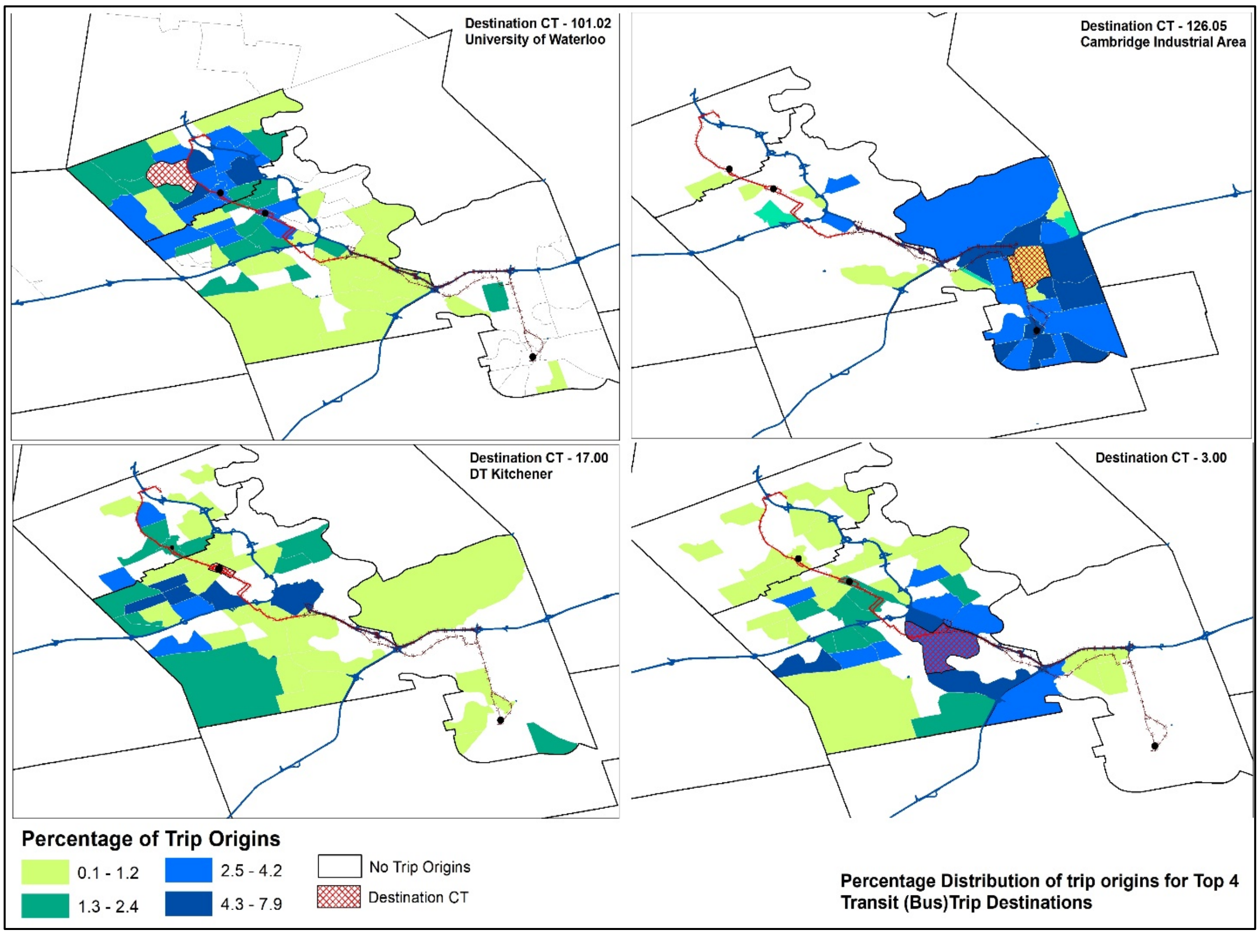
to undertake an analysis to see spatial change in transit destinations after the introduction of ION. Additionally, this pattern highlights the impact of high density development on transit usage, as the municipalities which have low transit ridership also have lower density, making them less serviceable by transit due to higher costs associated with this kind of development.



Map 13

d. Flow Analysis of the Top 4 Trip Destinations by Transit

Flow Analysis from the top four transit trip destinations follows a similar pattern to active transportation, although the trip origins are more geographically spread out. It is interesting to note that while active transportation trip destinations tend to attract intra CT trips, transit is majorly used for inter CT trips. This suggests that shorter commuting distances are more likely to encourage people towards active transportation. Additionally, the areas which have high number of trip origins destined for top 4 destinations, all lie in the tricity which have better connectivity compared to the peripheral townships. Thus, improvement in transit frequency and infrastructure can pull more people away from cars, towards transit. ION is scheduled to run every 8 – 10 minutes during the peak hours and has been seamlessly integrated into the existing transit network. However, not all CTs which generate higher percentage of trip origins towards the CTs which have high employment, have access to LRT or lie on the transit route. Thus, ION is not likely to impact the current transit users' mode choice behaviour.

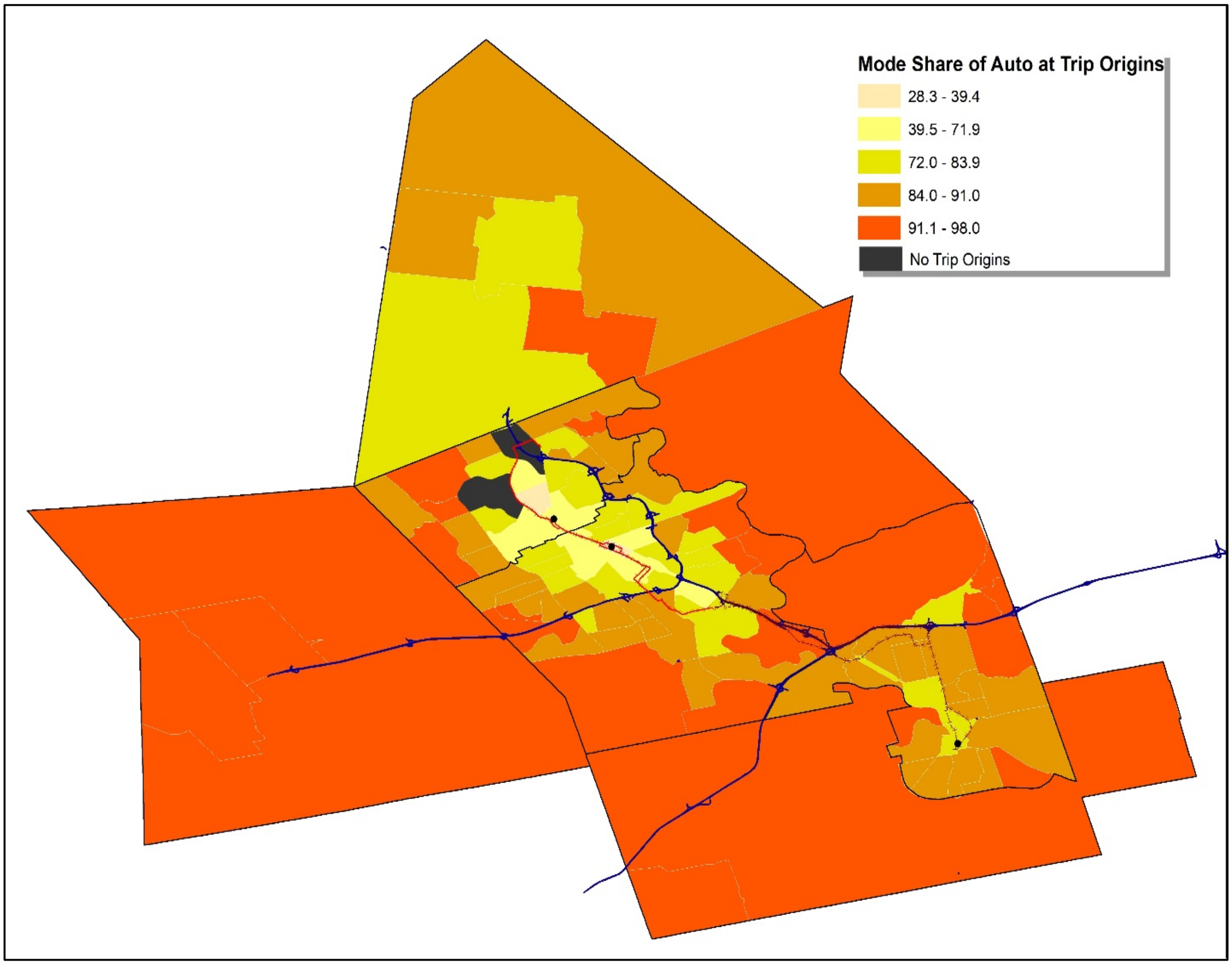


Map 14

4.2.3 Auto

a. Trip Origins:

Auto, incorporating CTV_D, CTV_P and MSM, dominates the overall mode share in the region, with over 85% trips. Mode Shares of driving at origins are highest in suburban areas which have low transit service and low intersection density. The central corridor, along which the ION is planned, has least number of auto trip origins, which still ranges between 28% – 40%.(Map 15) Ion has potential to convert these trips from auto to transit. However, it is not currently accessible to areas which have more than 90% auto modal shares. It is evident that as the distance from the central cores of the tricity downtowns increases, the percentage of auto trips increases. This pattern suggests a combined effect of lack of multiple mode options in the peripheral areas and increased commute distances. This hypothesis will be tested in the next section by assessing the destinations from top four origins by transit.



Map 15

Among the different auto modes, CTV_D (Car, Truck or Van as a Driver) wins over the other two modes by controlling 91.5% of the total auto share. (Figure 4.4). The share of Motorcycles, Scooter and Mopeds (MSM) for commuting is negligible, (0.1%), and that for CTV_P is very low at 8.5%, indicating that carpooling and car sharing are not popular among commuters in the region. It is interesting to note that CTV_P mode shares are higher in the census tracts on the Periphery of tricity. Flow Analysis from these Census Tracts reveals that most of these trips were longer distance, intercity, or to employment zones such as the University of Waterloo and

Cambridge Industrial Area. (Appendix, Figure 3 & 4) Thus, commuters are more likely to car pool to travel longer distances.

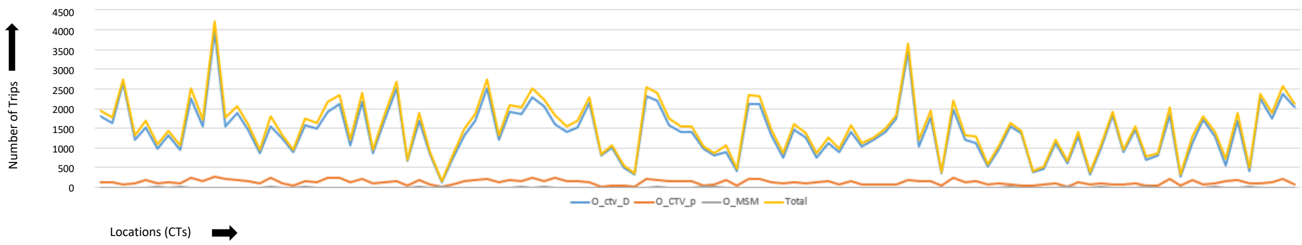
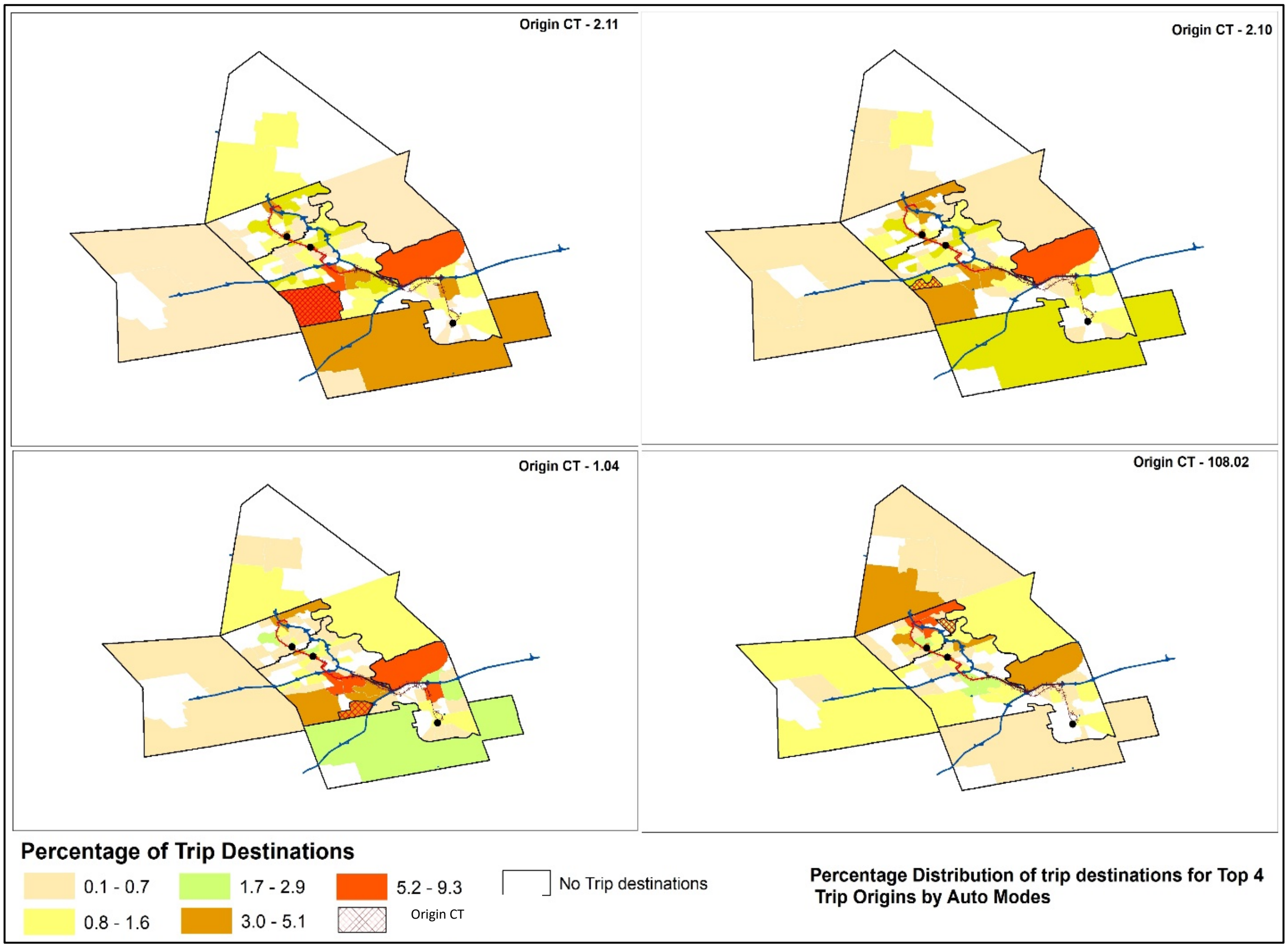


Figure 4.4: Number of Trips by different auto modes in total auto mode share

b. Flow Analysis of the Top 4 Trip Origins by Auto

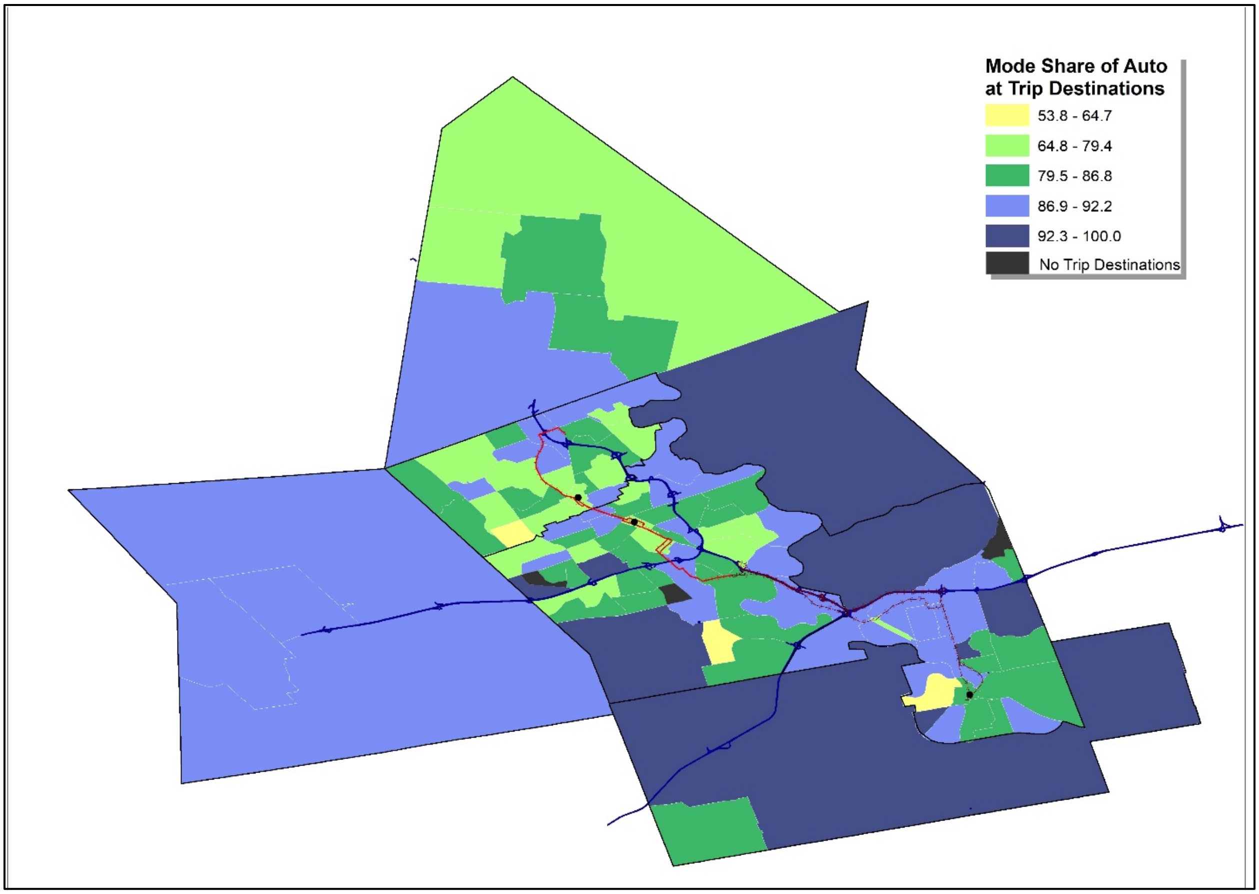
The top for trip origins by auto are primarily residential areas on the peripheries of the cities. The destinations from these trip origins are dispersed throughout the region, and higher percentages (5.2 – 9.3%), as per expectation, are directed towards the Industrial areas in Kitchener, Cambridge and along the 401. Interestingly, none of these trip origin CTs is served by the current ION route (constructed and planned), which makes the shift from auto to transit less likely. Additionally, the hypothesis in the previous section which stated the possibility of longer travel distances by auto, cannot be confirmed with confidence, as although there is low Intra CT activity, a similar pattern is observed in transit flows, where commuters opt for buses, despite travelling distances comparable to that by auto travellers. However, higher average household annual incomes and vehicular ownership are more indicative of higher auto modal shares in these areas. (Appendix, Figure 5 & 6)



Map 16

c. Trip Destinations

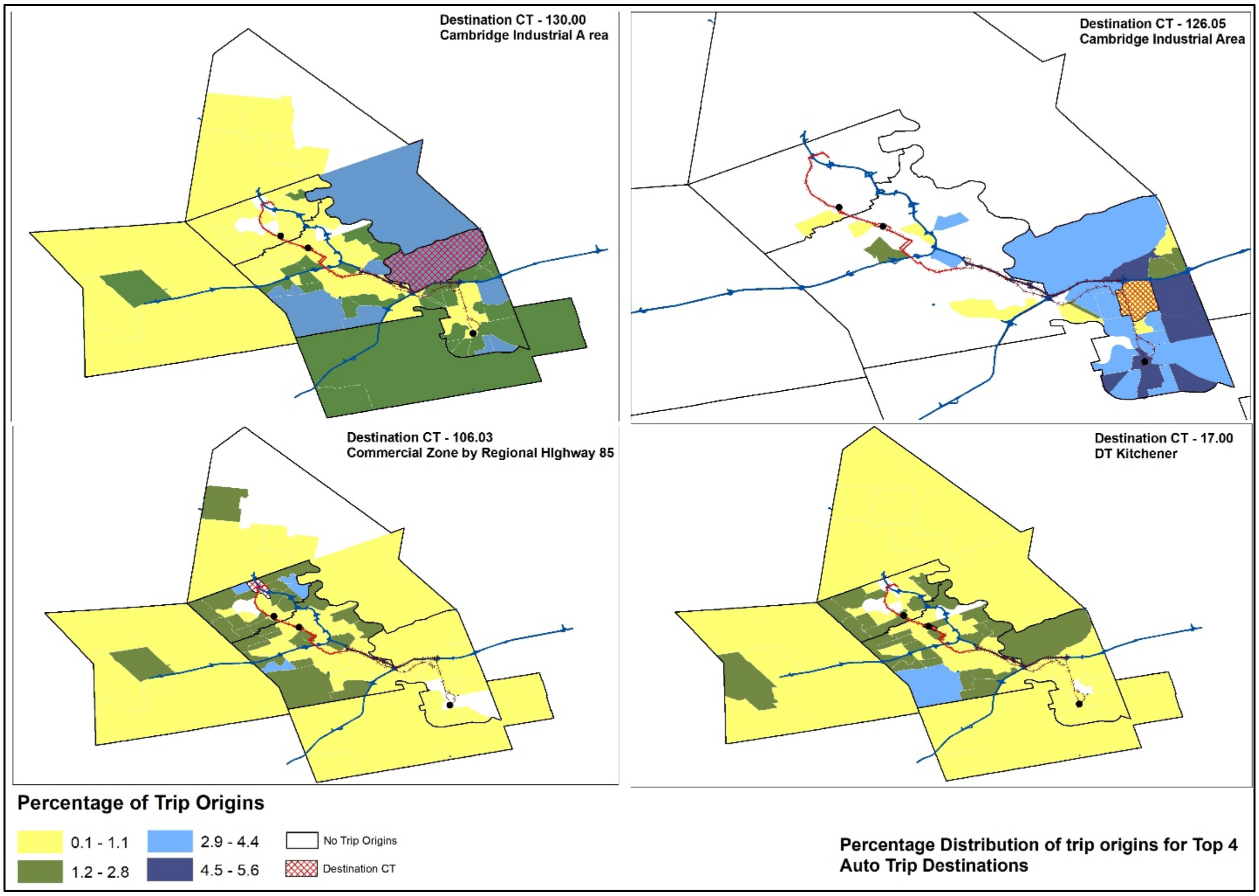
Trip Destinations by auto follow a similar pattern to trip origins. Higher mode shares are clustered in the peripheral zones zone and in industrial/business areas (Map 17). This further highlights the unsustainable commuter travel patterns in the region, as at least 53.8% trips received at all destinations are undertaken through auto modes and this this range extends to 100%, implying no commuter transit or cycling activity in these census tracts. These patterns suggest the need of better connectivity to high employment zones in the region.



Map 17

d. Flow Analysis of the Top 4 Trip Origins by Auto

The top four trip destinations by auto modes are the centres of employment – Cambridge Industrial Areas, DT Kitchener and Commercial Zone by Regional Highway 85, which are all connected by ion. Flow Analysis reveals that these destinations attract auto trips from throughout the region. The CTs abutting the LRT zones each generate less than 1.1% of the trips that towards these destinations, which have the potential to transition into transit trips due to better connectivity through ION. Auto traffic is generated in larger percentages (4.5 – 5.6%) from the periphery of the tricity and the surrounding smaller townships, which is expected to witness negligible change due to the introduction of LRT system.

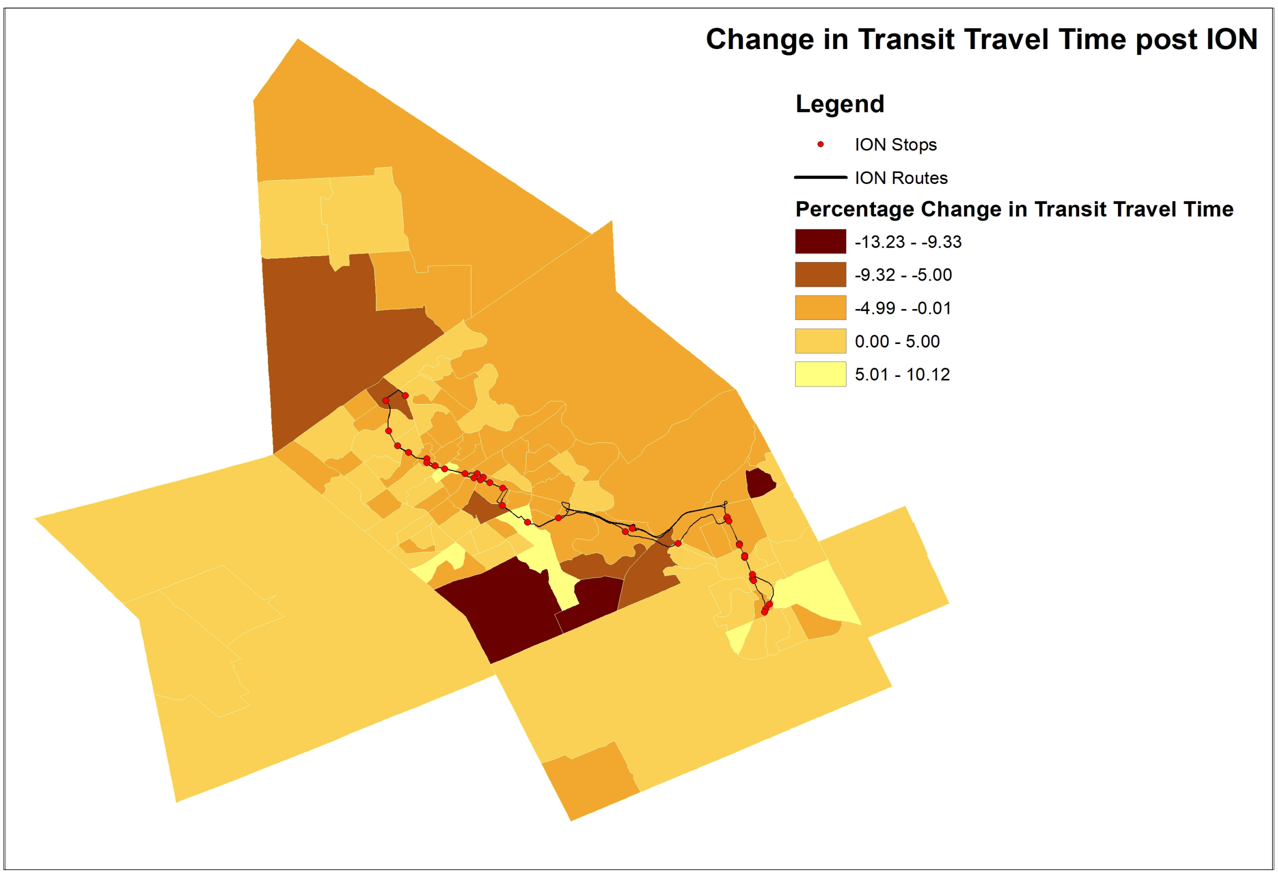


Map 18

e. Change in Transit Travel Time

The average change in transit time for the region overall is estimated to be about 0.14%. This is a result of the integration of LRT in the existing transit system which led to the realignment of bus routes. The change in transit travel time is based on the travel time from the centroid of the origin census tract to the destination census tract through the best combination of walking and transit, or access and invehicle time. The change in travel time (Map 19) shows that the census tracts surrounding the LRT witness the most decrease in travel time. Kitchener and Cambridge Industrial areas both witness decrease in travel time, mostly due to new transit routes. These improvements indicate a positive shift towards transit, especially in these areas as they are major employment zones and attract a high share of trips.

While the general spatial pattern of travel time decrease aligns with expectations, it is surprising to see North Dumfries in the higher range of travel time decrease. However, further investigation reveals that the average transit travel time from this CT in 2018 was 223 minutes, which decreased to 213 minutes post ION. Thus, the change is too small when compared to the overall travel time to cause a substantial shift towards transit. The scenario is same in Woolwich with an average transit travel time of 115 minutes, and Wilmot with 133 minutes.



Map 19

4.3 Model Results

The Methodology section elaborates on the process undertaken to specify and build a mode choice model for the region of Waterloo. This section analyses the results obtained from the final specification (table 4.2), and the prediction results thereof. Multiple specifications were tested for both, Multinomial and Nested Logit models, and the later proved to be a better fit for this exercise. The variables which proved to be significant in the specification were travel time, median household Income, Vehicle Ownership, Employment Density and population density. The results of the estimation are below:

Table 4.2: Multinomial Model Estimation Results

	Multinomial Logit Model	
Parameters for dependent variables*	*Coefficients*	*p-value*
Car Truck or Van as a Driver (CTV_D)*		
Travel Time	-0.03530	0.0667
Vehicle Ownership	0.14812	0.2194
Car Truck or Van as a Passenger (CTV_P)*		
Constant	-0.91657	0.0004
Travel Time	-0.10018	0.0000
Transit*		
Constant	0.52021	0.0292
Travel Time	-0.02810	0.0000
Income	-0.21328	0.0000
Walk*		
Constant	1.36651	0.0626
Travel Time	-0.03960	0.0000
Vehicle Ownership	-0.69367	0.0534
Employment Density	0.01070	0.0363
Population Density	-0.00035	0.0463
Bicycles*		
Constant	-2.51940	0.0000
Travel Time	-0.11904	0.0000
Log-Likelihood function	-2831.44041	
Chi-Squared	614.91	
		*Dependent Variable

The model reveals that the choice of mode 'Car Truck or Van as a Driver (CTV_D)' is largely influenced by travel time. The relationship with choice probability is inverse, meaning increase in travel time discourages commuters from travelling by CTV_D. Vehicle ownership, however, shows a positive relationship with the choice of this mode, implying availability of a vehicle increases the change of using for commuting. However, it is important to note that the p value for travel time is 0.0667, making it insignificant at 95% confidence interval where alpha for comparison is 0.05. Thus, the association between choice of CTV_D and travel time is weak. Thus, increase in travel time by CTV_D might not significantly decrease auto mode share. Additionally, Vehicle Ownership is not statistically significant in the choice of CTV_D, for both 95 and 85% confidence intervals, however, it is significantly different from the impact that vehicle ownership has on other modes, like cycling. Furthermore, these results lead to the speculation that various other behavioural factors, which have been emphasized in the theory play a role in the decision making of an individual and they are worth exploring for further identification of characteristics which promote driving as a mode choice.

Commuting by Car Truck or Van as a Passenger shows significant negative relationship with Travel Time, thus increase in travel time decreases the affinity to this mode. Currently, the average travel time for CTV_P is 15 minutes. Additionally, spatial analysis in the previous section revealed that this mode has higher shares in the peripheries of the tricity. This suggests that Car-pooling or CTV_P is the mode of choice in peripheries of urban areas, where the distance to employment zones is longer and yet the road network is not congested, implying relatively shorter travel times. Other variables which were expected to influence choice behaviour towards CTV_P such as marital status and family structure were statistically insignificant. This may also be attributed to the aggregate nature of the data. The results might be more refined with stated preference data, where the social dynamics of mode choice emerge to be discrete and reflect an impact on choice behaviour.

Transit Choice in the region is influenced by Travel time and Median Household Income. The p value of travel time coefficient is 0.00, significantly below 0.05, implying the travel time is a major consideration for the choice of transit. The negative sign indicates that increase in travel time decreases the affinity to transit. Additionally, Income too, has an inverse relationship with transit choice, as the increase in income decreases the probability of transit choice for commuting. Although vehicle ownership did not emerge as a significant variable, income can be related to higher affordability and thus propensity for vehicle ownership. Additionally, Intersection Density, which was included to reflect accessibility to transit, emerged as insignificant in the model results.

Choice of walking for commuting showed significant relationship with travel time, employment density, population density and vehicle ownership. Unexpectedly, intersection density remained insignificant for Walking as well, despite spatial analysis revealing a positive relationship between active transit and intersection density. This can be attributed to the aggregated nature of data, and the lack of variability between different origin destination pairs. Relationship with travel time is strong, as indicated by p value of 0.00, significant beyond 0.05, and decrease in travel time increases the probability of walking to work. Vehicle ownership has a weak and negative association with walking, which can again be related to income and affordability. Both population and employment density, too have a weak albeit positive relationship with choice. This further suggests that non-availability of car does not alone contribute towards choice of commuting by walk, and an individual might choose to walk to work based on spatial fabric of his/her location such as the landuse mix. Lastly, the results indicate that areas with denser development witness higher mode shares of walking.

The model explains cycling mode choice only on the basis of time, the coefficient of which has a significant value of 0.00. Again, decrease in travel time increases the probability of choice. However, among all the modes, cycling has the largest constant, which represents unobserved utility, although none of the factors included in the research showed significant p values during model runs. This hints that there are many other factors at play in the choice behaviour which have not been identified in this research.

The coefficients of travel time in the model results are comparable to those reported in the literature. This implies that the magnitude of significance of the variable (travel time) aligns with the expected values. For Car travel time, Hensher & Rose (2007) reported a coefficient of -0.03 for both car and transit, which is comparable to 0.03 and 0.028 obtained for car and transit respectively in this study. The relationship of vehicle ownership is weaker as it emerged insignificant for all modes except walking, with a coefficient of -0.69, while most researchers reported a value of below -1 (Eluru et al., 2014; Ashalatha et al., 2013;). This may be because of aggregated values, resulting in lower variability of the variable among different modes. Income shows a stronger relationship, with transit at -0.21, while Hensher and Rose (2008) report the same to be -0.007. This implies that the choice to take transit is strongly related with income or affordability in the region. Furthermore, population density has a weaker relationship with mode choice, related only to walking with a coefficient of -0.00035, while Buehler (2011) reports a much stronger relationship at 0.149. Overall, the results conform with findings from the literature, where majority studies reported travel time, vehicle ownership and Income as significant factors which impact the probability of mode choice.

Other factors which were included in the process such as gender, education status, living situation, marital status and presence of children had insignificant p-values during model runs. It may be a result of aggregate nature of data that did not allow enough variability between these factors for different mode choices. It is also important to note that the specification of the model was built around travel time, as it was the variable known to be impacted by the introduction of LRT. Thus, that model specification was chosen which had significant p-values for travel time. Hence, while this model does not show significance of these variables, contrary to the results reported in the literature, specific studies to understand the role of these factors might better highlight their part in determining choice behaviour in the region.

Simulation:

The average decrease in transit travel time for the region is estimated to be 0.14%. This model was simulated for 0.14%, 10%, 15%, and 20% decrease in travel time. (Table 4.2) It is essential to highlight that the base share and the simulation calculated after eliminating the 41 bad observations from the data sample. The base mode shares, however, remain relatively unchanged. It is also important to note that the simulation takes into account the entire LRT network – planned and running.

Table 4.3: Mode Shares in Simulated Scenarios

		Base Share	0.14% Decrease	10% Decrease	15% Decrease	20% Decrease
Mode Shares	CTV_D	84.17	84.04	83.28	82.81	82.30
	CTV_P	6.17	6.17	6.12	6.07	6.04
	Transit	6.54	6.62	7.46	7.98	8.55
	Walk	2.19	2.19	2.17	2.16	2.15
	Cycle	0.96	0.96	0.95	0.95	0.94

The results of various simulation scenarios indicate a positive shift towards transit ridership. 0.14% decrease in travel time, which is the estimated on-ground scenario; transit ridership is expected to increase by 0.09%. This shift is, however, not significant, and it decreases auto mode share by a mere 0.07%. In the first hypothetical scenario, with 10% reduction in transit travel time, increases ridership by 0.92%, pulling about 0.83% from CTV_D. For simulation at 15% decrease, the increase in ridership is 1.4%, while that for 20% is 2.01%. Additionally, the results reveal that introduction of LRT reduces the probability of car-pooling at 10%, 15% and 20% decrease in travel time by 0.92%, 0.98% and 0.13% respectively. The decrease is however not significant enough to cause considerable positive results in reality. Introduction of the ION has potential to cause a shift, albeit small, from active transportation towards transit as well. This can be attributed to the physical location and connectivity on the new system, which is through the areas which have high existing active transportation mode shares. Furthermore, post LRT there isn't a significant increase in transit ridership forecasted, because LRT is integrated into the transit system in a way that areas with existing higher transit mode shares have accessibility to the new system. Thus, the existing bus users tend to be the natural users of LRT, instead of a shift from other modes. Additionally, spatial analysis reveals that Census Tracts which have high trip origins by cars are not connected to their destinations through transit, which again reduces the probability of shift from other modes towards the new transit system.

5 Conclusion

The goal of this research was to develop an analytical basis for exploring the introduction of a new transit system in the region. This was achieved through three research objectives: (1) describing the current commuting patterns in the region of Waterloo, (2) Identifying the factors which affect commuting mode choice behaviour and finally (3) modelling these patterns to finally predict the potential changes after the introduction of the LRT. This exercise provides a methodical approach to inform transit related policies and investments in the region.

5.1 Summary of Findings

This study summarised the existing literature on commuting mode choice behaviour, which set the base for establishing prominent factors that impact mode choice. The existing transportation patterns of the region were studied for all three mode types – auto, active transportation and public transportation. This exercise revealed that the industrial areas in Cambridge, Kitchener and the University of Waterloo attract the highest number of commuters in the region. However, both the trip origins and destinations in the region are dispersed, the former more so than the latter, which is indicative of a good land use mix in the region. Transit and active transportation activity were most prominent in the downtown areas and the central corridor, where the first phase of the LRT has been constructed and functioning leading to the proposed second phase towards downtown Cambridge.

The theoretical base built in the literature review formed the foundation of the modelling exercise. Different model specifications were tested, both, for nested logit models and multinomial logit models. Nested logit models, despite their advantages, proved to be inappropriate for application in the region. As detailed in Chapter 3, this can be attributed to the aggregate nature of the data, and lack of stated preference data which would have better informed the nested logit model. However, this gave an opportunity to explore multinomial logit models, which were a better fit with respect to the available data.

The modelling exercise began with testing various variables for their influence on mode choice decision making in the region, namely, Socio Economic Characteristics (Gender, Marital Status for both Genders, Household structure – Married/ Common law with children, married/ common law without children, single parent or Not living in census family, Median household annual income, average household vehicle ownership and Education Status), Characteristics of Landuse and Built form (Employment Density, Population Density and Intersection density) and lastly mode specific variables - Travel time. Modelling these factors revealed that attributes like gender and education status which emerged significant in the literature review reflected little to no impact in the developed model, which can be a result of aggregate nature of the analysis. Ultimately, the variables which emerged significant and were included in the final model specification are – travel time, vehicle ownership, household income, population density and employment density.

The impacts of various factors in choice behaviour conformed with the observations in the literature review. Travel time had a significant and negative relationship with all modes, implying increase in travel time decreased the likelihood of choosing that mode. Income emerged as an important factor in the choice of transit. Various studies explored in the literature review reported that higher household income increases affordability which can be linked with higher vehicle ownership. The model, too, maintained that increase in affordability decreases the affinity towards transit. Walking choice behaviour in the region showed dependence on travel time, employment density, population density and Vehicle ownership, where again, the results align with the literature. This revealed that higher population and employment density is positively related to pedestrian commuter activity. Thus, compact developments with a good land use mix promote walking among commuters. Vehicle ownership has an inverse relationship with probability of walking as mode choice as increase in vehicle ownership reduces the probability of commuting on foot. Lastly, cycling was found to be sensitive only to travel time and had similar inverse effect as the other modes.

The developed model was used to assess the potential impacts of introduction of LRT to the existing transportation system. Simulations were conducted for varying decrease in transit travel time due to the new system. At 10% decrease, the model predicted an increase of 0.92% in the commuter transit modal share and 1.4% increase was projected for 15% decrease in travel time. The transit travel time from 2018 to 2019 was estimated to decrease by 0.14%, for which the model predicted 0.09% increase in transit ridership. Furthermore, at 20% decrease, the transit ridership is expected to increase by 2.01%. Additionally, these scenarios pull about 0.07 to - 1.8 percent from private automobiles, which is way below the 15% targeted by the region. This exercise has significant findings that can be implemented to push for greater public transportation shares in the region and have been discussed below.

5.2 Research Contribution and Recommendations

Mode Choice Modelling is an essential exercise in transportation planning, and is a continuously evolving process, informing decisions regarding major transit investments. With the widespread awareness about global warming and climate change concerns, it is now more important than ever, to devise ways which increase the share of sustainable transportation. Spatial understanding of current travel patterns and analytical predictions related to investments provide a robust basis for making informed transit planning decisions.

The research findings assist in developing valuable recommendations for future transit planning in the region. The LRT intends to attract development and boost transit ridership. In its current form, without accounting for the spatial and landuse changes that the LRT is generating, it increases transit ridership by about 0.09%. In the short term, it does not attract a major share of new riders towards transit but emerges as an infrastructure improvement for the existing riders. This is reflected in the simulation scenarios, where even 20% decrease in travel time projects an increase of only 2%. Furthermore, increase in transit shares post LRT pull from the short pedestrian trips as well. Spatial analysis further echoes similar concerns, as it revealed that the Census tracts which generate high volumes of commuter trips are not directly connected via the LRT to their respective destinations such as the Kitchener Industrial area and the Cambridge Industrial area. Existing body of literature on mode choice behaviour too, suggests that a greater number of transfers between origin and destination, decrease the probability that an individual will opt for transit, especially for commuting. It is thus recommended that LRT extension to connect these areas should be considered and feeder system should be enhanced by increasing the service and frequency of bus service in the periphery of the tricity. Gradual Extension of ION routes is a long-term recommendation and improvements in feeder system is endorsed to increase transit shares in the short term.

It is important to note that LRT is a tool for growth management and the conception of this project was not merely centered around moving people, but around shaping future developments in the region. It can thus, attract new, denser development and influence the location of jobs and housing in the future, making the urban area more compact. Denser development along the LRT corridor might generate potential for larger proportion of commuter trips taking place through the LRT or active transportation. Furthermore, the changes in travel time are not merely results of introduction of LRT, but realignment of transit system as a whole. Due to investments and development that the LRT corridor is now witnessing, It is recommended to undertake a similar exercise after the transit system has been in functioning for some time in the region, to understand and assess the actual impacts that LRT has had on ground by accounting for the spatial and landuse changes brought about by ION. This research should be a mixture of revealed and stated preference data and account for factors which were not included in this exercise such as behavioural characteristics of individuals making the choice. Additionally, this study focuses only on one dimension of peak time transit usage - commuters, aged between 15 – 64 years, while LRT caters to all age groups for various trip purposes. This study is thus, a first step towards understanding the impacts of LRT, and analysis on ridership as a whole needs further work and contribution.

5.3 Limitations

This study reveals various unique findings pertaining to travel patterns in the region of Waterloo and develops template to establish relationships between various choice considerations and mode choice behaviour in the region of Waterloo. However, this research, like most studies has certain limitations and the reader should inform himself/herself of these before interpreting the results in Chapter 4. These constraints and their impacts on the study have been discussed below.

A major constraint of this research is the availability of data. Special data compilation obtained from Statistics Canada formed the basis of the study. However, aggregated nature of data decreases the variability of attributes between different observations. Due to this, various factors such as gender and household structure did not emerge significant for choice behaviour, which are otherwise reported to have association with mode choice in the literature. Furthermore, regression models fit best with lesser number of variables. Thus, many variables which were significant, but their inclusion made travel time insignificant were excluded from the specification. The model specification was so chosen that it fit the data and had the ability to answer the research questions. Additionally, in the absence of availability of travel times, they were enumerated from GTFS data. Thus, the calculation was done from centroid of origin to the centroid of the destination Census Tract which might not accurately reflect the actual access and egress travel times. Thus, there is a potential danger of overestimation of transit travel time, although precautions were taken to check for this error by manual checking of a small sample of this data.

Another major constraint in the undertaking of this exercise was the time. This model was developed only using revealed preference data while an ideal model would use an amalgamation of both, stated and revealed preference to give more robust results. This would further have imparted more variability in different observations, potentially leading to more associations between considered attributes and mode choice behaviour. The lack of time however, discouraged collection of RP data and it is suggested that an interested researcher undertake this task in a future project.

The role of behavioural attitudes and personal preferences in mode choice behaviour are well documented in the literature (Ben-Akiva & Morikawa, 2002; Day, 2008; Bahamonde-Birke at al. 2017). However, the lack of revealed preference data did not enable the model to account for these factors. Thus, the model is unable to account for choice behaviour of individuals who, for example, live in a high population and employment density CT, have a short commuting trip, and own a vehicle and yet choose to take transit to work instead of walking or driving. Inclusion of individual preferences and attitude variables would have furthered this study but have not been included for analysis.

In consideration of the above limitations, the study has scope of further improvement. The recommendation would be undertaking a stated preference data collection exercise, which when coupled with the available data would provide valuable insights into mode choice behaviour. Additionally, the literature highlights the advantages of nested logit models over multinomial logit, but the former did not prove a good fit for the available data as the IVs of the regression contributed to be above 1, which indicated that the alternatives in the nest do not relate well together. It is recommended to attempt to model choice behaviour with nested logit specifications once additional data is collected.

5.4 Opportunities for future research

This study has two major outputs; firstly, understanding of Travel Patterns through spatial analysis; and secondly explanation of these patterns through analytical modelling. While this study provides important and significant findings, there is potential of further exploration to improve and perfect the model. Future research should seek to collect stated preference data and then undertake this modelling exercise to identify more attributes which determine the choice behaviour in the region. It would be fascinating to see if the simulation results with additional variables corroborate the findings of this study.

In the region of Waterloo, it would be interesting to reflect on the attitudes of individuals towards transit and active transportation and which factors that promote them. This study missed on factors like people's perception towards improvements in transit system, attitude towards bus v/s the LRT, safety or comfort, which would be important in the overall mode choice. There is the potential of further exploration in this aspect of choice behaviour to further inform the model and make it more unique to the Region of Waterloo. Additionally, this study focuses only on commuter travel behaviour, investigation of recreational and educational trip activity is another aspect that deserves more digging to understand travel patterns in a more holistic light.

This research focused on the commuting patterns of the region as a whole, however, LRT is more likely to intensify development and promote growth in the central transit corridor. A detailed study looking at the impacts of LRT only along the central corridor at the dissemination area level has the potential to provide more insightful results. The LRT corridor is witnessing investments, intensification and development which would be interesting to study. Furthermore, it would be interesting to look at the housing scenario along the LRT corridor, as gentrified areas often tend to squeeze out affordable housing.

Lastly, this study explored two modelling techniques – Multinomial and Nested logit models. Although the later finds more support in the literature, it did not work well with the nature of the available data. One recommendation would be to build on this study and estimate a nested logit model with additional data, and second would be exploration of fuzzy logic-based models to explore modelling further. These are relatively new, interesting and supported by artificial intelligence which increases their accuracy. It would be interesting to compare the results of these models with the findings of this study.

6 References

1. Abdel-Aty, M., & Abdelwahab, H. (2001). Calibration of nested mode choice model for Florida (p. 104) [Final Research Report]. University of Florida.

2. Acker, A., & Ng, W. S. (2018). Understanding Urban Travel Behaviour by Gender for Efficient and Equitable Transport Policies (International Transport Forum Discussion Papers 2018/01). https://doi.org/10.1787/eaf64f94-en

3. Al-Ahmadi, H. M. (2017). The Effects of Mode Perceptions on Intercity Mode-Choice Behavior in Saudi Arabia. European Journal of Business and Management, 10.

4. Alavi, S. (2016). The role of parking pricing and parking availability on travel mode choice. Ryerson University, 58.

5. Anwar, A. M. (2013). Determinants of Modal Choice Behavior Using Nested Logit Model: A Journey to Work Trip in Dhaka City. Dhaka Metropolitan Development Area and Its Planning: Problems, Issues and Policies. Dhaka: Bangladesh Institute of Planners (BIP).

6. Ashalatha, R., Manju, V. S., & Zacharia, A. B. (2013). Mode Choice Behavior of Commuters in Thiruvananthapuram City. Journal of Transportation Engineering, 139(5), 494–502. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000533

7. Bahamonde-Birke, F. J., Kunert, U., Link, H., & Ortúzar, J. de D. (2017). About attitudes and perceptions: Finding the proper way to consider latent variables in discrete choice models. Transportation, 44(3), 475–493. https://doi.org/10.1007/s11116-015-9663-5

8. Bai, T., Li, X., & Sun, Z. (2017). Effects of cost adjustment on travel mode choice: Analysis and comparison of different logit models. Transportation Research Procedia, 25, 2649–2659. https://doi.org/10.1016/j.trpro.2017.05.150

9. Barff, R., Mackay, D., & Olshavsky, R. W. (1982). A Selective Review of Travel-Mode Choice Models. Journal of Consumer Research, 8(4), 370. https://doi.org/10.1086/208877

10. Bastarianto, F., Irawan, M., Choudhury, C., Palma, D., & Muthohar, I. (2019). A Tour-Based Mode Choice Model for Commuters in Indonesia. Sustainability, 11(3), 788. https://doi.org/10.3390/su11030788

11. Beltman, C. (2014). Intrapersonal variability in mode choice behaviour. University of Twente, 86.

12. Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: Theory and application to travel demand. MIT Press.

13. Ben-Akiva, M., & Morikawa, T. (2002). Comparing ridership attraction of rail and bus. Transport Policy, 9(2), 107–116. https://doi.org/10.1016/S0967-070X(02)00009-4

14. Bierlaire, M. (1995). A Robust Algorithm for the Simultaneous Estimation of Hierarchical Logit Models. Transportation Research Group, Dept. of Mathematics, University of Namur, Belgium.

15. Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. Annals of Operations Research, 144(1), 287–300. https://doi.org/10.1007/s10479-006-0015-x

16. Böcker, L., Dijst, M., & Prillwitz, J. (2013). Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review. Transport Reviews, 33(1), 71–91. https://doi.org/10.1080/01441647.2012.747114

17. Brandenburg, C., Matzarakis, A., & Arnberger, A. (2004). The effects of weather on frequencies of use by commuting and recreation bicyclists. In Matzarakis A., de Freitas C.R. and Scott D. (Eds) Advances in Tourism Climatology, Freiburg., 10.

18. Brownstone, D., & Small, K. A. (1989). Efficient Estimation of Nested Logit Models. Journal of Business & Economic Statistics, 7(1), 67. https://doi.org/10.2307/1391839

19. Bruch, E., & Feinberg, F. (2017). Decision-Making Processes in Social Contexts. Annual Review of Sociology, 43(1), 207–227. https://doi.org/10.1146/annurev-soc-060116-053622

20. Buehler, R. (2011). Determinants of transport mode choice: A comparison of Germany and the USA. Journal of Transport Geography, 19(4), 644–657. https://doi.org/10.1016/j.jtrangeo.2010.07.005

21. Chee, W. L., & Fernandez, J. L. (2013). Factors that Influence the Choice of Mode of Transport in Penang: A Preliminary Analysis. Procedia - Social and Behavioral Sciences, 91, 120–127. https://doi.org/10.1016/j.sbspro.2013.08.409

22. Cheng, L., Chen, X., Wei, M., Wu, J., & Hou, X. (2014). Modeling Mode Choice Behavior Incorporating Household and Individual Sociodemographics and Travel Attributes Based on Rough Sets Theory. Computational Intelligence and Neuroscience, 2014, 1–9. https://doi.org/10.1155/2014/560919

23. Cheng, S., & Long, J. S. (2007). Testing for IIA in the Multinomial Logit Model. Sociological Methods & Research, 35(4), 583–600. https://doi.org/10.1177/0049124106292361

24. Corpuz, G. (2007). Public Transport or Private Vehicle: Factors That Impact on Mode Choice. 30th Australasian Transport Research Forum, 11.

25. Creemers, L., Cools, M., Tormans, H., Lateur, P.-J., Janssens, D., & Wets, G. (2012). Identifying the Determinants of Light Rail Mode Choice for Medium- and Long-Distance Trips: Results from a Stated Preference Study. Transportation Research Record: Journal of the Transportation Research Board, 2275(1), 30–38. https://doi.org/10.3141/2275-04

26. Day, N. (2008). The Joint Modelling of Trip Timing and Mode Choice. University of Toronto, 127.

27. Ding, L., & Zhang, N. (2016). A Travel Mode Choice Model Using Individual Grouping Based on Cluster Analysis. Procedia Engineering, 137, 786–795. https://doi.org/10.1016/j.proeng.2016.01.317

28. Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial logit: A comparison of choice models for voting research. Electoral Studies, 23(1), 107–122. https://doi.org/10.1016/S0261-3794(03)00040-4

29. Edara, P. K. (2003). Mode Choice Modeling Using Artificial Neural Networks. Virginia Polytechnic Institute and State University.

30. Eluru, N., Chakour, V., & El-Geneidy, A. M. (2012). Travel mode choice and transit route choice behavior in Montreal: Insights from McGill University members commute patterns. Public Transport, 4(2), 129–149. https://doi.org/10.1007/s12469-012-0056-2

31. Errampalli, M., Okushima, M., & Akiyama, T. (2013). Development of the Microscopic Traffic Simulation Model with the Fuzzy Logic technique. SIMULATION, 89(1), 87–101. https://doi.org/10.1177/0037549712443716

32. Ewing, R., & Cervero, R. (2010). Travel and the Built Environment: A Meta-Analysis. Journal of the American Planning Association, 76(3), 265–294. https://doi.org/10.1080/01944361003766766

33. Forsey, D., Habib, K. N., Miller, E. J., & Shalaby, A. (2013). Evaluating the impacts of a new transit system on commuting mode choice using a GEV model estimated to revealed preference data: A case study of the VIVA system in York Region, Ontario. Transportation Research Part A: Policy and Practice, 50, 1–14. https://doi.org/10.1016/j.tra.2013.01.033

34. Foth, N. (2013). Transit Accessibility, Social Equity and Job Type: Commuting to Work in Toronto 1996 & 2006 [Thesis]. McGill University.

35. Ganji, M., Eftekhar, H., Shokri, F., Ismail, A., & Atiq, O. K. R. (2013). The Comparison of Mode Choice Sensitivity to the Reduction of Travel Time and Cost in Multimodal Trip. Research Journal of Applied Sciences, Engineering and Technology, 6(9), 1680–1684. https://doi.org/10.19026/rjaset.6.3890

36. Gärling, T., & Axhausen, K. W. (2003). Introduction: Habitual travel choice. 11.

37. Ghareib, A. H. (1996). Evaluation of Logit and Probit Models in Mode-Choice Situation. Journal of Transportation Engineering, 122(4), 282–290. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:4(282)

38. Guo, Z., Wilson, N. H. M., & Rahbee, A. (2007). Impact of Weather on Transit Ridership in Chicago, Illinois. Transportation Research Record: Journal of the Transportation Research Board, 2034(1), 3–10. https://doi.org/10.3141/2034-01

39. Habib, K. N. (2014). Household-level commuting mode choices, car allocation and car ownership level choices of two-worker households: The case of the city of Toronto. Transportation, 41(3), 651–672. https://doi.org/10.1007/s11116-014-9518-5

40. Hasnine, S., Lin, T., Weiss, A., & Habib, K. M. N. (2017). Is It the Person or Urban Context? The Role of Urban Travel Context in Defining Mode Choices for School Trips of Post-Secondary Students in Toronto". CD-ROM of 96th Annual Meeting of Trans. Res. Board.

41. Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by Bicycle: An Overview of the Literature. Transport Reviews, 30(1), 59–96. https://doi.org/10.1080/01441640903187001

42. Hensher, D. A. (1986). Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model. The Review of Economics and Statistics, 68(4), 657. https://doi.org/10.2307/1924525

43. Hensher, D. A., & Button, K. J. (2008). Handbook of transport and the environment.

44. Hensher, D. A., & Rose, J. M. (2007). Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. Transportation Research Part A: Policy and Practice, 41(5), 428–443. https://doi.org/10.1016/j.tra.2006.09.006

45. Hensher, D. A., & Ton, T. T. (2000). A comparison of the predictive potential of arti®cial neural networks and nested logit models for commuter mode choice. 18.

46. Hess, S. (2005). Advanced discrete choice models with applications to transport demand. Centre for Transport Studies, Imperial College London, 303.

47. Hess, S., Daly, A., & Batley, R. (2018). Revisiting consistency with random utility maximisation: Theory and implications for practical work. Theory and Decision, 84(2), 181–204. https://doi.org/10.1007/s11238-017-9651-7

48. Hodgson, G. M. (2012). On the Limits of Rational Choice Theory. Economic Thought, 15.

49. Hu, X., Zhao, L., & Wang, W. (2015). Impact of perceptions of bus service performance on mode choice preference. Advances in Mechanical Engineering, 7(3), 168781401557382. https://doi.org/10.1177/1687814015573826

50. Idris, A. O., Habib, K. M. N., & Shalaby, A. (2014). Dissecting the Role of Transit Service Attributes in Attracting Commuters: Lessons from a Comprehensive revealed Preference–Stated Preference Study on Commuting Mode-Switching Behavior in toronto, Ontario, Canada. Transportation Research Record: Journal of the Transportation Research Board, 2415(1), 107–117. https://doi.org/10.3141/2415-12

51. Kedia, A. S., Saw, K. B., & Katti, B. K. (2015). Fuzzy logic approach in mode choice modelling for education trips: a case study of indian metropolitan city. Transport, 30(3), 286–293. https://doi.org/10.3846/16484142.2015.1081279

52. Khan, O. (2007). Modelling passenger mode choice behaviour using computer aided stated preference data. Queensland University of Technology, 341.

53. Klöckner, C. A., & Matthies, E. (2004). How habits interfere with norm-directed behaviour: A normative decision-making model for travel mode choice. Journal of Environmental Psychology, 24(3), 319–327. https://doi.org/10.1016/j.jenvp.2004.08.004

54. Koppelman, F., & Bhat, C. (2006). A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. U.S. Department of Transportation Federal Transit Administration.

55. Kumar, M., Sarkar, P., & Madhu, E. (2013). Development of fuzzy logic based mode choice model considering various public transport policy options. International journal for traffic and transport engineering, 3(4), 408–425. https://doi.org/10.7708/ijtte.2013.3(4).05

56. Ladhi, S. C., Ghodmare, S. D., & Sayankar, B. B. (2018). Review on Assessment of Mode Shift Behavior due to Introduction of New Mass Transit System. International Esearch Journal of Engineering and Technology (IRJET), 05(04), 7.

57. Legrain, A., Eluru, N., & El-Geneidy, A. M. (2015). Am stressed, must travel: The relationship between mode choice and commuting stress. Transportation Research Part F: Traffic Psychology and Behaviour, 34, 141–151. https://doi.org/10.1016/j.trf.2015.08.001

58. Liu, C., Susilo, Y. O., & Karlström, A. (2017). Weather variability and travel behaviour – what we know and what we do not know. Transport Reviews, 37(6), 715–741. https://doi.org/10.1080/01441647.2017.1293188

59. Liu, Y., Chen, J., Wu, W., & Ye, J. (2019). Typical Combined Travel Mode Choice Utility Model in Multimodal Transportation Network. Sustainability, 11(2), 549. https://doi.org/10.3390/su11020549

60. Lo, S. H., van Breukelen, G. J. P., Peters, G.-J. Y., & Kok, G. (2013). Proenvironmental travel behavior among office workers: A qualitative study of individual and organizational determinants. Transportation Research Part A: Policy and Practice, 56, 11–22. https://doi.org/10.1016/j.tra.2013.09.002

61. Lucas, K., Blumenberg, E., & Weinberger, R. (Eds.). (2011). Auto motives: Understanding car use behaviours (First edition). Emerald.

62. Maat, K., van Wee, B., & Stead, D. (2005). Land Use and Travel Behaviour: Expected Effects from the Perspective of Utility Theory and Activity-Based Theories. Environment and Planning B: Planning and Design, 32(1), 33–46. https://doi.org/10.1068/b31106

63. Markley, S. (2007). Spatially-Oriented Discrete Choice Predictions: A Case Study of French Supermarket Preferences. Université de Toulouse I, France, 21.

64. McNally, M. G. (2007). Handbook of Transport Modelling: 2nd Edition, Chapter 3 (D. A. Hensher & K. J. Button, Eds.; Vol. 1). Emerald Group Publishing Limited. https://doi.org/10.1108/9780857245670

65. Ministry of Finance, Ontario. (2016). 2016 Census Highlights: Commuting to Work. Statistics Canada.

66. Ministry of Municipal Affairs and Housing. (2017). Growth Plan for the Greater Golden Horseshoe. 120.

67. Moniruzzaman, M., & Farber, S. (2018). What drives sustainable student travel? Mode choice determinants in the Greater Toronto Area. International Journal of Sustainable Transportation, 12(5), 367–379. https://doi.org/10.1080/15568318.2017.1377326

68. Nitta, Y. (1998). Transportation mode change model to special bus incorporating generalized time. Proceedings of the 8th international conference on transport and mobility for elderly and disabled people, Perth, Australia.

69. Ortuzar, J. de D. (1982). Fundamentals of discrete multimodal choice modelling. Transport Reviews, 2(1), 47–78. https://doi.org/10.1080/01441648208716481

70. Ortúzar S., J. de D., & Willumsen, L. G. (2001). Modelling transport. Wiley.

71. Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., & Lee, A. (2011). Importance of traveler attitudes in the choice of public transportation to work: Findings from the Regional Transportation Authority Attitudinal Survey. Transportation, 38(4), 643–661. https://doi.org/10.1007/s11116-011-9336-y

72. Pucher, J., Dill, J., & Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. Preventive Medicine, 50, S106–S125. https://doi.org/10.1016/j.ypmed.2009.07.028

73. Pulugurta, S., Arun, A., & Errampalli, M. (2013). Use of Artificial Intelligence for Mode Choice Analysis and Comparison with Traditional Multinomial Logit Model. Procedia - Social and Behavioral Sciences, 104, 583–592. https://doi.org/10.1016/j.sbspro.2013.11.152

74. Ramanuj, P. S., & Varia, H. R. (2018). Application of Soft Computing Techniques in Mode Choice Analysis. Pramana Research Journal, 8(11).

75. Region of Waterloo. (2010). Regional Transportation Masterplan.

76. Region of Waterloo. (2014a). Active Transportation Master Plan.

77. Region of Waterloo. (2014b). The ION Story.

78. Region of Waterloo. (2017). 2016 Labour Force Report.

79. Region of Waterloo. (2019). Regional Transportation Masterplan, 2018.

80. Region of Waterloo Community Profile. (2018). 17.

81. Rodríguez, D. A., & Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. Transportation Research Part D: Transport and Environment, 9(2), 151–173. https://doi.org/10.1016/j.trd.2003.11.001

82. Sabir, M., Koetse, M. J. J., & Rietveld, P. (2008). The Impact of Weather Conditions on Mode Choice: Empirical Evidence for the Netherlands. Proceedings of the BIVEC-GIBET Transport Research Day 2007, 512-527).

83. Saneinejad, S., Roorda, M. J., & Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. Transportation Research Part D: Transport and Environment, 17(2), 129–137. https://doi.org/10.1016/j.trd.2011.09.005

84. Sartori, J. J. P., & Robledo, C. W. (2013). Preferences for vehicle ownership and mode choice for commuting trips in Cordoba city. 13th World Conference on Transportation Research, Rio De Janerio, Brazil.

85. Schmöcker, J.-D., Quddus, M. A., Noland, R. B., & Bell, M. G. H. (2008). Mode choice of older and disabled people: A case study of shopping trips in London. Journal of Transport Geography, 16(4), 257–267. https://doi.org/10.1016/j.jtrangeo.2007.07.002

86. Sekhar, Ch. R. (2014). Mode choice analysis: the data, the models and future ahead. International journal for traffic and transport ENGINEERING, 4(3), 269–285. https://doi.org/10.7708/ijtte.2014.4(3).03

87. Sims, R., & Schaeffer, R. (2014). Climate Change: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [IPCC]. Cambridge University Press.

88. Statistics Canada. (2017). Journey to work: Key results from 2016 census.

89. Stover, V., & McCormack, E. (2012). The Impact of Weather on Bus Ridership in Pierce County, Washington. Journal of Public Transportation, 15(1), 95–110. https://doi.org/10.5038/2375-0901.15.1.6

90. Su, F., Schmöcker, J.-D., & Bell, M. G. H. (2009). Mode Choice of Older People Before and After Shopping: A Study with London Data. Journal of Transport and Land Use, 2(1). https://doi.org/10.5198/jtlu.v2i1.69

91. Sun, B., Ermagun, A., & Dan, B. (2017). Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. Transportation Research Part D: Transport and Environment, 52, 441–453. https://doi.org/10.1016/j.trd.2016.06.001

92. Tao, S., He, S. Y., & Thøgersen, J. (2019). The role of car ownership in attitudes towards public transport: A comparative study of Guangzhou and Brisbane. Transportation Research Part F: Traffic Psychology and Behaviour, 60, 685–699. https://doi.org/10.1016/j.trf.2018.12.005

93. Train, K. (2003). Discrete choice methods with simulation. Cambridge University Press.

94. Tyrinopoulos, Y., & Antoniou, C. (2013). Factors affecting modal choice in urban mobility. European Transport Research Review, 5(1), 27–39. https://doi.org/10.1007/s12544-012-0088-3

95. Urban Transportation Task Force. (2012). The High Cost of Congestion in Canadian Cities. Council of Deputy Ministers Responsible for Transportation and Highway Safety Secretariat.

96. van Loon, R., Rietveld, P., & Brons, M. (2011). Travel-time reliability impacts on railway passenger demand: A revealed preference analysis. Journal of Transport Geography, 19(4), 917–925. https://doi.org/10.1016/j.jtrangeo.2010.11.009

97. Wang, Y., Li, L., Wang, Z., Lv, T., & Wang, L. (2013). Mode Shift Behavior Impacts from the Introduction of Metro Service: Case Study of Xi'an, China. Journal of Urban Planning and Development, 139(3), 216–225. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000148

98. Werner, C. M., Brown, B. B., Tribby, C. P., Tharp, D., Flick, K., Miller, H. J., Smith, K. R., & Jensen, W. (2016). Evaluating the attractiveness of a new light rail extension: Testing simple change and displacement change hypotheses. Transport Policy, 45, 15–23. https://doi.org/10.1016/j.tranpol.2015.09.003

99. Xie, C., Lu, J., & Parkany, E. (2003). Work travel mode choice modeling using data mining: decision trees and neural networks. 82nd Transportation Research Board Annual Meeting, January 12-16, 2003, Washington, D.C, 28.

100. Yang, M., Li, D., Wang, W., Zhao, J., & Chen, X. (2013). Modeling Gender-Based Differences in Mode Choice considering Time-Use Pattern: Analysis of Bicycle, Public Transit, and Car Use in Suzhou, China. Advances in Mechanical Engineering, 5, 706918. https://doi.org/10.1155/2013/706918

101. Zhang, M. (2004). The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong. Journal of the American Planning Association, 70(3), 344–360. https://doi.org/10.1080/01944360408976383

7 Appendices

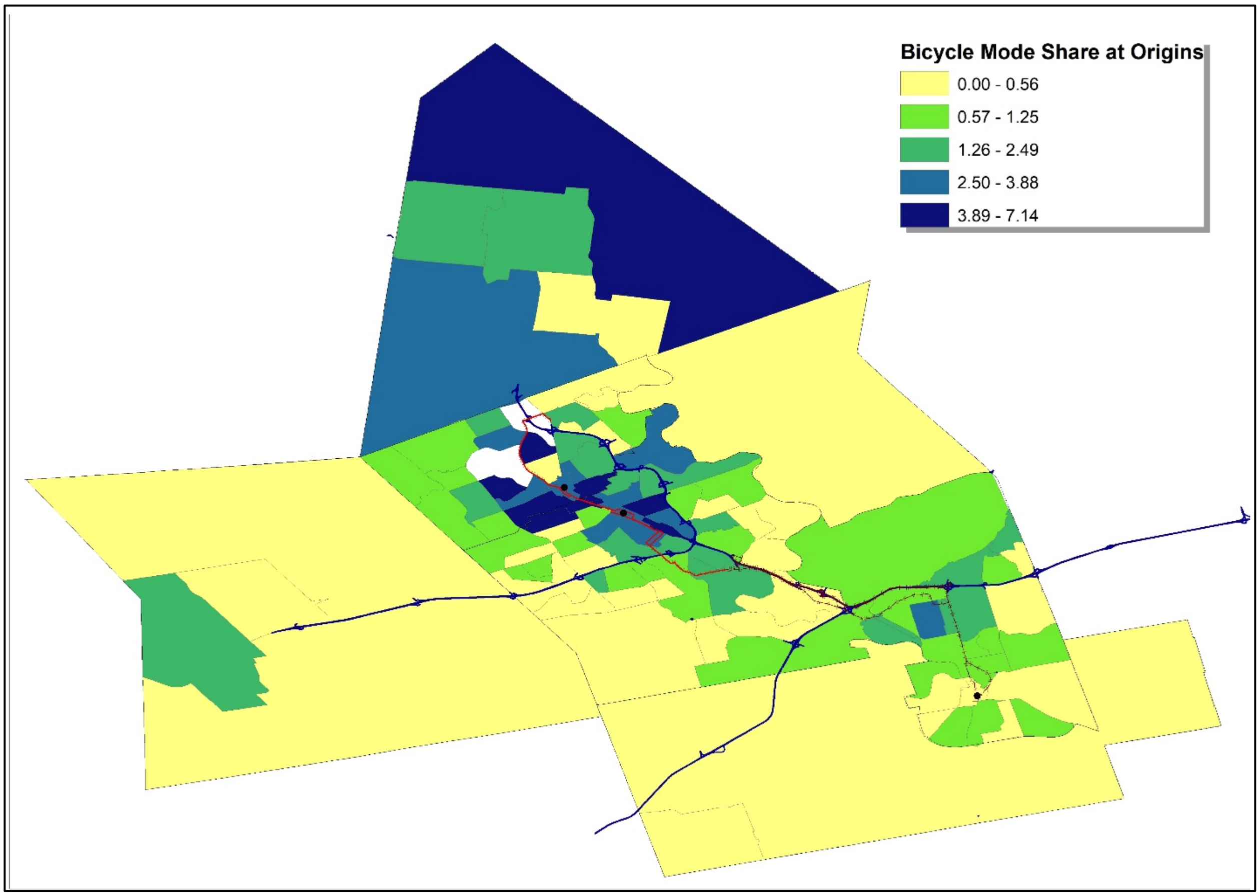


Figure 1

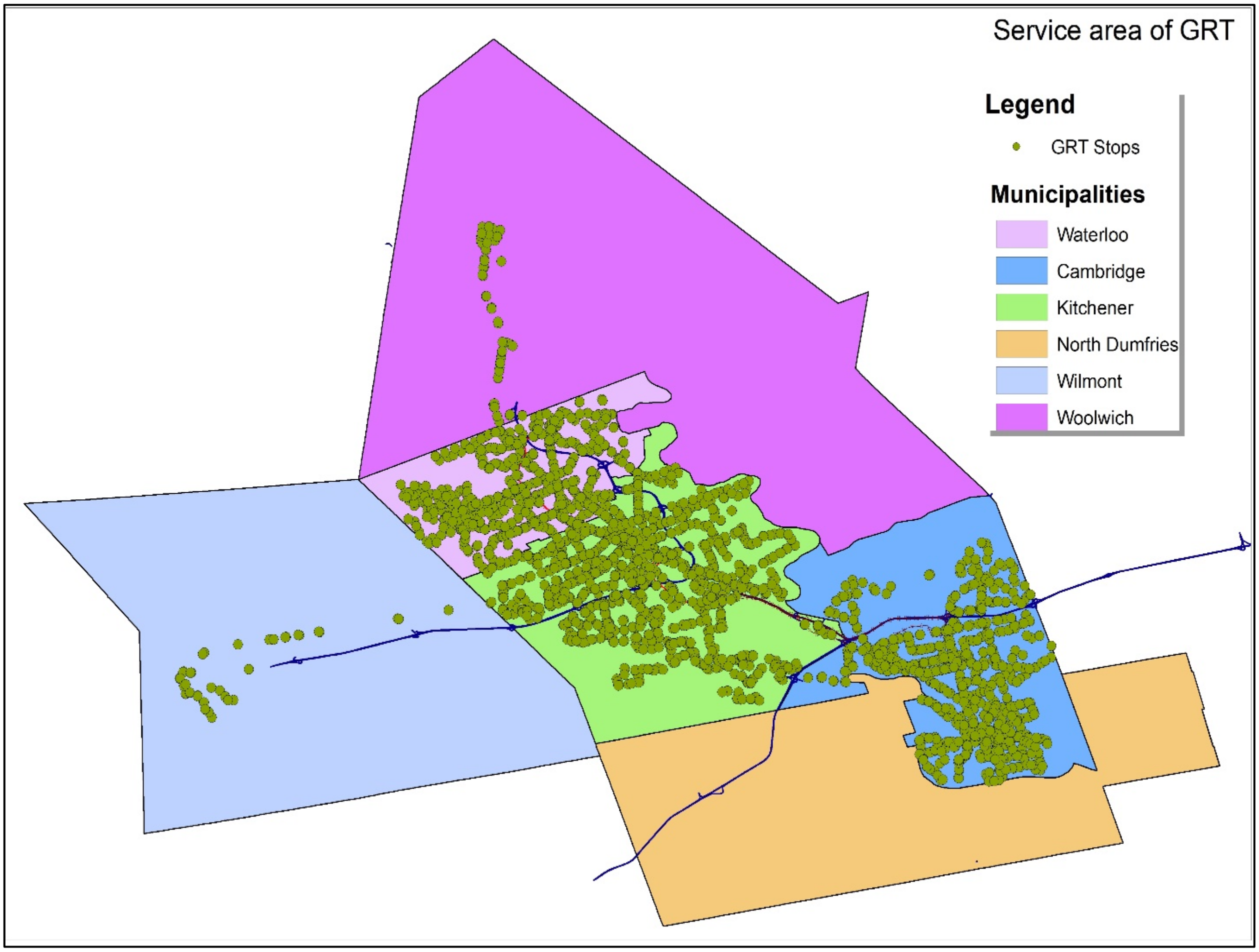
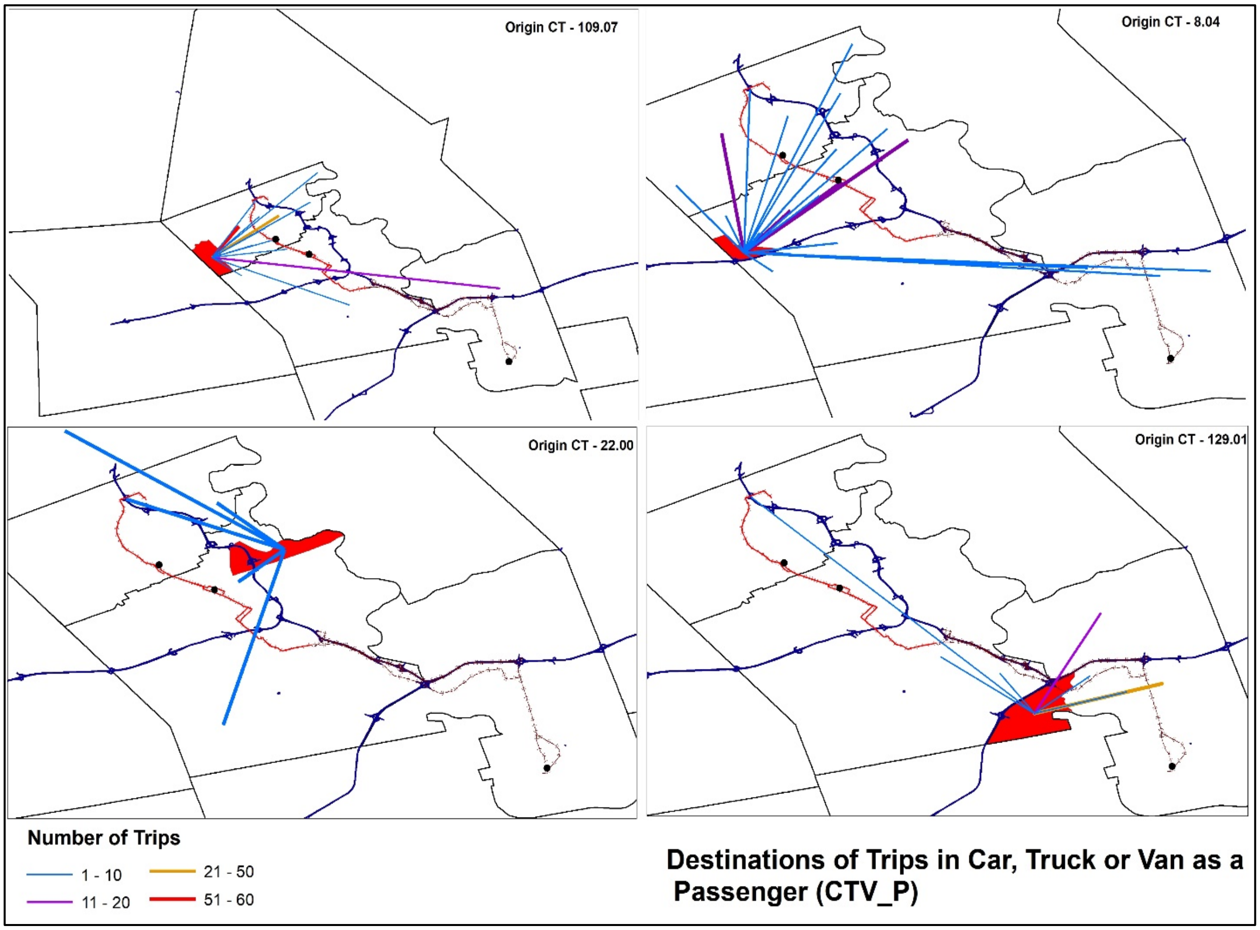

Service area of GRT
Legend
GRT Stops
Municipalities
Waterloo
Cambridge
Kitchener
North Dumfries
Wilmont
Woolwich

Figure 2



Origin CT - 109.07
Origin CT - 8.04
Origin CT - 22.00
Origin CT - 129.01
Number of Trips
1 - 10
11 - 20
21 - 50
51 - 60
Destinations of Trips in Car, Truck or Van as a Passenger (CTV_P)

Figure 3

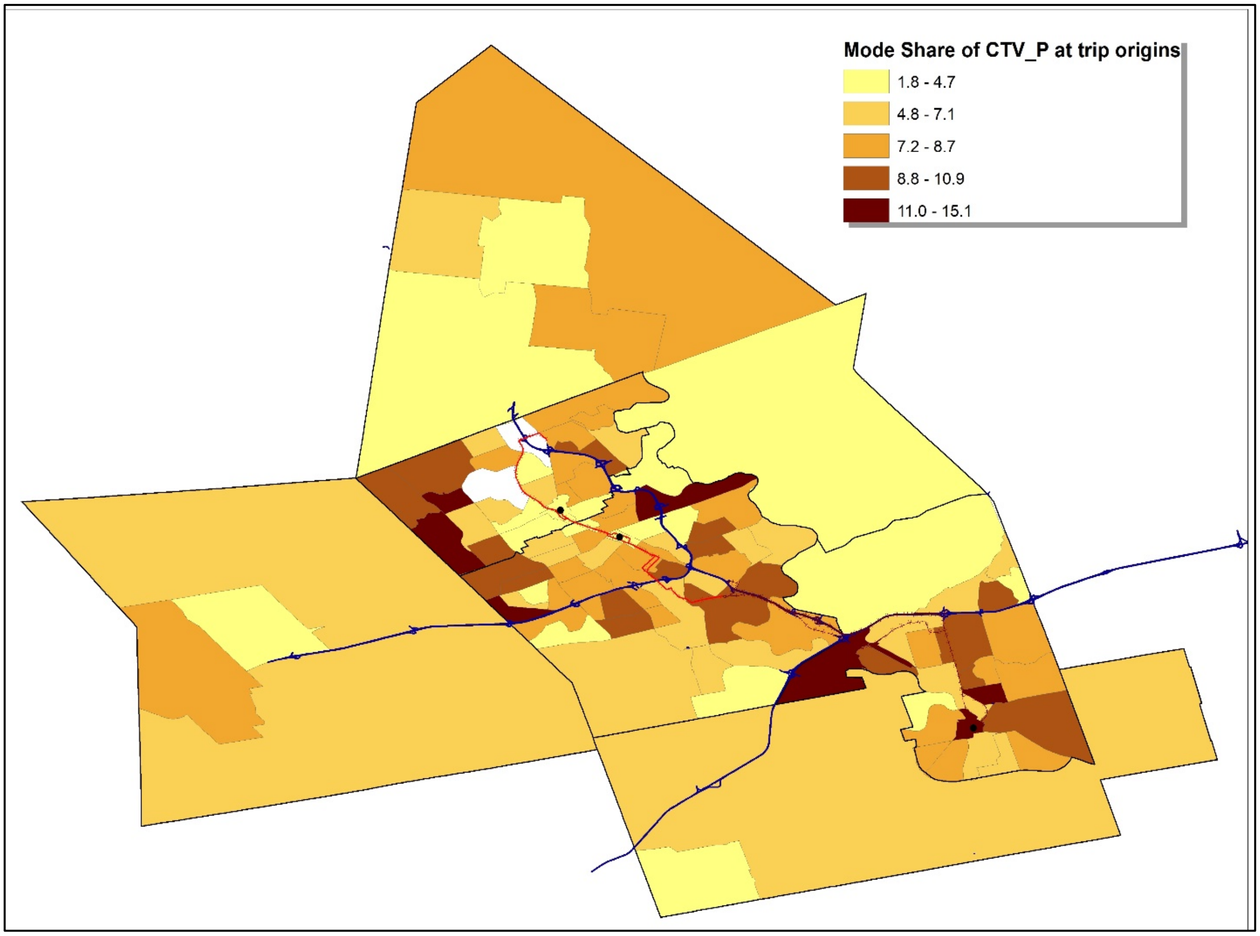

Mode Share of CTV_P at trip origins
1.8 - 4.7
4.8 - 7.1
7.2 - 8.7
8.8 - 10.9
11.0 - 15.1

Figure 4

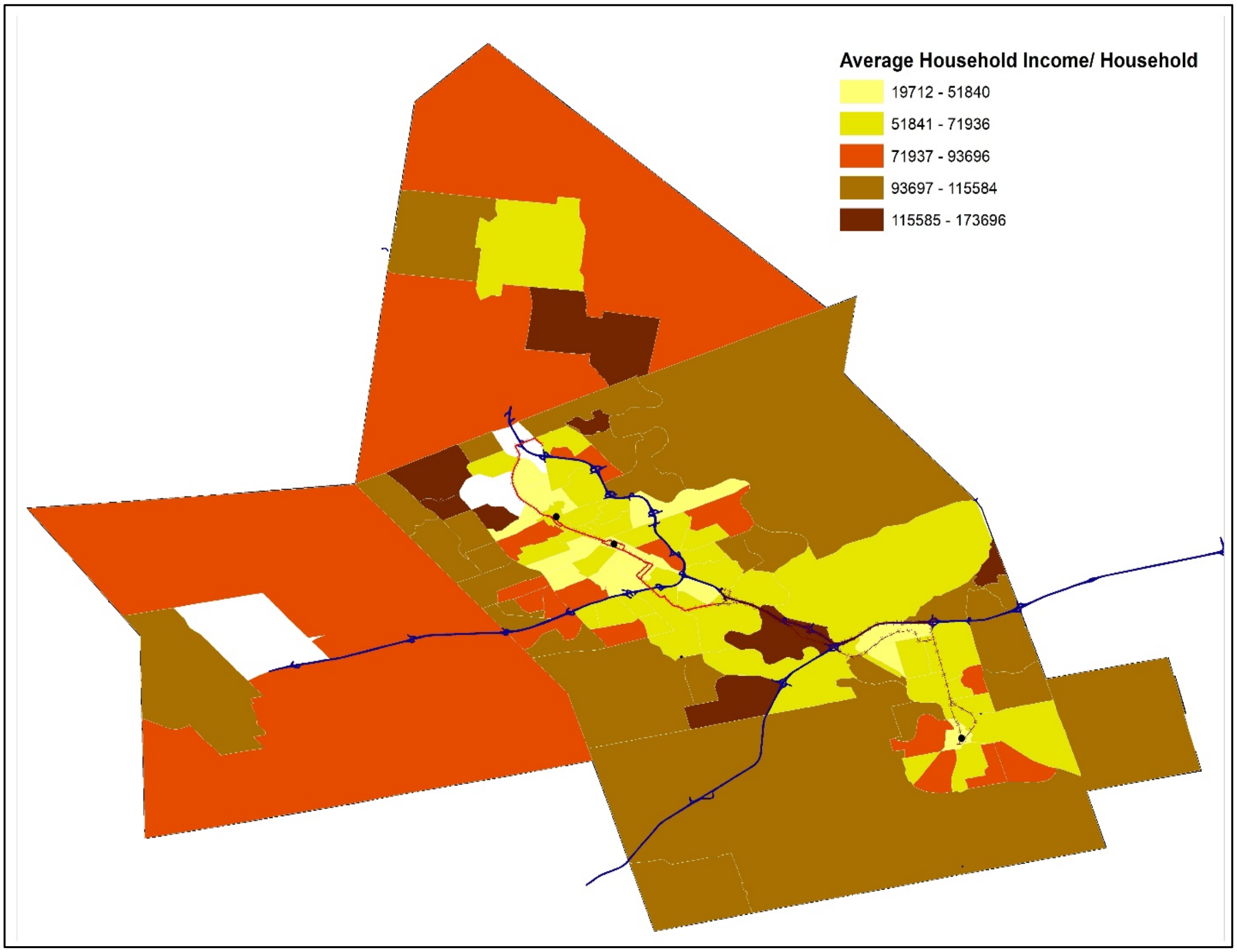

Average Household Income/ Household
19712 - 51840
51841 - 71936
71937 - 93696
93697 - 115584
115585 - 173696

Figure 5

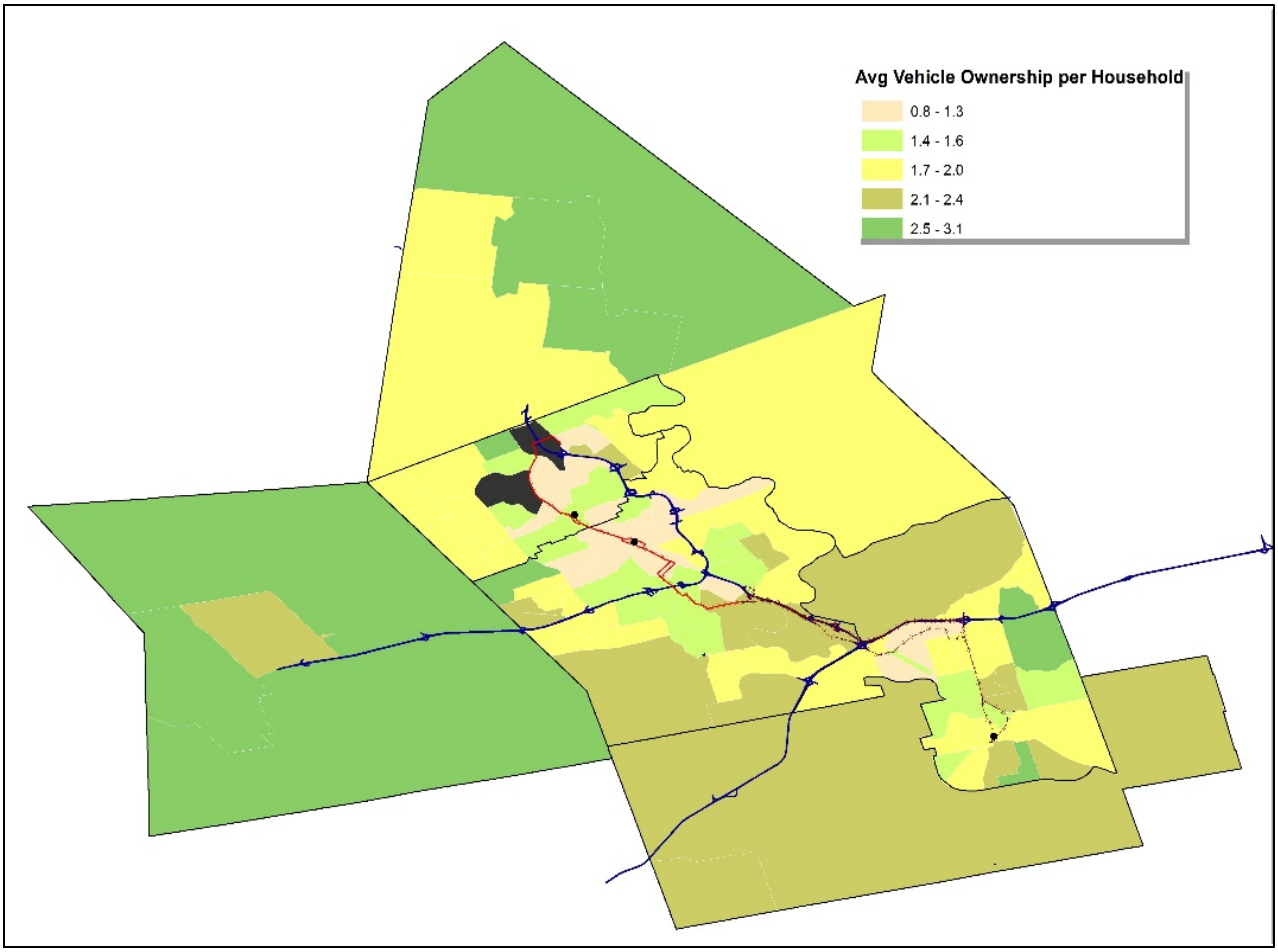

Avg Vehicle Ownership per Household
0.8 - 1.3
1.4 - 1.6
1.7 - 2.0
2.1 - 2.4
2.5 - 3.1

Figure 6

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -2831.44041
Estimation based on N =    5027, K =  14
Inf.Cr.AIC  =   5690.9 AIC/N =    1.132
----------------------------------------
            Log likelihood R-sqrd R2Adj
ASCs  only  model must be fit separately
               Use NLOGIT ;...;RHS=ONE$
Note: R-sqrd = 1 - logL/Logl(constants)
----------------------------------------
Chi-squared[10]          =    614.91857
Prob [ chi squared > value ] =   .00000
Response data are given as proportions.
Number of obs.=  5068, skipped   41 obs
```

CHOICE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
TT1	-.03530*	.01925	-1.83	.0667	-.07303	.00243
VO1	.14812	.12062	1.23	.2194	-.08829	.38453
B_CARP	-.91657***	.25846	-3.55	.0004	-1.42313	-.41000
TT2	-.10018***	.02097	-4.78	.0000	-.14128	-.05908
C_BUS	.52021**	.23850	2.18	.0292	.05275	.98767
TT3	-.02810***	.00547	-5.14	.0000	-.03881	-.01738
INC3	-.21328***	.03137	-6.80	.0000	-.27476	-.15179
D_WALK	1.36651*	.73384	1.86	.0626	-.07180	2.80482
TT4	-.03960***	.00406	-9.76	.0000	-.04755	-.03165
EM4	.01070**	.00511	2.09	.0363	.00068	.02071
PO4	-.00035**	.00017	-1.99	.0463	-.00069	-.00001
VO4	-.69367*	.35915	-1.93	.0534	-1.39760	.01026
E_BIKE	-2.51940***	.33816	-7.45	.0000	-3.18218	-1.85661
TT5	-.11904***	.02204	-5.40	.0000	-.16223	-.07585

Figure 7 Multinomial Model Estimation Results