Framework for the Integration of a Parameterized Logit Captivity Model for Morning Commuting in the Greater Toronto and Hamilton Area with an Agent Based Dynamic Traffic Micro Simulation

By

Adam Weiss BASc

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science

Department of Civil Engineering
University of Toronto

© Copyright by Adam Weiss 2013
Framework for the Integration of a Parameterized Logit Captivity Model for Morning Commuting in the Greater Toronto and Hamilton Area with an Agent Based Dynamic Traffic Micro Simulation

Adam Weiss

Master of Applied Science
Department of Civil Engineering
University of Toronto

2013

Abstract

This thesis proposes a framework that combines a mode choice model with a large scaled agent-based multimodal traffic microsimulation. Both components are discussed with respect to their development as separate entities. The mode choice model uses a formulation that explicitly considers latent modal captivity despite using conventional travel survey data. An existing multimodal microsimulation traffic assignment model used in the study area is enhanced and partially calibrated for use with the MATSIM traffic assignment tool. Both of the components are then tested independently in terms of statistical and behavioral validity and a conceptual procedure to test the implications of the mode choice model on mode switching behaviour within the traffic assignment model is presented. Other applications of both the travel assignment model and mode choice model are discussed. In order for the framework to become operational, further development with respect to the traffic assignment model is required.
Acknowledgments:

A number of people were instrumental in keeping me sane and providing suggestions and general support over the course of the completion of this project. Without the assistance of these people this work would not have been possible and as such I am heavily indebted to them for their support and assistance.

First and foremost, I would like to express my gratitude to Dr. Khandker Nurul Habib for his supervision of this thesis work and financial support. His guidance, patience and encouragement throughout the process of completing this work made this project possible and I am thankful to him for the opportunity to have completed this work. Another big thanks to Dr. Matthew Roorda for taking the time to review my work.

In terms of technical support, this project would not have been finished without the support of two individuals: Peter Kucirek and Mohamed Mahmoud. The assistance provide by Peter in terms of both understanding the intricacies of Java programming and of how the MATSIM platform functions were most helpful given my lack of experience with either. The GIS technical support provided by Mohamed during the course of this thesis provided many of the results used within as well as an opportunity for me to become familiar with GIS software.

A more general thanks to my colleagues in the transportation group: Jared, Jenessa, James, Ishwar, Fang, Siva, and the rest. The regular coffee breaks throughout the last year have been essential to staying awake and keeping sane. Finally, I would like to thank my family and extended circle of friends for their love and support over the last year. In particular thanks to my parents for listening to me talk about my research, even when you didn’t quite understand it and for providing proof reading of this thesis.
# Table of Contents

ACKNOWLEDGMENTS: .................................................................................................................. III

LIST OF TABLES .......................................................................................................................... VI

LIST OF FIGURES ........................................................................................................................ VII

LIST OF APPENDICES ................................................................................................................ VIII

1. INTRODUCTION ............................................................................................................................ 1
   1.1. STUDY AREA .................................................................................................................. 2
   1.2. DATA SOURCES .............................................................................................................. 3
   1.3. THESIS OUTLINE .......................................................................................................... 5

2. MODE CHOICE AND PARAMETERIZED LOGIT CAPTIVITY MODEL .................................... 7
   2.1. LITERATURE REVIEW ............................................................................................... 7
   2.2. MODEL FORMULATION ............................................................................................. 14
       2.2.1. RANDOM UTILITY THEORY BACKGROUND .................................................. 14
       2.2.2. THE PLC MODEL FORMULATION ................................................................. 16
       2.2.3. THE DOGIT AND IIA ...................................................................................... 19
   2.3. MODEL PARAMETER DISCUSSION ............................................................................... 19
   2.3.1. TTS DATA DISCUSSION ..................................................................................... 20
   2.3.2. MODEL DISCUSSION .......................................................................................... 23
   2.3.3. OTHER POTENTIAL APPLICATIONS ................................................................. 26

3. ASSIGNMENT THEORY AND DEVELOPMENT OF THE MULTIMODAL NETWORK AND SCHEDULE ................................................................. 27
   3.1. FOUR STAGE MODELLING VS. ACTIVITY MODELLING .............................................. 27
   3.2. LITERATURE REVIEW ............................................................................................... 30
       3.2.1. MATSIM OVERVIEW ....................................................................................... 30
       3.2.2. GTHA MATSIM ............................................................................................... 35
       3.2.3. MATSIM APPLICATIONS IN THE REST OF THE WORLD ............................. 40
   3.3. DEVELOPMENT OF MULTIMODAL NETWORK AND SCHEDULE .......................... 44
   3.4. TRANSIT ROUTER-NETWORK IMPROVEMENTS ....................................................... 53

4. INTEGRATION THEORY ............................................................................................................... 60
   4.1. EXISTING GTHA/EMME FRAMEWORK ..................................................................... 60
   4.2. MATSIM UPDATING PROCEDURE ............................................................................. 62
   4.3. PROCEDURE ................................................................................................................ 65

5. MATSIM MULTIMODAL SIMULATION AND VALIDATION ................................................. 69
   5.1. POPULATION CONSIDERATIONS ............................................................................... 69
       5.1.1. MULTIMODAL LEGS ......................................................................................... 70
   5.2. CHANGES TO GTHA PARAMETERS AND PROCEDURES ........................................ 71
       5.2.1. STUCK VEHICLES ............................................................................................ 71
       5.2.2. FLOW, SCALE AND TRANSIT CAPACITY FACTORS ..................................... 72
   5.3. SIMULTANEOUS VS. SEQUENTIAL MULTIMODAL ASSIGNMENT ............................ 74
   5.4. GENERAL RUN RESULTS ........................................................................................... 75
   5.5. TRIP DISTANCE VS STRAIGHT LINE TTS DISTANCE .............................................. 75
   5.6. SCREENLINE COUNTS ............................................................................................... 76
   5.7. LINE BOARDING AND STATION COUNTS ............................................................... 85
5.7.1. LINE BOARDINGS RELATIVE TO TTS DATA ......................................................... 85
5.7.2. STATION COUNTS .......................................................................................... 88

6. FUTURE WORK AND OUTLOOK ........................................................................... 92
6.1. LIMITATIONS ASSOCIATED WITH PLC CHOICE MODEL ................................. 92
6.2. LIMITATIONS ASSOCIATED WITH MATSIM ASSIGNMENT ............................. 93
6.2.1. 2006 POPULATION DATA VERSUS 2011 TRANSIT SCHEDULE DATA ........... 93
6.2.2. GENERALIZED UTILITY ROUTING/PLAN SCORING PARAMETER
       CALIBRATION ..................................................................................................... 93
6.2.3. SUMMARY OF OTHER AREAS FOR IMPROVEMENT ...................................... 94
6.2.4. FUTURE WORK ASSOCIATED WITH MODE CHOICE SWITCHING
       FRAMEWORK ...................................................................................................... 96

7. CONCLUSIONS ...................................................................................................... 96

8. REFERENCES .......................................................................................................... 99
List of Tables

Table 3.1: MATSIM Input Files Required for Multimodal Simulation ......................... 35
# List of Figures

Figure 1.1 The Greater Toronto And Hamilton Area (GTHA) ........................................... 3  
Figure 1.2 A Conceptual Framework For The Integration Of A Mode Choice Model And A Traffic Assignment Model ................................................................. 5  
Figure 3.1 MATSIM Iterative Assignment Procedure For Route Updating Only ............. 31  
Figure 3.2 Transit Stops For 6 Transit Agencies Within the GTHA ................................. 46  
Figure 3.3 A Representation of the Transit Stop to Link Mapping Problem ..................... 50  
Figure 3.4 A Depiction Of Transfer Rules Using Overlapping Approach ...................... 56  
Figure 3.5 A Depiction of Reverse Overlap Conditions for Transfer Rules .................. 58  
Figure 4.1 A Schematic of the Equilibrium Level Of Service Updating Procedure for Policy Testing ........................................................................................................ 67  
Figure 5.1 Percentage of all transit Types by Access and Egress Mode ......................... 71  
Figure 5.2 A Plot of Average Trip Distance by Time of Day ........................................... 76  
Figure 5.3 Westbound Toronto Peel Boundary Screenlines .......................................... 78  
Figure 5.4 Eastbound Toronto Peel Boundary Screenlines ........................................... 78  
Figure 5.5 Northbound Toronto York Boundary Screenlines ....................................... 79  
Figure 5.6 Southbound Toronto York Boundary Screenlines ....................................... 79  
Figure 5.7 Eastbound Toronto Durham Boundary Screenlines ..................................... 80  
Figure 5.8 Westbound Toronto Durham Boundary Screenlines ..................................... 80  
Figure 5.9 Eastbound Peel Halton Boundary Screenlines ............................................. 81  
Figure 5.10 Westbound Peel Halton Boundary Screenlines ......................................... 81  
Figure 5.11 Eastbound York Durham Boundary Screenlines ....................................... 82  
Figure 5.12 Westbound York Durham Boundary Screenlines ....................................... 82  
Figure 5.13 Eastbound York Peel Boundary Screenlines ............................................. 83  
Figure 5.14 Westbound York Peel Boundary Screenlines ............................................. 83  
Figure 5.15 Boardings Over 24 Hour Period for all Agencies by Mode ....................... 86  
Figure 5.16 Boardings Over 24 Hour Period Where Boarding Discrepancies Occur ....... 87  
Figure 5.17 Station Boardings Over 24 Hour Period Finch to Union .............................. 88  
Figure 5.18 Station Boardings Over 24 Hour Period Union to Downsview .................... 89  
Figure 5.19 Station Boardings Over 24 Hour Period Kipling to Bay .............................. 89  
Figure 5.20 Station Boardings Over 24 Hour Period Bay to Kennedy ............................ 90
List of Appendices

Appendix A Parameterized Logit Captivity Model Specification .................................................. 101
Appendix B Configuration File ........................................................................................................ 104
1. INTRODUCTION

The importance of a safe, accessible, and fast transportation system for the vitality of an urban region cannot be overlooked. The economic vitality, quality of life for those who live in the region, and attractiveness of the region to outside visitors and investors are all highly dependent on the region’s transportation system. The challenges associated with maintaining and improving such a system can be significant, particularly with the influence of competing interest groups and financial restrictions, which further constrain decision making. To alleviate some of the challenges and uncertainty associated with making these decisions, transportation modelling plays an important role. Rather than blindly guessing at the impact or outcome of a decision, it is possible to test beforehand what the approximate effect of a decision may be. This allows decision makers to evaluate and compare different potential projects in terms of each project’s potential effectiveness at improving the transportation system. Furthermore, these decision tools provide a means of justification for decision makers to support their decisions to stakeholders.

This work presents a modest improvement on existing modelling techniques used for transportation planning and modelling in the Greater Toronto and Hamilton region from two standpoints. This work presents two main components:

1. A novel mode choice model that incorporates the tendency of travellers to latently exclude all but a single alternative from their choice sets, thereby becoming captive users of that alternative.

2. An improvement on the agent-based multimodal stochastic traffic assignment model developed for use with the MATSIM framework. This agent-based software provides a detailed and disaggregate alternative to the traditional traffic assignment done in the four stage modelling approach, which has historically dominated transportation modelling and planning.
These two models represent travel demand and supply respectively. The interaction between travel supply and demand is important to understanding the overall impact on travel patterns in response to changes to the transportation system. From a modelling perspective, it is therefore necessary to develop a method to combine these two models into a cohesive framework. This thesis will therefore present a conceptual framework for the integration of these two components through an iterative feedback loop. Once completed this integration may also be used as a test scenario for the integration between a traffic assignment model and a full activity based model. The remainder of this introductory chapter will provide some brief background on the study area in question and will outline the data sets that were used to create both the choice model and the traffic assignment model.

1.1. STUDY AREA

The Greater Toronto and Hamilton area (GTHA), as shown in Figure-1.1, is positioned to the northwest of Lake Ontario in the province of Ontario. The region is the largest urban area within Canada. The City of Toronto alone is the fourth most populous city in North America, recently having overtaken the City of Chicago. The GTHA’s current population sits above 6 million residents, with forecasts projecting population growth to approximately 8.6 million inhabitants by 2031 (Metrolinx, 2008). The GTHA has eight local transit systems, each with their own separate transit agency and a single regional transit service operating under the administration of Metrolinx, a provincial government agency that was created to improve the coordination and integration of different modes of transportation within the GTHA. Metrolinx created a regional transportation plan for the GTHA entitled The Big Move (Metrolinx 2008). This document identifies a number of transportation related challenges facing the region as it continues to grow. The combination of population growth, auto-centric development practices, both current and historically inadequate spending and funding allocation towards transportation infrastructure and a disconnect in policies between the transit service providers in the region are factors contributing to increased reliance on private automobiles for travel, which in turn creates congestion. These issues are all indicators that an increase in transportation infrastructure and public transit spending are required so that the GTHA remains an economically competitive, attractive, vibrant and healthy region. This further
brings home the importance of transportation and travel modelling given the rapid growth expected in the region over the next 20 years.

Figure-1.1 The Greater Toronto and Hamilton Area (GTHA) (Metrolinx 2008)

1.2. DATA SOURCES

General Transit Feed Specification

General Transit Feed Specification (GTFS) data is a standardized data storage structure to store public transit schedule information and corresponding geographical information regarding the location of transit infrastructure. Public transit agencies publish their schedule data using this data format, which allows program developers to create applications for this data (Google Developers, 2012). The standardized format of the GTFS data allows these applications to be adapted to the data provided by other transit agencies, eliminating the need for developers to parse different schedule formats. The traditional applications that make use of GTFS data relate to providing schedule information to transit riders, however there are a number of other applications for this data format. A relatively comprehensive list of these applications was compiled by Antirm and Barbeau and is available to the reader for further review (2013). The most relevant application to the work done for this project was found in the realm of
multimodal network development through the use of map matching algorithms to an existing base network. In particular, several applications that use GTFS data to develop a multimodal assignment procedure with the assignment software selected for this work will be reviewed in chapter 3.

EMME NETWORK
The development of the multimodal assignment model required a base network on which to map the GTFS schedule data. The base network that was selected for modification to allow for multimodal assignment is an existing geocoded planning level network developed for the GTHA. The existence of this base network meant that the multimodal network did not need to be developed from scratch (which can be an arduous process for a region the size of the GTHA). Furthermore the base network had already been heavily adjusted to accurately reflect the traffic patterns in the GTHA when combined with the TTS travel demand data discussed below. The base network consists of links and nodes, which are virtual representations of the physical road infrastructure. Links typically represent roadways and nodes often represent intersections, or in the case of a multimodal network, can represent transit stops or switches in a rail line. Each node is associated with an identification number as well as a coordinate value and each link is associated with an identification number, a “from” node, a “to” node (representing the start and end nodes of each link), a length, a free flow speed, a capacity, and a list of modes which are permitted to use that link.

TRANSPORTATION TOMORROW SURVEY
The Transportation Tomorrow Surveys (TTS) is a household based trip diary conducted in the GTHA every 5 years (DMG, 2008a). The survey began in 1986; however, the current shape used for the latest data collection process was initiated in 1996. The TTS survey is a sample survey with a sample size equivalent to approximately 5 percent of the population of the GTHA. The latest available data set was from 2006. This survey information forms the basis for both the estimation of the choice model described in section 2 and the travel demand used for travel assignment presented in section 3 and 5. This consistency of data allowed the integration framework described in section 4 to be
spatially and temporally realistic from a travel demand standpoint, as the travel demand extracted from the TTS data remains consistent between both components of the feedback model. A more detailed discussion regarding the utilization of this data is presented in both chapters 2 and 3 as it pertains to the estimation of the choice model and the travel demand used in the traffic assignments respectively.

1.3. THESIS OUTLINE

The conceptual framework discussed in this thesis will work to combine a mode choice model and a traffic assignment model using a framework similar to the basic relationship presented in figure 1.1. In this framework, the mode choice model produces modal probabilities, which can be fed into the traffic assignment model, and the traffic assignment model produces network level of service attributes, which can in turn be fed back into the mode choice model. This cyclic relationship forms the basis of the framework, which works to capture mode switching behaviour in response to changes to the transportation system.

Figure 1.2 A Conceptual Framework for the Integration of a Mode Choice Model and a Traffic Assignment Model
Each of the individual components of the framework and the framework as a whole are discussed in five distinct sections within this thesis. The first section (chapter 2) presents the mode choice model, which using standard travel survey data captures latent captivity to a single mode of travel. The second section (chapter 3) presents the background information for the MATSIM agent-based multimodal microsimulation platform, as well as the work that has been done to further develop the MATSIM GTHA implementation. The third section (chapter 4) presents a detailed examination of the aforementioned framework for examining mode-switching behaviour in response to changes to the transportation system. The use of the captivity model is ideal for this purpose, as it should assist in distinguishing captive user who will not switch modes from uncaptive users, who are free to switch modes based on changes to the system. The fourth section (chapter 5) outlines the validation of the traffic assignment model by comparing the simulated output with both collected survey data and real world counts. The fifth section of this report summarizes the limitations and outcomes of both the mode choice model and the traffic assignment model and makes recommendation for future work to improve the framework, which integrates these models.
2. MODE CHOICE AND PARAMETERIZED LOGIT CAPTIVITY MODEL

This section will present a literature review covering the challenges and approaches associated with choice set definition in choice modeling while providing a more in depth discussion regarding the development of a specific model formulation which addresses these challenges. A choice model that uses this formulation is then presented and discussed in terms of reasonableness and potential policy applications.

2.1. LITERATURE REVIEW

Discrete choice modelling is a widely used tool in the field of transportation demand analysis, and is a popular method for modelling mode choice. Despite the popularity of traditional discrete choice methods, there are a number of limitations associated with the standard model formulations most commonly used. In the standard multinomial logit (MNL) formulation, a decision maker has a choice set defined for them a priori by the modeller in question. If the model in question is being estimated using stated preference (SP) data, the SP experiment explicitly states the options that are available to an individual making a choice (Train, 2009). This gives the modeller full knowledge of the choice set for each decision maker. However, when revealed preference (RP) data is being used, the choice set of the decision maker is not known a priori. Therefore the modeller must make assumptions regarding the choice set of each decision maker, as they only know for certain that the choice that was selected is included in the decision maker’s choice set. To overcome this challenge, modellers will often use a form of educated guessing through the use of rule-based approaches. These rules are used to define what alternatives are available to each decision maker within the population based on the characteristics of both the decision maker and the potentially available alternatives. Modellers may also circumvent the problem entirely by allowing the full universal choice set (all possible alternatives as defined by the modeller) to be made available to all decision makers (Swait & Ben-Akiva, 1986a). Rule based approaches work well for defining the hard feasibility of a mode. For example, a person without a driver’s license will not have the option to drive when making a trip. However, when less clear rules are used (such as a cut-off distance past which non-motorized travel is no longer an option), outliers in the population may still use these modes past the cut-off point. Furthermore, although the rules may catch many of the totally infeasible modes, the rules are not
capable of capturing all of the reasons why a mode is available or not available for a decision maker, resulting in an improved although not entirely accurate choice set for that individual. It therefore follows that rule-based approaches causes modellers to incorrectly specify the choice set of some individuals. This oversight has been found to be a serious form of choice model misspecification, which results in behaviorally inconsistent and ultimately incorrect models (Stopher 1980, Williams & Ortuzar 1982). The work of Stopher used a set of deterministic rules to identify the captive subsection of a population and then developed two binary choice modes, one for only the non-captive group of the population, and one for the full population, with both the captive and non-captive groups in the sample. In both model formulations, it was assumed that both options were available. Stopher found that the estimated coefficients were smaller and less significant in the model that used both groups as compared to the coefficients in the non-captive only model. This finding suggests that a population with captive decision makers will skew the statistical reliability and overall explanatory power of the model should the modeller assume those decision makers are not captive. Williams and Ortuzar performed a similar analysis using data randomly pulled from an existing model. Williams and Ortuzar investigated the impact of assuming full choice set availability when in fact the availability follows a probabilistic distribution. Their findings were that although the model fit the base case data quite well, when used for policy analysis, the results were often biased towards alternatives that were probabilistically less available.

Further to this, Swait and Ben-Akiva (1986a) found that ignoring the choice set formation causes a significant sacrifice in forecasting robustness for urban areas with rapidly changing social and economic structures or standard changes to household structures over longer periods of time. More generally, improvements to the transportation network will not only change the attributes of alternatives within an individual’s existing choice set and therefore the tendency to select one alternative mode of travel over another, but also potentially open up new choice alternatives to an individual which were not originally available. Likewise as household demographics and land use patterns change and evolve, travel patterns and modal availability will similarly evolve, which suggests the need to consider choice set generation for medium to long term policy analysis models.
A general framework to address the concerns related to choice set generation was first proposed by Manski (1977). Manski establishes discrete choice econometrics as a two-stage process: the definition of the choice problem and then the choice selection. Manski notes that much of the research into choice theory has been focused on the second stage of the process, with the first stage being dealt with using decision rules as discussed above. As a more complete alternative to the traditional approach Manski proposes the following procedure. This model formulation was later named the independent availability model (IAM). First, a universal choice set of size N is defined. This choice set contains all feasible alternatives for any member of the population that is being used to develop the model. The universal choice set is then used to form a set of subsets containing every possible combination of the alternatives within the universal choice set including as few as one alternative and as many as all of the N alternatives. These different combinations of alternatives form the alternatives in a choice set selection model for each individual. The probability that an individual selects an alternative A is therefore a function both of the probability that an individual selects a choice set contains alternative A, and then, given that choice set, the probability that that individual selects A from all alternatives within said choice set.

The IAM approach has a number of problems associated with it, most notably an explosion of the number of feasible choice sets available as the number of alternatives in the universal choice set increases. This is because for an N alternative universal choice set as presented in this work, the number of choice sets is equal to $2^N - 1$. This means that for a smaller universal choice set of say 3 alternatives the number of unique choice sets is a manageable 7, however for a universal choice set with 7 alternatives the number of individual choice sets explodes to 127. This number of potential choice sets is not only cumbersome to deal with but becomes computationally intensive to estimate, particularly when larger data sets are used (Swait & Ben-Akiva 1987b). With the advance of modern computing the computational intensity of such an estimation procedure has likely dropped drastically, however the cumbersome nature of such a large choice set still present a problem. In the case that all potential choice sets are used in the model, the
calculation of the probability of selecting an alternative becomes particularly unwieldy as the number of choices in the universal choice set grows. The probability that an individual chooses an alternative is equal to the sum of all the choice set probabilities that contain that alternative multiplied by the probability of selecting that alternative given the selected choice set. This creates an equation with 64 terms for each probability equation in the case of the 7 alternative universal choice set. While still manageable conceptually, conventional approaches only have a single term, which may limit the applicability of a complete choice set approach for planning purposes in practice. Therefore, while conceptually realistic, for real world policy analysis, this complete choice set model formulation has little to no applicability, particularly considering there are alternative model formulations, which are not quite so challenging to deal with.

In place of gathering full choice set information, different model formulations that examine different potential choice sets have been developed as alternatives to the model proposed by Manski. One such formulation was the DOGIT model, which was concurrently proposed by two research teams in the late 1970s. The model proposed by Ben-Akiva initially in 1977 and then further elaborated on by Swait and Ben-Akiva (Swait 1984, Swait & Ben-Akiva 1987a) was an attempt to address the concerns with the full IAM approach developed by Manski by limiting potential choice sets of an individual to a single choice (the captive choice sets) or the full choice set. Gaudry and Dagenais (1979) proposed an identical formulation to the one presented by Ben-Akiva and postulated that the model serves as a technique for dodging (hence the name “DOGIT”) the independent and irrelevant alternative (IIA) assumption inherent in standard multinomial logit models (MNL). For more information regarding the IIA assumption with respect to the DOGIT model, please refer to section 2.2.1 The DOGIT formulation asserts that an individual may be captive to any mode in the universal choice set, or may select freely from any mode within the universal set. Furthermore, the work of Bordley (1990) suggests that the DOGIT model is a form of the Colombo/Morrison econometric model for consumer-customer loyalty; the probability that a consumer will purchase a product in the current time period is conditional on whether that item was purchased in a previous time. Bordley’s work demonstrates that the DOGIT formulation is identical to
an unconstrained Colombo/Morrison (used for customer loyalty rather than captivity) model under special circumstances, thereby suggesting that unconditional captivity to a single alternative is not required for the use of the DOGIT model. This has major implications for cases where captivity is not necessarily guaranteed. For example, when Gaudry applied the DOGIT to mode choice in Montreal (1980), he inferred that complete captivity did not occur in the population used to develop both a set of MNL and DOGIT models, because the more traditional MNL formulation was found to be statistically superior to the DOGIT formulation for the data set. While Gaudry was likely correct in inferring from his results that no hard captivity exists, there is a possibility that incomplete captivity did exist within the population. Furthermore, Gaudry’s findings could likely be explained by the lack of parameterization of the captivity terms. Furthermore the systematic utility component was specified to be a function of only the price and level of service attributes of the competing alternatives. Socio-economic attributes of decision makers are potentially strong contributors to captivity. For example, income will often dictate if the automobile mode is available as households in lower income brackets may not be able to afford to purchase an automobile. Therefore it is reasonable to assume that Gaudry’s exclusion of socioeconomic attributes influenced his findings. The inclusion of these socioeconomic variables might have lead to finding partial captivity, making the use of the DOGIT statistically permissible in this case, as per the work of Bordley.

Swait and Ben-Akiva (1987b) further elaborated on the approach discussed in their early work by parameterizing the captivity probability terms. They empirically prove that the captivity odds parameters can be fully specified as functions of independent variables related to either the alternative or the decision maker. Swait and Ben-Akiva referred to the parameterized form of DOGIT model as the Parameterized Logit Captivity (PLC) model. The PLC model, like the DOGIT, retains the circumvention of the IIA property inherent in the standard MNL model as well as the N+1 choice sets for N alternatives in the universal choice set.

Both Swait and Ben-Akiva and Gaudry and Willis developed empirical studies using
these formulations, often with the express purpose of comparing the statistical and behavioral validity of these models with more standard multinomial formulations.

The work of Gaudry and Willis (1979) examines the use of the DOGIT formulation for fare type for transit usage and intercity travel mode using a Box Cox transformation on the attributes in the systematic utility. While acknowledging the behavioral potential of the DOGIT model to capture captivity effects, the main discussion points in this work relate to the development of the model as well as the results of statistical comparative tests across the models that were developed. The overall conclusions drawn from the findings were that the DOGIT model formulation is appropriate in some circumstances. The DOGIT narrowly outperforms the standard MNL formulation for the transit fare data set, but for intercity mode choice the DOGIT underperforms compared to MNL, with the authors suggesting that the alternative specific constants in the systematic utility of the DOGIT likely account for captivity. As with the work on mode choice done by Gaudry (1980), it is plausible that through a parameterization of the captivity effects, the DOGIT model’s explanatory power would improve.

Swait and Ben-Akiva developed three different models for a 1977 data set for the city of Maceio in Brazil (1986b): standard MNL, a PLC model and an IAM as presented by Manski. The paper also used market segmentation techniques through the use of energy consumption as a proxy for income in all of the model formulation. The model specification used four alternatives, so the IAM approach is still reasonable with $2^4-1 = 15$ alternatives. Highlights of the results include the importance of income for the study city in determining choice sets and understanding that different model formulations work better for different market segmentation groups (captivity works well for low income, independent availability for high income). This suggests that rather than blindly assuming a choice set formulation such as the captivity model or the independent availability model, some justification for the rationale behind the use of one choice set generation formulation over another should be provided.

Swait and Ben-Akiva also performed a similar analysis for the city of Sao Paulo, Brazil
(Swait, Ben-Akiva 1987a). In this study, they developed both a PLC model as well as a standard MNL model using the 1977 origin destination (O-D) travel data from Sao Paulo. The authors found that the PLC model was statistically superior to standard MNL for the Sao Paulo data. Notably, the PLC model contains different results compared to the MNL in terms of sensitivity analysis with respect to changes in variables such as travel time. In many cases, as expected, there is a smaller sensitivity to change reported for the PLC model. This lower sensitivity to change is reasonable as the model accounts for captive travellers, and as such the change in probability from a decrease in utility will have smaller impacts. In other cases the authors report a higher sensitivity to changes in level of service attributes, which may be explained by the higher magnitude of the parameters in the PLC model coupled with lower captivity odds parameters for specific modes. Other potential explanations relate to the unstable local socio-economic context in Sao Paulo at the time of the data collection.

Since the initial development and applications of either PLC or DOGIT choice model formulations there have been limited attempts to further pursue these models. It is possible that the unavailability of commercial software to estimate such models has restricted their application despite the potential benefits of these approaches relative to more standard approaches. The DOGIT formulation has a closed form likelihood function and can therefore be easily estimated with classical maximum likelihood techniques. This means that many general software packages including GAUSS, R, SAS, STATA are capable of estimating these models. Although the work done with the DOGIT and PLC formulations has been limited, it has not been nonexistent. McCarthy (1997) estimated a logit captivity model examining mode choice for intercity travel using aggregate market shares. His work affirms that overlooking the potential for captivity amongst decision-makers results in biased parameter estimation within choice models, particularly for sensitivity analysis, which is relevant for the evaluation of policy scenarios. More recently, Chu (2009) applied a captivity model which used an ordered generalized extreme value (OGEV) model in the place of the standard logit, which was originally proposed by Fry and Harris (2005). This model was applied to departure time choice modeling for commuting trips in the metropolitan New York City area. The results
suggest that this formulation is superior to standard OGEV models as a result of the inclusion of the captivity model and is certainly a more behaviorally representative of the decision making process relative to the use of a standard DOGIT model with a MNL specification when used to model departure time. Intuitively, the captivity formulation for departure time choice has some limitations in terms of behavioral representations of departure time. Namely, decision makers are more likely to be captive to depart before or after a certain time, rather than being constrained to depart in a single time period. The use of larger time segments alleviates this as many decision makers may have multiple constraints, which may force decision makers into a single time slot. Chu (2010) also applied a logit captivity model for disaggregate destination choice model to distinguish between compulsory work trips and other trips for rural or suburban travel. This work shows that the consideration of captivity in destination choice is associated with significantly improved network modeling capabilities. Future work might involve parameterizing the captivity terms in this destination choice model using socio-economic and demographics variables.

2.2. MODEL FORMULATION

Before establishing the formulation of either the DOGIT or the PLC models, a review of standard random utility theory and MNL model formulation is provided. This discussion is simply an overview and by no means addresses all concerns regarding random utility theory and the implications that it has for choice modelling.

2.2.1. RANDOM UTILITY THEORY BACKGROUND

The PLC model proposed here is in accordance with the Random Utility Maximization (RUM) based approach. This RUM based approach is derived as follows (Train 2009): an individual decision maker n is faced with a choice between a set of alternatives $J_n$. Each choice, $j$, within $J_n$ would result in the decision maker receiving a certain amount of utility, $U_{jn}$, or net benefit from selecting that choice. Assuming the decision maker is rational in terms of their decisions, they will select the alternative $j$ from within $J_n$, which provides them with the greatest amount of utility $U_{jn}$. In the case of modeling the actions of decision makers the utility of each choice $j$ is latent and therefore unknown. This
means that the modeller estimates the utility $V_{jn}$ for all $j$ within $J$ (or the systematic component of utility) of an individual selecting a choice $j$ based on known attributes of both that choice, and the alternative in question. Because $V_{jn}$ is an estimate, it does not equal $U_{jn}$. To account for the discrepancy the utility is broken down into two components such that $U_{jn} = V_{jn} + \varepsilon_{jn}$ where $\varepsilon_{jn}$ is an error term (or the random component of utility). Each error term $\varepsilon_{jn}$ for all $j$ within $J$ is used to represent those attributes of both the decision maker and the alternative $j$, which were not considered in the systematic component of the utility $V_{jn}$. It therefore follows that the modeller will not know the values or the form of $\varepsilon_{jn}$ and therefore treats these terms as random variables. The joint probability density function of the random vector $\varepsilon'_{n} = \{\varepsilon_{1n}, \varepsilon_{2n}, ..., \varepsilon_{Jn}\}$ can be denoted as $f(\varepsilon_{n})$, which can be used in probabilistic equations that aid in defining choice model formulations. Namely:

$$P_{ni} = \text{Prob}(U_{ni} > U_{nj} \forall j \neq i)$$

$$= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i)$$

$$= \text{Prob}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i)$$

This represents the probability that the difference between each random term $\varepsilon_{jn}$ and $\varepsilon_{in}$ is smaller than the observed different between $V_{ni}$ and each $V_{nj}$. This means that the above probability equation can be written:

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i) f(\varepsilon_{n}) d\varepsilon_{n}$$

Where $I(\bullet)$ is a boolean value equal to one if the inequality in the parentheses is true, and zero if the inequality is false. Depending on the formulation of $f(\varepsilon_{n})$ the discrete choice model can take different forms. For certain formulations of $f(\varepsilon_{n})$, the integral takes a closed form, resulting in a relatively easy estimation procedure whereas formulations without closed form integral solutions require numerical simulation to solve. In the case of standard multinomial logit models the formulation assumes that each error term $\varepsilon_{jn}$ is independently and identically distributed using the Type I extreme value distribution. The
probability density function (PDF) for the Type one extreme value distribution \( f(\varepsilon_{jn}) \) is equal to:

\[
f(\varepsilon_{jn}) = e^{-\varepsilon_{jn}} e^{-\varepsilon_{jn}}
\]  

(5)

Resulting in a cumulative probability density function (CDF) of:

\[
F(\varepsilon_{jn}) = e^{-\varepsilon_{jn}}
\]  

(6)

Because each individual error term is identical and independent, the joint PDF for all error terms is the product of \( f(\varepsilon_{jn}) \) for all alternatives \( j \) within the choice set of individual \( n \) and likewise the joint CDF is again the product of each individual CDF for all \( j \) within the choice set of the individual.

This assumption regarding the joint distribution of all error terms results in the well-known formulation as follows:

\[
P_{ni} = \frac{e^{V_i}}{\sum_j e^{V_j}}
\]  

(7)

### 2.2.2. THE PLC MODEL FORMULATION

The PLC takes on a similar form to the standard MNL formulation, as the specification of the utility in the systematic utility functions \( V \) will be identical in cases were the PLC and the standard MNL are being compared. This is achieved by setting the captivity odds parameters equal to negative infinity in the MNL model specification, thereby the probability of selecting the captive option goes to zero. This implies that the full choice set component of the PLC model takes on the same form as the standard MNL. In turn, alternative availability is defined by a set of constraints. These constraints are each defined as a function of both the decision-maker’s attributes as well as the attributes of the alternative whose availability is in question. These constraints are defined as some
function of decision maker attributes or alternative attributes being less than some unknown random variable unique to that decision maker. By specifying the distribution of this random variable for each constraint, the probability that an alternative is available is equal to the joint cumulative probability distribution of all the constraints for each alternative. As alluded to in section (2.1), if all possible alternative availability permutations are considered, the model becomes cumbersome so modellers will manipulate the constraints such that the only options available to a decision maker are the captive options (a single choice) or the full choice set the model resolved into the formulation seen below:

\[
P_{ni} = \frac{\exp(D_i)}{1 + \sum_{j=1}^{J} \exp(D_j)} + \frac{1}{1 + \sum_{j=1}^{J} \exp(D_j)} \frac{\exp(\mu V_i)}{\sum_{j=1}^{J} \exp(\mu V_j)}
\]  

In this case the \(D_i\) term represents the individual n’s mode specific latent captivity to mode \(i\) and \(V_i\) represents the systematic utility of mode I in the standard multinomial logit formulation. To further accommodate systematic heteroskedasticity, the logit scale parameters, \(\mu\), was parameterized. The scale parameter is applied to all parameter estimates in order to scale each of the estimated coefficients to account for the variance in the unobserved portion of the utility. In most typical modelling applications, the scale parameters and the initial coefficient estimates are not reported, rather the product of these two values is presented to facilitate model use. In this case, because the model uses the average income in each decision maker’s geographic zone to account for differences in variance across the population, the initial parameter coefficients, \(\beta\), as well as the scale parameter, \(\mu\), are reported separately for each decision maker. The parameterized components of the equation are:

\[
D_m = \delta_m + \sum \alpha y
\]

\[
\mu = \exp(\gamma z)
\]

\[
V_m = ASC_m + \sum \beta x
\]
Here, $\delta_m$ represents the alternative mode specific captivity constant and $(\alpha y)$ refers to a linear parameterized function. The captivity component of choice probability is specified such that captivity to a specific mode is independent from captivity to any other mode. The $V_m$ term represents the systematic utility of an alternative assuming a standard MNL formulation and the probability of the decision maker not being captive and is therefore free to make a rational choice between all alternatives. The total probability of selecting a mode is the sum of the captive choice probability and the rational choice probability, (where the alternative with the highest utility will be chosen) conditional on no captivity. The conditional probability of no captivity is based on the latent “captivity” to the rational choice, or the full choice set and the captive choice to each specific alternative. If we define the term $D_f$ where $f$ is the full choice set alternative, to be zero, then the exponent of this term becomes 1. This term can be found in the numerator of the second term in equation (8). Such formulation allows that for a number $M$ of alternative modes, we can estimate $M$ number of alternative mode specific constants $\delta_m$, as we hold the non-captivity alternative constant $\delta_f$ and therefore $D_f$ at zero.

The scale parameter, $(\gamma z)$ refers to a linear parameterized function of variables $z$. Since the scale parameter of the logit choice component multiplies the numerator and denominator, the alternative mode specific scale parameters are not identifiable. Therefore, it is not possible to estimate any alternative mode specific constants for the scale parameters. The scale parameters are used to accommodate variables that vary across the population and so are used to capture systematic heteroskedasticity in choice behaviour based on these variations.

In the case of the MNL choice component, the ASC$_m$ represents the alternative mode specific constant. Maintaining the model identification restraints, we can estimate $(M-1)$ number of alternative specific constants for total $M$ alternatives under consideration. The additional $(\beta x)$ component represents a linear parameterized function of variables $x$. 
2.2.3. THE DOGIT AND IIA

Finally, as mentioned in section 2.1 and discussed in the work of Gaudry and Dagenais (1979), the DOGIT formulation is unconstrained by the IIA assumption present in the standard MNL formulation. The IIA assumption is the result of having statistically independent error terms for all alternatives. The implication of this assumption is that the ratio of the probabilities of any two choices remains constant for an individual even if alternative attributes change or new alternatives are added or existing alternatives are removed from the choice problem. This assumption is not always valid, as the introduction of a new alternative, which is similar to an existing alternative, will likely only decrease the probability of selecting the similar existing alternative rather than the other non-similar alternatives. The classic example of IIA violation is the red bus/blue bus scenario, which is outlined in Train (2009). Because of the formulation of the DOGIT model, the ratio of the probabilities of any two alternatives, $i$ and $k$, can be defined as

$$\frac{P_i}{P_k} = \frac{e^{V_i} + e^{D_i} \sum_j e^{V_j}}{e^{V_k} + e^{D_k} \sum_j e^{V_j}}$$

(12)

If a new alternative were added to the universal choice set $J$, or an existing alternative were changed or removed, then the $\sum_j e^{V_j}$ term will change to reflect the modification. In order for the ratio between the two terms to remain constant, the $e^{D_i}$ and $e^{D_k}$ terms must be equal to zero (in this case the model collapses to the standard MNL), or multiple changes to the $\sum_j e^{V_j}$ term result in the same the exact same value as the original $\sum_j e^{V_j}$, or the ratio of the $e^{D_i}$ and $e^{D_k}$ terms terms must be equal to the ratio of the exponents of the sum of the systematic utility terms $V_j$. The latter two conditions are special conditions of the DOGIT where some alternatives are constrained by the IIA assumption and others are not.

2.3. MODEL PARAMETER DISCUSSION

Using the PLC model framework presented above, a PLC model for the GTHA was estimated using 2006 TTS data provided by the Data Management Group at the
University of Toronto (2008). A discussion of the estimation procedure and an analysis of the results of the model estimation are presented here.

2.3.1. TTS DATA DISCUSSION

Before providing a detailed discussion of the actual model itself, it is relevant to provide some further background on the TTS data used to develop the model to provide sufficient context and rational for choices within the model specification. In general, the TTS survey classifies commuter modes into 7 distinct options. These 7 different modes form the alternative set used in the model. As discussed in section 2.1, a probabilistic choice set generation model that uses 7 alternatives and an IAM formulation would result in a total of 127 choice choice sets. Therefore the use of the PLC formulation greatly simplified both the computational intensity of the model estimation, as well making the model much easier to work with, as there are only 8 possible choice sets to deal with, and only a single MNL model

The commuting modes are:

1. Auto driving (AD)
2. Auto passenger (AP)
3. Transit with walk access (TWA)
4. Local transit with auto access (park & ride) (TAA)
5. GO transit with local transit access (GTAA)
6. Go transit with auto access (park & ride) (GAA)
7. Non-motorized (NMT)

Unfortunately, the TTS survey does not collect individual income or household-level income information, which is a severe limitation of the dataset. Income is an important factor in determining vehicle ownership at the household level, as well as the willingness of an individual to pay for a faster mode of travel, both of which affect the modal availability and mode choice of a traveler. The exclusion of income parameters in mode choice models will leave much of the utility of a given mode unexplained, weakening the explanatory capacity of the model. To account for this limitation and capture income
effects on mode choice, occupation specific variables for travel cost were used in the utility function of each choice in the model presented below.

The four occupation groups are:

1. General office
2. Manufacturing
3. Professional
4. Retail/Service

Furthermore, zonal average and median income were used in the scale parameterization of the logit choice model component of the model. The dataset also includes a series of personal and household specific variables, many of which were used in the model development. Appendix A provides a description of all the parameters estimated for the choice model and the corresponding variables that were used to estimate those parameters.

In order to capture the effects of the home location attributes, average and total zonal characteristics were determined based on 2006 Canadian census data (2008). The survey collects information on start time, origin and destination zone, distance between origin and destination and mode of travel for each trip. To determine, transportation level-of-service attributes for all potential trips by each mode (travel time and travel cost by different mode of transportation), deterministic user equilibrium (DUE) traffic assignment models were used. This is particular important as TTS does not collect trip durations, nor is it possible to know the hypothetical trip durations for unobserved modes given the RP nature of the TTS survey. These traffic assignment models were calibrated and validated using EMME/2-based DUE traffic networks for 2006. This DUE assignment procedure is used extensively by local and regional planning agencies and is therefore considered to be best practice for the region. Although this method of assigning level of service attributes has aggregation problems, as will be discussed in greater detail in chapter 3 and 4, no disaggregate agent based multimodal simulation platform has been validated to the extent of the existing EMME/2 assignment procedure. Therefore this
approach was presented as reasonable and the outputs from the DUE assignment were used for model estimation.

Although the PLC/DOGIT models makes use of probabilistic choice set generation, the use of rule-based approaches in conjunction with probabilistic choice sets is reasonable, particularly for so called hard rules of availability. These feasibility rules are commonly employed by the planning agencies in the GTHA and were therefore used to develop the PLC model (Miller, 2007). These rules consist of:

1. Having a driving license and belonging to a household with at least one private automobile makes auto driving feasible.
2. Auto passenger mode is available to all travelers.
3. Transit modes are feasible if the corresponding origin-destination zone pair has reasonable transit travel time (less than 150 minutes in one direction).
4. The non-motorized mode is considered feasible if the distance between the origin-destination pair of the commute is less than or equal to 10 kilometers. For walking the maximum threshold distance is dropped to 3 kilometers and for biking the maximum threshold distance is 10 kilometers.
5. With respect to access modes to transit, it is assumed that the commuters access their closest (by straight-line distance) feasible station with on-site parking.

The limitation of the third and fourth rule based approaches are of course the inclusion of empirically developed cut off points for travel time and distance which may cause the choice sets for outliers to be incorrectly specified. Despite this, the cut off points for both of the rules are conservative and were therefore used a priori in the choice set development phase.

The fifth rule-based approach involves determining transit access station based solely on straight-line distance from the starting location. This approach is behaviorally incorrect as there are many other factors which influence station choice, however this falls outside of the scope of this research and was therefore accepted as a reasonable simplification of the
behavior of travellers, particularly given the aggregate nature of the level of service attributes used for the model estimation. Ideally, a more robust location choice model should be used to determine the access station for each traveller, particularly if disaggregate level of service attributes are to be used in the model estimation. This will increase the behavioral accuracy of the model and permit planning agencies to test the potential effectiveness of station improvements with respect to overall modal share.

After a data cleaning process where missing values were eliminated and the feasibility rules had been applied, 55,927 individual morning commuting trip records remained for the 2006 TTS data set. Once expanded using the TTS weighting factors to account for the entire population, the total number of daily morning commuting trips collected equaled 1,069,252. These weighting factors were developed by the DMG using occupation dwelling counts collected from the 2006 Canadian census (DMG 2008b).

2.3.2. MODEL DISCUSSION

The estimated model proved to be fairly reasonable with the majority of the parameters being both significant and of the correct sign in the systematic utility component of the model. As discussed above, a full list of the parameters and their corresponding variables can be found in Appendix A. The employment classes were all relatively equal in terms of travel cost disutility except for travellers who work in manufacturing who are much more cost adverse than travellers in other employment classes. The relatively constant parameters for the other three employment categories is not ideal as it likely suggests that the model failed to capture the impact of income on mode choice. This model limitation likely stems from the use of generic employment categories with a wide range of income variability within each type. Also it is interesting to note that gender played an important role in determining access mode to regional transit, with women being more likely than men to make use of the park and ride access mode and men more likely to access the regional rail service via local transit than women. This could potentially be caused by the division of household tasks for the standard husband, wife and children family living in the suburbs. In this scenario, the wife will traditionally pick up much of the child care responsibility, which may involve dropping the children off at school for the day, thereby tying up the household vehicle. This forces the husband to access the regional rail service
via transit rather than taking the vehicle, making men more likely to use transit to access GO. Other potential explanations include safety concerns for women at more remote transit stations or transit stations with a considerable walk from the platform to the parking facilities. These factors might influence women to avoid park and ride transit entirely in favour of an alternative mode. To better understand this household bargaining procedure, a comprehensive household activity based model with the different household agents competing for household resources could be developed. Another potential method to examine this would be an alternative model formulation that examines the probability of selecting either the full choice set or a choice set equal to the full choice set less one alternative.

Of more interest are the captivity parameters, which are, for the most part, significant and of the correct sign. The non-motorized parameters present an interesting trend with both the youngest and oldest population segments being less likely to be captive to non-motorized transit. The parameter for travellers in the 18-to-24 age category was found to be insignificant, and was of the incorrect sign given that people in this age category are generally more physically capable of performing non motorized travel, and are limited to non motorized due to lower earning potential. A potential reason for the negative sign (relative to those in the 65 or older age category which was held at zero) is the potential growing trend of young people remaining at home with their parents for much longer than has been historically the case, particularly in suburban areas. While true, many young people still move out of their parents suburban homes for work or school purposes, potentially to a more urban area with greater biking infrastructure and support and amenities that are much more accessible via transit. These conflicting trends will create a high degree of variability in terms of the choice behavior of this age segment, which explains the insignificance of the parameter. To test this hypothesis, the population could be segmented geographically based on density, then, a density parameter could be added to the captivity utility for non-motorized travel. Also viable would be the development of a set of separate models, each of which could be estimated for a single subsample living in one of the density ranges specified. Testing this hypothesis falls outside the scope of this work and therefore was not included here. Another possible explanation for this
phenomenon is an inconsistency within the data set. In models estimated for 1996 and 2001 for purposes discussed below, the age parameter for the under 26 population segment was found to be positive and significant, suggesting that all other age groups are more likely than those over the age of 55 to consider non motorized travel only, despite the negative and non significant parameter for those under the age of 26. The low likelihood of travellers over the age of 55 to only consider non motorized travel may likely stem from mobility issues for the more elderly population as well as more senior positions at their place of employment which require a greater number of work based trips during the work day, thereby requiring the commuter have access to an automobile at the work place.

Intuitively, an increase in auto ownership increases the captivity to transit with automobile access as well as automobile driving and automobile passenger modes. Transit pass ownership for either local transit or regional transit increases the captivity to the respective transit mode. Auto ownership and transit pass ownership do have predictable impacts on captivity to these modes however these variables present a problem for long-term forecasting and policy analysis. Both auto ownership and transit pass ownership likely act as proxies for captivity and fail to highlight the underlying reason for it; a commuter will own a vehicle or purchase a monthly metro pass because of a need to commute to work. However, the purchase of the vehicle or pass itself is simply a response to the underlying reason behind their captivity to a specific mode. The true reasons behind their captivity are much more complex and difficult to collect and understand. This could potentially limit the long term forecasting applicability of the model without first considering the underlying root cause of either the single component choice set or understanding how transit pass and auto ownership functions.

For example if a potential policy reduced the attractiveness of driving for all modes through either an increase in travel time or travel cost, there would be intuitively a drop first in the systematic utility of driving, and as time progressed in the captivity parameters as vehicle owners sell their vehicles. As the current model requires the vehicle ownership values to predict captivity probability and overall probability, the model needs to further
incorporate the underlying reasoning behind the latent captivity. An alternative that will permit the model in its current form to function would be to tie the model into an auto ownership and transit pass ownership model. One such auto ownership model is being developed concurrently to this work for the GTHA, (Duviestein 2013) and a transit pass ownership model for the region has already been developed (McElroy 2009).

Other potential underlying reasons for captivity are likely related to land use factors such as housing location relative to both employment location and/or transit station location, number of expected work based trips, and intra-household interaction where resource allocation and bargaining play an important role in determining the choice set for an individual. Ideally, a more robust version this model would be estimated. This model should include these factors, thereby allowing the model to be used for independent long-term forecasts.

2.3.3. OTHER POTENTIAL APPLICATIONS
As alluded to above, models for 1996 and 2001 were concurrently developed for the region, also using TTC data for those years. These models and data sets were used in conjunction with the existing 2006 model and data set to develop a pooled parameterized logit captivity model for the GTHA between 1996 and 2006. Although not relevant to the subsequent chapters of this thesis, which focus on the 2006 study year alone, this model did present an opportunity to examine the evolution of latent captivity within the GTHA. This pooled model was presented at the annual Transportation Research Board (TRB) conference on January 16, 2013 in Washington DC. The corresponding paper submitted for presentation at the TRB conference is currently under review for publication in Transportation Research Record Part A.
3. ASSIGNMENT THEORY AND DEVELOPMENT OF THE MULTIMODAL NETWORK AND SCHEDULE

This section will examine the development of an activity based disaggregate multimodal assignment framework for the GTHA. The section will begin with a discussion on a comparison of more traditional four stage modelling approaches to activity based approaches, followed by an introduction to the activity based assignment tool used in this work, MATSIM. Then this section presents a review of existing uses of MATSIM within the GTHA as well as multimodal MATSIM implementations internationally. Finally the section will conclude with a discussion regarding the challenges and procedures associated with the development of the GTHA MATSIM multimodal assignment procedure.

3.1. FOUR STAGE MODELLING VS. ACTIVITY MODELLING

With the choice model from section 2 established, potential planning applications of this model may now be examined. Mode choice models are typically the third step in the four-stage planning model, which is used to test the impact and influence of different transportation policies on travel patterns and behaviors within an urban region. The four-stage model is a travel forecasting tool developed to evaluate large-scale infrastructure projects over longer periods of time where both the travel demand and supply are expected to undergo changes. The four-stage model initially creates trip production and attraction for a set of spatial zones within a region (either for the current day or for future projected population growth and land development), and then based on these trips, travel between all potential zone pairs is distributed. Given the trip distribution, each individual making a trip is assigned a mode using either a standard logit model or a more complex model, such as the PLC model described in section 2. Finally the zonal productions and attractions are assigned to travel on the network (for each mode) based on a set of rules regarding their route choice.
While reasonable for long-term travel forecasts, the four-stage model is limited in terms of its applicability to more subtle and complex behavioral policies or for short or medium term planning issues such as management of existing infrastructure. This limitation is in part because four-stage models use aggregate travel production originating from and aggregate travel attraction traveling to the centroids of geographic zonal regions, rather than the behavior of individuals traveling throughout the day to and from unique activity points, each with unique departure times.

An aggregate approach is more reasonable for long-term forecasts, as there is less certainty regarding the increase in travel demand. However, for short-term analysis, where highly detailed information is available, the use of aggregate approaches for travel modelling is inefficient in terms of the use of data, as it aggregates the information into zones. This aggregate approach further limits the application of a four-stage modelling approach for refined policies targeted at smaller segments of the population, as the incremental changes to the travel patterns of these individuals will not be accurately portrayed. The sequential nature of the approach also gives earlier stages of the model little to no ability to adjust to the outcomes of later stages. Although attempts have been made to account for this through the use of iterative feedback loops from a later stage to a previous stage, smaller policies are still much harder to capture realistically and effectively.

This challenge can be illustrated by analyzing how these feedback frameworks function. Demand models examine individual travel patterns and supply models are often at lower resolution and assume large segments of people are all traveling from the same origin to the same destination over the same time period. When both models are set up as an iterative feedback loop, where the demand and supply models feed each other inputs this can be problematic. The supply side model will underutilize the agent-based output produced by the demand models.

Modern demand models treat travelers as unique agents, each with their own unique origin and destination. When inputted into the supply model, the agents will simply be
aggregated to depart from their zone origin centroid and arrive at their zone destination centroid. Furthermore, all agents departing within the same time period start their trip simultaneously rather than using the unique departure times that a departure time choice model produces. Conversely, the demand models will receive aggregate level of service attributes (travel time, trip distance, cost, etc.), which will reduce the accuracy of the predicted travel behavior of the individuals.

These challenges are not as important for longer-term forecasts, as there already exists a degree of uncertainty with population growth patterns. In addition, the potential long-term policies being tested are generally large-scale infrastructure improvements, which will have sufficiently large impacts not to be lost in the aggregation procedure. There are still concerns regarding the use of these approaches for a set of complex and interactive policies, such as congestion pricing/highway tolling or integrated transit fare systems across multiple agencies. These issues become even more problematic for shorter-term policy analysis, or more subtle incremental policy changes, as the changes from a policy will be much smaller. Given that highly detailed data exists for these scenarios, the use of the aggregate four-stage model is not appropriate and an alternative model framework is required.

To account for these challenges, many modelers have turned to agent-based assignment approaches, which disaggregate individuals into unique starting locations and destinations. This is a departure from the traditional four-stage model as trip generation, trip distribution, and trip assignment typically all use aggregate zonal approaches to determine travel behavior. In agent based-travel modelling approaches, travel is instead viewed as a means to an end: people travel to perform activities, such as going to work, shopping, leisure or returning to home from one of these activities. The sequence of activities and the trips between them form activity chains, which populate each agent’s itinerary for the time period for which the model is being developed. Agent-based approaches often use unique vehicle agents and advanced car following models for traffic simulation, which can be computationally intensive, resulting in long run times on large
networks, even with modern super computers. This makes large urban region level scenarios infeasible using the standard agent based approaches.

An alternative to the prevailing agent based approach is an agent based traffic simulation which uses a simplified queuing approach to emulate traffic flow while maintaining unique origins and destinations for all users. This method sacrifices realism for the sake of run time, resulting in a good compromise between an fully aggregate approach used in traditional four stage models and a full microsimulation, used for testing operations over smaller study area or study time period.

One piece of software that uses a queue-based approach is MATSIM: the multi agent transportation simulation (Nagel & Axhausen, 2013). This software allows for large agent-based simulations of multimodal traffic flow in a fraction of the time relative to more typical agent based traffic simulations using car following approaches. These factors make MATSIM ideal for larger planning level questions, while still maintaining the disaggregate nature of the supply side. The remainder of this section will focus on the MATSIM framework itself and projects which use the MATSIM assignment framework, with a specific focus on projects for the GTHA study area as well as those which have developed a full multimodal assignment procedure within MATSIM.

3.2. LITERATURE REVIEW
The literature review section for the traffic assignment is broken into three main components, the first being a basic overview of the MATSIM framework itself from a functionality standpoint, the second looking over existing MATSIM applications within the GTHA, and the third being a brief overview of MATSIM applications which employ transit assignment.

3.2.1. MATSIM OVERVIEW
MATSIM itself is a dynamic toolbox, which allows for a modular approach to performing traffic assignment or agent based travel replanning depending on what the user of the tool is attempting to accomplish. In the case of this work the main goal was to use existing modules present within MATSIM to create a multimodal traffic assignment
baseline. Other researchers can then use this baseline scenario to test specific policies of interest by adding on their own modules and modifying the transportation supply to fit their needs. As such, this work only uses existing route assignment procedure used by MATSIM rather than extending it in any meaningful way. It is still prudent however to briefly examine how route assignment is actually performed within the MATSIM platform as it informs many of the decisions made in the development of the assignment procedure and has implications for the mode choice integration theory discussed in chapter 4.

MATSIM uses a stochastic iterative procedure to assign each agent to the network as described in figure-3.1. It should also be noted that MATSIM also supports iterative replanning of other plan aspects, including departure time, non-work activity choice, activity location choice and mode choice. All of these replanning modules or strategies (including route choice) can be used separately or in conjunction with each other based on what is specified by the modeler. These modules include defining a set of potential plans and selecting a plan at random, optimization of plans (as is seen in the standard route choice presented in figure 3.1) or a more behaviorally realistic approach using an econometric model, where different plans are created and then scored, with the score of a plan reflecting the probability that plan of being selected by the agent (Balmer et al. 2008).

![Figure 3.1 MATSIM Iterative Assignment Procedure for Route Updating Only (adapted from Kucirek 2012)](image-url)
The standard and simplified version MATSIM route choice replanning strategy is as follows:

1. MATSIM finds the path with the shortest travel time for each travel leg, (or more generally, the path for each travel leg that results in the highest possible score or generalized utility; this score function can be predefined by the modeller) for all trips being conducted by each agent in the plans file.

2. The shortest paths are executed on the network and travel times for each agent are collected. MATSIM operates by assuming a first in first out queue-based system to exit links developed by Nurhan Çetin (2006). Vehicles must remain on the link for the free flow travel time and only a certain number of vehicles are permitted to leave the link per time step. Links also have internal capacity constraints, based on the physical space available on the actual road; vehicles will not be able to enter a full link even if they are scheduled to leave their current link during the next time step.

3. A percentage of agents will then consider the new travel times and adjust their planned routes accordingly, with the remaining agents maintaining their existing paths.

Using the updated plans from step three, the second and third steps are repeated iteratively until the final iteration is reached, at which point it is assumed that travel on the network has reached a form of equilibrium. It typically takes 30 to 50 iterations to achieve a relatively stable equilibrium state if route choice is the only plan feature being updated based on existing applications within the GTHA which only examine route updating (Gao 2009). If other aspects of the plan are being modified it may take closer to 200 iterations (Erath et al. 2012). It should be noted that the standard MATSIM procedure does not check for convergence at the end of each iteration; rather, the modeller specifies the number of iterations performed. It is possible to configure MATSIM to perform as many iterations as is required until some pre-specified convergence criterion is reached, however establishing that criterion can be challenging for reasons discussed in the subsequent paragraph. A more complete discussion of the
MATSIM plan selection and replanning modules functionality is presented and discussed in section 4.

It is important to note that the resulting traffic flow from the iterative plan updating procedures is not necessarily user equilibrium traffic flow as defined by the well known Waldrop’s first principle, but rather, an approximation to user equilibrium. The theory of user equilibrium states: each traveller will pick the path that optimizes their travel time, such that, once the system reaches a state of equilibrium; each traveler will not be able to improve their travel time by selecting a different path. User equilibrium theory is applicable in cases where the travel demand is fixed for every hour, such that travellers departing at the same from the same location going to the same destination will all have the same travel time (though not necessarily the same path as two paths may result in equal travel time). In these cases, it is fairly straightforward to implement user equilibrium assignment, or something that closely approximates it, for example, the iterative Frank Wolfe Algorithm used in the EMME traffic assignment model. In agent-based assignment, travelers depart from unique locations at unique times and travel to unique destinations. This disaggregation makes traditional user equilibrium approaches difficult to implement and an alternative simulated method of performing route assignment, such as the plan updating procedure used in MATSIM, is required. Furthermore, unique departure times mean that establishing when the simulation has reached equilibrium state is also challenging, as there are no real tests for user equilibrium for unique departure times (Vovsha, 2008). While the method utilized by the MATSIM framework does not result in actual user equilibrium per se, multiple studies have found favorable comparisons between the approach used in MATSIM and more standard user equilibrium (Gao, 2009) (Fourie, 2010)

It is also important to note that route choice only makes up a single component of the overall plan score, which has implications for the implementation of MATSIM within the GTHA. Travellers in reality may adjust their departure time or activity duration to avoid periods of heavy congestion or change their travel mode entirely. The initial implementation presented in this section only examines route choice updating. Only
considering route choice is reasonable in this case as TTS survey data being used to perform the assignment. This means that departure time choice for each trip is recorded and does not need to be modified by the MATSIM simulation. However if the assignment model presented here is to be used for policy analysis, it is strongly recommended that a joint activity duration model and departure time model of some sort be implemented, either through existing replanning modules contained in the MATSIM core, or through region specific models which can be integrated in a similar fashion to the choice model integration framework discussed in section 4.

The actual plan scoring uses a utility or score, which is calculated with a utility function which is defined by the modeller. This utility function typically uses travel time and travel cost, time spent performing activities, and penalties for arriving to activities late or departing from activities early. The parameters used in the scoring function were initially calibrated by Wenli Gao (2008), with the public transit parameters calibrated by Peter Kucirek (2012) for the GTHA and as such the majority of the parameters will not be altered from the work of Kucirek for this work. The work of Gao and Kucirek will be discussed in greater detail in section 3.2.2.

The required inputs for the MATSIM assignment are outlined in table 3.1. A more detailed discussion of these files and their components can be found in chapter 6 of the thesis of Kucirek (2012).
Table 3.1 MATSIM Input Files Required for Multimodal Simulation Table

<table>
<thead>
<tr>
<th>File Name</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plans file</td>
<td>Contains a list of agents each with their own Id and at least one plan</td>
</tr>
<tr>
<td></td>
<td>Each plan has a list of activities and travel legs</td>
</tr>
<tr>
<td></td>
<td>Each activity has a type (home, work, leisure, etc.), an end time, and a coordinate</td>
</tr>
<tr>
<td></td>
<td>Each travel leg has a mode of travel</td>
</tr>
<tr>
<td>Network file</td>
<td>Contains a list of nodes and links</td>
</tr>
<tr>
<td></td>
<td>Each node has a coordinate and an Id</td>
</tr>
<tr>
<td></td>
<td>Each link has a start node, an end node, an Id a free-flow speed, vehicle capacity, number of lanes, modes and for slime links a type for links such as highways</td>
</tr>
<tr>
<td>Transit Schedule file</td>
<td>Contains list of transit stop facilities and transit lines,</td>
</tr>
<tr>
<td></td>
<td>Each transit stop facility has a coordinate, a reference link, and an Id</td>
</tr>
<tr>
<td></td>
<td>Each transit line has a set of Transit Routes and an Id</td>
</tr>
<tr>
<td></td>
<td>Each transit route has an ordered list of transit stops, the arrival time at each transit stop offset from the departure from the first stop, an ordered list of the links which the vehicle travels on to get to the stop and a list of departures with times offset from the first departure</td>
</tr>
<tr>
<td></td>
<td>Each transit stop is attached to both a transit stop facility and a route.</td>
</tr>
<tr>
<td>Vehicle File</td>
<td>Contains a list of transit vehicles</td>
</tr>
<tr>
<td></td>
<td>Each vehicle is assigned a type and an Id</td>
</tr>
<tr>
<td></td>
<td>Each vehicle type has a length, width, passenger car equivalence, boarding time per person, alighting time per person, door operation information (can passengers enter and exit simultaneously), seated capacity, and standing capacity.</td>
</tr>
</tbody>
</table>

3.2.2. GTHA MATSIM

Three graduate students from 2007 to 2012 have developed the existing GTHA MATSIM framework. Two of the three projects focused primarily on a comparison of the existing
transportation planning route assignment tool most commonly used in the GTHA: EMME. The third project looked to integrate an existing travel demand model developed for the GTHA into the MATSIM framework. EMME (Equilibre Multimodal, Multimodal Equilibrium) was initially developed in the 1970s at the University of Montreal and typically is used for the route assignment portion of a four-stage model. Within the GTHA, the GTA model (REF), a four-stage model used by planning agencies within the region, makes use of EMME for the assignment portion of the model.

The initial work done using the MATSIM assignment package in the GTHA was performed by Wenlei Gao (2009, Gao, Balmer & Miller 2010). These works compared the standard four-stage model for automobile trips as done with the EMME/2 software package and a new MATSIM scenario developed by Gao. Both of these models examined a 2001 GTHA travel network and made use of 2001 TTS travel demand information. The work examines four indicators of simulation performance: average travel time, average travel distance, link volumes and link speeds. Travel time during off peak hours was typically similar between the two procedures, however the simulation in MATSIM resulted in much higher AM and PM peak period travel times, suggesting that MATSIM is more adept at capturing the fluctuation between peak and off peak travel relative to the more uniform travel times across the day simulated by EMME. Trip distances were also similar except for during the midday off peak period where many short trips occur. This difference in trip distance results likely stemmed from all trips in EMME starting and ending at zonal centroids; which results in trip lengths of zero for trips where the origin and destination centroid are the same. Furthermore, the use of the zonal centroids as start and end points in EMME results in a slightly smaller average trip length in the MATSIM simulation MATSIM. Link volumes from both assignment packages were compared to screenline traffic counts. The findings of this comparison were that while both packages gave comparable traffic counts to the screenlines; MATSIM had slightly lower counts relative to EMME. Gao proposed that the lower counts observed with the MATSIM assignment procedure were caused by MATSIM specifying capacity as a hard limit on the number of vehicles allowed through a link per hour. The TTS data used also excludes a number of auto-based trips; therefore the total
number of vehicles simulated is lower than the actual number of vehicles on the road. Conversely EMME can put more traffic through a link than the link’s defined vehicle per hour limit, based on how the volume delay-functions for each link are coded. Link speeds along specific sections of highway are compared to speeds collected on the actual highways themselves. Both EMME and MATSIM simulated higher speeds than those which were collected, however MATSIM was much closer to the recoded values than EMME. The higher speeds from both simulation runs are likely attributed to both the simulations only considering automobile traffic and neglecting all other vehicles travelling on the network. Because transit and in particular, freight, will contribute heavily to traffic flow on freeways and therefore the higher speeds predicted by the simulations compared to reality are expected. Ultimately, Gao shows that the MATSIM platform is viable and similar to both real world data as well as to the existing standard for regional planning within the GTHA from an assignment perspective. Despite this Gao highlights MATSIM’s sensitivity to input quality. This sensitivity is caused by MATSIM performing full 24-hour assignment simultaneously resulting in more traveller interaction and therefore the potential for an increase in the number of travel delays experienced by agents within the simulation.

The second project was performed by Hao (2010, Hao, Hatzopoulou & Miller 2010) and involved the integration of TASHA, a microsimulation model for forecasting travel demand with MATSIM. This work was accomplished within the GTHA using the 2001 travel demand data and corresponding network. The end goal of this work was the development a high-resolution vehicle emissions model. Although the development of an emissions model is a potential application of MATSIM, the relevance of the work of Hao to this thesis is the development of TASHA-MATSIM integration. The Travel/Activity Scheduler for Household Agents model (TASHA) (Roorda & Miller 2003) constructs trip chains for individual agents based on four main steps: activity generation, activity location choice, activity scheduling and household level mode choice. This creates a full day itinerary for each agent, which can then be used in a traffic assignment or route choice software package such as EMME or MATSIM. The final two steps of TASHA are performed iteratively, with new feedback values taken from route assignment software.
This iterative procedure is done until the schedules normalize. Because TASHA was initially developed for integration with EMME, the MATSIM output needed to be aggregated to get zone-to-zone network level of service attributes. Ultimately the integration of TASHA with MATSIM compared favorably to the existing integration between TAHSA and EMME, while maintaining the agent based output from both models, which is ideal for the emission modelling that was done in Hao’s work as well as potentially the more complex policies as discussed in section 3.1. Although the integration of a mode choice model with MATSIM is not the same as the integration of a full activity based model such as TASHA, both the mode choice model integration and the TASHA integration provide a strong first step towards the development of a full econometric activity based model designed explicitly for use with MATSIM or other disaggregate traffic assignment software packages.

In the first two projects involving MATSIM, the only mode considered was the personal automobile. While entirely reasonable for applications such as emissions modelling as the bulk of all traffic in the GTHA is the result of personal automobile use, failure to consider full multimodal passenger assignment inhibits the testing of the effectiveness of mode shift policies and does not allow for the modeller to examine the impact of the construction of new transit infrastructure. Furthermore, the results of auto-only traffic assignment fail to capture the true nature of the transportation system and as such are less accurate than their multimodal counterparts. As such, an investigation into developing a full multimodal (auto & Transit) MATSIM assignment model was undertaken in 2011. It should also be noted that freight traffic also contributes to the overall transportation system and as such a long-term goal of the GTHA MATSIM platform should be the integration of freight traffic.

Initial attempts to develop multimodal transit assignment using the MATSIM platform in the GTHA were performed by Kucirek (2012). The procedure involved obtaining existing transit infrastructure and schedule information, and performing a map matching procedure with the existing automobile network. The work performed by Kucirek was done using temporally different data sets, acting more so as a proof of concept, rather
than a scenario which is useable for testing policies. Furthermore, the procedure meshed a low resolution planning auto network and higher resolution transit infrastructure data without increasing or decreasing the resolution of either set of data to match the other. This resulted in transit stops being aggregated, thereby heavily altering the transit infrastructure relative to the real world infrastructure. This complicated the validation of the assignment as the assignment was performed on only an approximation of the true network. Kucirek identifies this as a major issue associated with this initial attempt to create a multimodal assignment framework and identifies the need to modify the network to match the transit infrastructure resolution.

Fortunately much of the work performed by Kucirek involving GTFS data pre-processing was reusable for the work done in this thesis. Most notably, the ‘generate GTFS frequencies’ procedure outlined and developed by Kucirek was imperative for the development of both his work and the multimodal assignment approach presented here. This frequency generation procedure was necessary as the GTFS data in the region presents the trips made by each individual bus throughout the day. Given this information, the MATSIM schedule would generate individual routes for each individual transit vehicle trip created. This would result in a considerably slower run time as MATSIM would have to process a large number of routes, each with only a single departure. By organizing trips with identical stop profiles and uniform headway, it was possible to categorically group these trips into identical routes, which greatly increased the speed of the schedule generation and overall MATSIM assignment runtime.

Kucirek identified a number of limitations associated with running MATSIM transit for the GTHA, which required attention and action for the purposes of this current work. Due to time constraints Kucirek only included the Toronto Transit Commission (TTC) transit agency, which operates within the city of Toronto, omitting the 8 other agencies that operate within the GTHA. The consequences of this are two-fold, the first being obviously that transit riders who do not use the TTC will not be modelled, and second, transit riders who access the TTC using other transit services, or those who use the TTC to access other transit services will also not be considered. The second concern will
directly affect simulated TTC ridership levels, as there are a significant number of riders who use other transit agencies in conjunction with the TTC for their trips. Furthermore, Kucirek only considered walk access to transit, excluding all other access modes. This resulted in a large number of trips being excluded from the simulation, most notably, auto access trips to the Subway. By not considering these trips, not only will transit ridership be much lower, but also auto traffic that results from the access trip to the transit station will also not be considered, resulting in lower overall congestion. This becomes a bigger problem when the GTHA regional transit service is added to the simulation as a large proportion of their trips involve using an automobile to access train stations. In the case of the work done by Kucirek, excluding non-walk access transit trips was permissible as he was only examining the TTC operations where park and ride makes up a much smaller percentage of all access trips. These factors will all decrease the total ridership levels in MATSIM relative to reality or the EMME multimodal simulation. Despite these factors, there were a higher number of reported boardings on some TTC buses. Kucirek hypothesized that this result was likely caused by a lack of bus speed calibration (buses generally had faster cycle times in MATSIM relative to their scheduled travel times), which results in increased attractiveness of transferring to a bus for a short trip that would be walked in reality. Modelling transit vehicle dwell time (during which passengers board and alight), increasing the number of auto vehicles on the network (thereby increasing overall congestion which will further slow down the transit vehicles) and further network/MATSIM parameter calibration were all identified as potential solutions to this problem.

3.2.3. MATSIM APPLICATIONS IN THE REST OF THE WORLD

Marcel Rieser (2010) developed the framework for MATSIM transit integration and performed and first major application of multimodal MATSIM. The Ph.D. dissertation of Rieser gives an overview of the existing MATSIM functionality, as well as a discussion of the implementation procedure for MATSIM multimodal assignment. The initial attempt (Rieser, Grether & Nagel, 2009) at integrating public transit within the MATSIM framework was developed without the explicit simulation of transit. Instead, for each agent, two potential plans were created, one car plan and one non-car plan. All trip legs
within non-car plans were assumed to take twice as long as traveling by car on an empty network (at freespeed) from their origin to their destination. Agents who selected non-car plans were teleported from their starting location to their destination and assigned a penalty related to the predetermined travel time. The initial results of this experiment were tested on a 10% sample population for Zurich Switzerland and were reasonable, although limited due the approximation of the public transit rather than performing a proper simulation.

The work of Rieser then examines in some detail the procedure for explicitly simulating fully mixed multimodal traffic flow. The simulation procedure defines two new agent types, the Transit Driver and the Transit passenger different from the single agent used in standard auto assignment (here named the auto driver agent). Through the use of Transit Drivers, which are initialized to have plans much in the same way that a standard MATSIM agent has a plan, transit rider agent only need to consider the starting and ending transit stop of their transit trip leg, as the transit driver agent will in effect manage the remaining transit stops and nodes in between for the rider. This allows transit vehicles to be simulated much in the same way that standard passenger vehicles are. Rieser also examines the development of a simplified transit router, which creates a virtual transit-only network, allowing for transfers between different transit routes. Based on the transit router-network it is then possible to use a standard shortest path algorithm to find the best route for an agent as their activity ends. Rieser identifies this approach as a limitation: often travelers will modify their departure time to take a transit service with more frequent headway and suggests this may be an area for future development in terms of transit routing. Rieser then examined the procedure necessary for developing a full multimodal assignment for the region of Zurich, Switzerland, with a mode choice component using the built in MATSIM scoring module. This method of mode assignment resulted in under predicting auto use and over predicting transit ridership, suggesting that a different approach, potentially using econometric demand models may ameliorate the situation.
Based on this work the main MATSIM development team under the supervision of Dr. Kay Axhausen and Dr. Kai Nagel has dedicated significant resources to developing a multimodal activity based assignment model for the city of Singapore (Erath et al. 2012). The authors of this working paper have developed a full activity based model for integration within MATSIM. This involved generating a synthetic population to match census information, as travel information for the full census was not available. Furthermore this procedure involved developing a map-matching algorithm to merge the existing planning level network with the transit infrastructure and schedule data. Finally this work discusses the challenges associated with using such a large data set on a regional network and potential techniques used to overcome the challenges that this poses.

The developers of the activity model used a synthetic population 25% the size of the full population of Singapore, which addressed two main concerns discussed by Kucirek. First, this higher population will address the issue of transit speed being higher than observed values. By increasing the size of the population there will be an increase in the use of smaller side roads by cars. This increase in side road traffic is because when simulating with smaller populations there is less route choice heterogeneity amongst the population, as there are simply insufficient vehicles to fill up all the roads. This heterogeneity will result in many side roads being empty or close to empty, resulting in faster than realistic transit vehicle speeds along these roads. By using a larger sample population, there will be an increase in side road automobile traffic, which will result in much slower speeds for transit vehicles that service these side roads, thereby increasing the realism of the transit simulation. Furthermore this approach attempts to mitigate the concerns regarding transit capacity. Because transit capacity is scaled down to represent the population, a transit vehicle with a real world capacity of 60 will be scaled down to a capacity of 3 for a 5% population sample. This also means that each agent in the 5% sample represents 20 real people. Should a single agent try to board a transit vehicle that is already full, the single agents failure to board the transit vehicle is the equivalent of 20 people not being permitted to board. Because of MATSIM’s use of hard capacity constraints, there is no room for error (such as a penalty for all passengers on an overcrowded bus), and that
agent must wait at that stop for the next transit vehicle. This creates a data expansion problem, as it is unlikely that 20 people were unable to board the transit vehicle at that single stop. If a 25% population sample is used, then the transit vehicle capacity is scaled down to 15, and a single scaled agent represents four people in reality. This scaling range is much more reasonable as when a single scaled agent is unable to board a transit vehicle; this means that at most, 3 “extra” people did not board the transit vehicle at that stop.

The 25% synthetic sample population developed by the Singapore team was then used in the iterative plan, and replan procedure similar to that shown in figure 3.1, however also included a number of other plan aspects in the replanning process. The replanning in the case of this work examines activity duration departure time, mode choice, and secondary (non work or home) activity location choice, as well as the standard route choice discussed above. This procedure uses a summed plan utility that fits well within the existing MATSIM framework when used exclusively for route choice. However, because only a percentage of the population is replanned after every iteration, and the replanning is based on level of service attributes from an unstable non-equilibrium network, this approach may take longer to capture, or fail to detect at all the impacts of subtle policies the same way that replanning mode, departure time and location choice might for the entire population after route choice has reached a state of complete equilibrium. How this full itinerary replanning procedure functions will be further elaborated on in section 4.

This work also used the map matching methods described in the work of Ordonez and Erath (2011) to semi automate the matching of transit stop locations to the planning network. In this case, the methods proposed work for cases where the network is at higher resolution than the stops themselves, which was not the case for the GTHA network and as such a different procedure as discussed in section 3.2 was established to perform the map matching.
The actual assignment procedure for Singapore was found to be highly computationally intensive, even on modern computers outfitted for simulation. This resulted in the suggestion of a number of innovative methods to reduce run time by the Singapore MATSIM team. These suggestions include not performing the mobility simulation after each iteration and modifying the replanning module to use the existing traffic patterns to select the best (highest score) plan, rather than random plan mutation which should reduce the number of iterations required for equilibrium (again, a more elaborate discussion on this is presented in section 4). Although not applicable for the route choice component of this work, these factors are important long-term considerations, most notably because the GTHA multimodal network and population are larger than those used in Singapore, which will increase the computation time for full scale activity based simulations.

Given the limitations of the existing MATSIM multimodal assignment within the GTHA and the success achieved by the MATSIM Singapore team, there has been a push to improve the MATSIM multimodal assignment procedure for the GTHA. The remainder of this chapter will focus on the development of the multimodal network for assignment and procedures to expedite the assignment such that it could be completed within a reasonable timeframe without encountering the identified problems associated with using smaller population samples.

3.3. DEVELOPMENT OF MULTIMODAL NETWORK AND SCHEDULE

This section of work was performed in partnership with both Mohammed Mahmoud and Peter Kucirek. Without their assistance and knowledge of both GIS and MATSIM respectively, this work would not have been possible.

The development of the multimodal network as presented here required two components: an existing network with information regarding the road infrastructure, and information regarding the location of transit stop infrastructure and the transit schedule information which services the transit stops. Although not strictly required, information regarding the link speed and spatial location of dedicated right of way transit (regional trains, subways
and streetcars with partial right of way) is also beneficial for the development of the network.

In particular, the concurrent development of the transit schedule and multimodal network was required such that these two inputs to the simulation were compatible with each other. Initial work performed by Kucirek was met with some degree of success, however limitations in terms of network resolution addressed in section 3.1.2 were identified. A semi-automated map matching procedure was developed, using an existing auto-only planning level network and the transit schedule information for 6 of the 9 transit agencies that operate within the GTHA. This discussion is preceded by an analysis of the existing work that has been established for running MATSIM within the GTHA as well as a review of the development of running MATSIM for multimodal assignment and a discussion of some international applications.

The MATSIM auto assignment network for the GTHA as used by Gao and Hao was created from the existing planning level EMME network for the GTHA. Furthermore, Kucirek also used this network in for his initial attempts at developing a map matching procedure. This network is a regional planning level network and therefore does not contain information about small side streets or smaller intersections. As will be discussed subsequently, this presents a challenge in terms of matching the network to the higher resolution GTFS data.

Given the existing work done using GTFS data sets both in Singapore and in the GTHA for MATSIM multimodal applications, as well as the readily available GTFS data for 6 of the 9 transit agencies within the GTHA, GTFS data was selected as the data source used to develop the MATSIM transit network and schedule. In total there are 25431 stops split between all 6 operators as seen in figure 3-2. This work used only 6 of the 9 operators within the GTHA, as the remaining three operators (Burlington Transit, Oakville Transit and Durham Region Transit) do not have GTFS data available. Therefore, the procedure was developed such that if any of the remaining three transit agencies generate GTFS
data for their transit services, the procedure to introduce the new data into the system would be relatively straightforward, although potentially time consuming.

**Figure 3.2 Transit Stops For 6 Transit Agencies Within the GTHA**

*Figure created by Mohamed Mahmoud*

Because these transit stops are often at variable resolution relative to the planning level auto only network being used (some stops occur along a link, rather than at a node, while other stops may be separated by many links), traditional map matching approaches as discussed in Ordonez and Erath are not applicable. Furthermore, the approach employed by Kucirek simply circumvented the problem (due to time constraints) by matching all transit stops to their nearest link, in turn decreasing the accuracy of the simulation. It was therefore determined that a new map matching procedure which increased the existing auto network resolution needed to be developed. The following procedure created by the author along with the assistance of Mohamed S. Mahmoud and Peter Kucirek for this case is presented here. This procedure formed the content of a conference paper to be
presented at the transportation association of Canada annual conference held in Winnipeg Manitoba in September 2013.

This work attempts to modify the base network to increase the network resolution where applicable such that the network resolution matches the transit stop infrastructure resolution. This procedure involved the creation of four transit stop categories to facilitate the network resolution expansion.

1. Type one stops are stops located at surface terminals or interconnected transfer-points between different transit lines and are clustered together, with each stop cluster representing a single bus bay.
2. Type two stops are surface (bus and streetcar) transit stops located around nodes that are already present in the geocoded network.
3. Type three stops are surface (bus and streetcar) transit stops which are located along an existing network link that is present in planning level network but the link does not have nodes segmenting it at the location of the actual stops.
4. Type four stops are surface (bus and streetcar) transit stops located along road segments that are not defined in the planning-level network. These undefined road segments are usually either new residential subdivisions or minor local roads with transit services not deemed to have sufficient traffic flow to be coded into the planning-level network.

Stops classified as two, three and four stops were occasionally found clustered together into stop groups. Each stop in a stop group represents a stop for one direction of travel at the location of the stop group.

Before the map matching procedure was performed, a subset of the GTFS stops were identified and excluded from the semi-automated procedure. These stops were excluded in order to deal with them on a case-by-case basis. This set included transfer stations (type one stops), stops along transit lines with an exclusive right of way, express routes stops (where the procedure outlined in (Ordonez & Erath, 2011) may be more applicable) and night transit stops and (these services were deemed to be too erratic in their stop patterns and headways to be included in the final simulation). Once these special case
stops had been removed from the full set, stop type identification was performed. The
procedure for type one stops was to identify them manually based on proximity to known
transfer points (subway and regional rail stations, shopping malls, etc.). Once identified,
the type one stops were clustered into stop-groups based on their spatial proximity to
each other. The stop group itself is treated as a single point on the network, with its
location determined based on the centroid of all its constituent stops. To connect these
type-one stop-groups to the rest of the network, up to four new links were potentially
added which connected the rest of the stops to the network: one for each potential mode
(subway, streetcar, bus or train). To determine if a mode specific link was to be added,
the mode of each stop assigned to the stop group was determined and a corresponding
link was added. This link was connected to the nearest node to the stop group with an
outgoing link, which has that mode present in its allowed mode.

The type one stops were then removed or set aside and the remaining stops were spatially
aggregated into similar stop clusters to the type one stops, again based on proximity.
Once clustered, each of these new stop-groups was snapped to a corresponding planning-
level local road only (no highway or ROW transit links or nodes) network node if the
stop-group was within a specific distance. These stops were labeled as type two stops. All
remaining stop-groups were then snapped to a corresponding planning-level network link
if the perpendicular distance between the stop-group and link fell within a specific
distance. These stops were labeled as type three. Links with type three stops along them
were split at these locations creating two new links, replacing the old link. These new
links were assigned new lengths, while maintaining all other link characteristics. Any
remaining stops that were not snapped to the network were flagged as being type four
stops. These stops were examined manually, adding new nodes and links to the network
where applicable or re-snapped to the network with a more lenient tolerance. The type
four stops and stop-groups were then snapped to the new network nodes. The culmination
of this procedure, combines the original planning level network, the stop-groups coded to
existing nodes, the new nodes added as a result of splitting links and the new nodes and
links added to accommodate type one and four stops. The combination of these additions
to the base network forms a hybrid multimodal network, herein referred to as the hybrid
network. Finally, any routes, which had not been mapped to the network (predominantly express routes), were mapped using the map-matching tool developed by Orodonez and Erath.

The hybrid network now contains nodes, links and stop groups. Within each stop group is a set of GTFS stops, each of which in turn needs to be mapped to the end of a link based on the requirements of the MATSIM framework. In order to ensure that the link mapping for that mode was correct, four different network types were developed: a train network, a streetcar network, a subway network and a bus network. The GTFS stops were then sorted based on their mode and a new mode specific stop to node mapping was generated. In cases where multiple links terminate at a single node, a procedure must be developed to determine the correct link to assign any and all stops mapped to that node. For example in figure 3.3-(a) the node/intersection has four links terminating at it and three stops within the node’s corresponding stop group. As can be seen in figure 3.3-(b), the north, south and westbound links each have a stop associated with them, whereas the eastbound link does not have a stop associated with it. This results in the final required formulation seen in figure 3.3-(c). Because it is infeasible to determine the correct link for each stop by visual inspection for all GTFS stops, an automated procedure was developed. This procedure involved determining the stop sequence of each transit line and then using a modified version of the Dijkstra’s shortest path algorithm for the MATSIM framework to find the links between nodes whose stop groups have two consecutive stops in a transit line. The network used for this shortest path calculation was the full mode specific network where the mode was the mode of the transit route in question. As mentioned above, this prompted the need to generate a bus only, subway only, streetcar only and train only network for this procedure (a bus stop cannot be mapped to a subway link). The second stop in the pair of consecutive stops was then matched to the final link in the path between the two stops. This procedure was performed for all transit lines in the GTFS schedule data such that each and every GTFS transit stop was assigned to the end of a link.
The initial stop of a transit route was often the final stop in the route serving the opposite direction thereby ensuring that most if not all stops were accounted for. In cases where two subsequent stops in the same route were mapped to the same node due to the stop grouping procedure and neither of the stops had yet been mapped to a link (as would be the case for the first and second stop in a transit route) a problem arose. These two stops were stored as a pair and then if either of them were mapped to a link subsequently (via the route traveling in the opposite direction), the stop that had yet to be mapped to a link from the pair would be mapped to the same link. Once this procedure had been completed, if there were still stops without any links assigned to them (the first stop in a route with only one direction for example) these stops were simply assigned to the shortest link terminating at the node to which it had been assigned.

Fortunately, the link to stop mapping and the link path between consecutive stops is information that is explicitly required as an input in a schedule file for MATSIM to perform transit assignment and thus once this information had been extracted, it was used to generate the MATSIM schedule file. The schedule file generation also requires the
actual links that the transit vehicle travels on, which are fortunately a by-product of the stop to link matching procedure. This further emphasizes the importance of using mode specific networks; initial attempts to use the full multimodal hybrid network for transit paths resulted in a bus traveling along the subway and regional rail network rather than the links representing streets.

A subsequent validation step to this approach, which was performed with the help of three undergraduate summer research students (Di Niu, Tarek Abul-Fotouh and Mohannad Mohamed). This work was performed in order to validate the paths generated as a by-product of the link matching procedure. Multiple errors are possible including the shortest path between two stops not being the path taken by the transit vehicle to travel between those two stop, and a stop being mapped to the wrong node entirely during the map matching procedure. In order for the summer research students to work effectively, the routes that were potentially problems were identified using two main criteria. Routes that had more than 5 links between any two stops were flagged as having the potential to have incorrect routes, while stops, which were incorrectly mapped, could be identified using a similar procedure. Express routes are typically more spaced out than typical stops and as such the shortest path between two routes may not always be the path that the vehicle takes (which is almost exclusively the case for routes with smaller stop spacing). Conversely, a large number of links between stops on traditional routes could indicate that the transit vehicle is making a detour in order to access an incorrectly mapped stop. Depending on how incorrect the stop mapping is, the transit vehicles may have to diverge significantly from its expected path, resulting in an increase in the number of links between stops. The procedure to solve both problems is relatively straightforward and is as follows:

1. Visualize the routes that were flagged as being potential problems using GIS line shape files.
2. Compare these line shape files against the route maps provided by transit agencies on their respective websites. It should be noted that the routes, which were flagged for having potential problems, often do not actually have any problems associated with them. Therefore the visual validation of the routes and
identification of specific errors is an important component, as the flagging does not identify errors, but rather, potential errors.

3. For routes where the route path was incorrect, identify intermediate nodes on the full network that lie between the two stops with the incorrect travel path. When the schedule is regenerated, and encounters that pair of stops with the incorrect path in the path generation process it will create a path from the first node to the recorded intermediate node and then create a second path from the recorded intermediate node to the final node. In some cases multiple intermediate nodes were required for more complex paths.

4. For routes where stops were mapped to the incorrect location during the map matching procedure, the correct node for each stop can be identified and located on the full network. This will allow the stop to node mapping created in the map matching procedure to be updated with this new node for the culprit stop. Therefore, when the schedule is regenerated with this new stop to node map mapping procedure, the stops will be located at the correct node and therefore the path should hypothetically be correct.

Unfortunately due to time constraints it was not feasible to perform the fourth section of this schedule validation procedure. Therefore, it is recommended that this validation procedure be performed, potentially concurrently with the addition of the remaining transit agencies within the region.

Furthermore, the final schedule required a moderate amount of pruning and cleaning particularly on routes with a small number of departures. The cleaning approach involved removing any routes with no departures or only a single stop and identifying routes with low departures as candidates for merging with another higher departure route. A route was considered a low departure route if it had fewer than 5 departures. Once the low departure routes had been identified, a set of candidate high departure routes for merging was established. Stop sequences were used to determine if a high departure route was a candidate for each low departure route to be merged into. In cases where there was not an exact match (either the high departure route or the low departure route had a few extra stop) a partial match where one of the routes contained all of the stops of the other was
considered to be an acceptable compromise. In cases where multiple matches were found, the candidate high departure route with the largest number of departures was selected. In cases where no candidate high departure route was found, the low departure route in question was removed from the network. While not ideal, these low departure routes both slow down the overall runtime of the simulation and have only a small impact on the simulated transit connectivity and therefore this was deemed to be an acceptable cleaning procedure.

Once the network and schedule had been completed, it was possible to perform the actual assignment, however before this was performed several factors were considered to improve simulation accuracy and reduce the computational intensity and therefore the run time of the procedure.

3.4. TRANSIT ROUTER-NETWORK IMPROVEMENTS
The next thing to consider with respect to running the actual assignment is the routing of the agents on the network. MATSIM makes use of a shortest path algorithm, which considers the results from the previous iteration in replanning a percentage of the population during the subsequent iteration. While relatively straightforward for auto assignment, determining the transfer rules for transit assignment can result in an exponential increase in the number of links required in the network for transfers between transit services. The routing for transit assignment requires the generation of a virtual transit routing network, which creates connections between the unique stops on each route (on which transit vehicles travel), and transfers between two stops on different routes (on which agents transferring between transit services travel). The standard approach used in MATSIM, first developed by Rieser (2010), creates a buffer around each transit route stop and creates a transfer link to almost any other transit route stops that fall in that buffer. No transfers are created from the first stop on a route (as there will be no riders on the transit vehicle at that point) or to the last stop on a route (as that bus will no longer visit any stops so there is no reason for an agent to transfer to that stop). Furthermore, no two stops along the same route will be connected as they are already connected via the actual path of the transit vehicle. This approach typically leads to a fairly large number of transfer links in most applications. In the case of the network used
for the GTHA multimodal assignment there are 409628 transfer links created assuming a buffer of 100 meters around each stop, compared to 48537 actual road or rail links for the transit vehicles to operate on, resulting in a total transit routing network with 458165 links in total.

This may or may not have implications, depending on the shortest path algorithm that is used in the assignment. The standard Dijkstra’s shortest path algorithms performance at worst using Big O notation is $O(|n|^2)$ where n is the number of nodes present in the network on which the routing algorithm is being performed. Fortunately this performance is for the case where all nodes are connected to all other nodes, which does not occur for even the worst-case scenario of the transit router-network within the MATSIM implementation due to the rules established above regarding how transfers are created. The more general performance for cases where not all nodes are connected is equal to $O(|L| \times |n| \times \log(|n|))$, (Fredman & Tarjan, 1987), where L is the number of links, and n is the number of nodes in the network. It therefore follows that removing either nodes or links from the routing network should improve the computational efficiency of performing the shortest path calculation significantly.

Because the number of nodes for the transit router-network is equal to the sum of all stops on each route, it is not feasible to reduce this number past a certain point without sacrificing the realism of the simulation. Conversely, it is quite feasible to reduce the number of transfer links created by the standard transit router generation procedure. Two different procedures were developed to reduce the number of links, while still maintaining the realism in terms of connections that would be used by users in the real world.

The first step in both approaches was to limit the connections created between stops. This was accomplished by only allowing transfers between two stops that occur at the same node. This created some problems, particularly for cases where there were known transfers between two stops which were spatially far apart. Therefore the approach was supplemented by manually creating inter-nodal transfer links on the virtual road network.
used for routing in the auto assignment. These transfer links create connection between nodes, which have stops with known transfers. These transfers typically occur at intermodal or interagency transfer stations, or specific instances (such as transfers between subway lines). This approach is not perfect and is prone to error, particularly on a network the size of the GTHA; it is difficult to know all of the potential cases where transfers over longer distances are liable to occur and therefore some of these transfers may have been missed. Despite these limitations, it is reasonable to assume that the majority of the major transfer points have been accounted for.

The first approach used a similar approach to the one described above; allowing any and all transfers to take place between any two stops at a single node, assuming the transfer from transit stop was not the first stop on the route or the transfer to transit stop was not the last node on the link. Because MATSIM groups transit routes with similar stop profiles into transit lines, it was determined that any two routes on the same line should not have transfers between them. This is because the majority of routes within the same line either represent branch extensions or the opposite direction to other routes in the line. As the shortest path algorithm will not use U-turns, it was determined that no intra-line transfers were required. For the Toronto schedule, this approach resulted in 13689 precoded transfer links, 359356 newly created transfer links and 48537 actual road or rail links, resulting in a router-network with 421582 links in total. The number of road and rail links remains constant throughout all three methods discussed here).

The second approach (which was eventually selected to be used for the assignment) attempts to identify sections of overlap between any two routes with potential transfers, and only allow transfers between the first two stops in the overlap section and the last two stops in the overlap section. This case is depicted in figure 3-5. This approach will reduce the number of transfers created between transit lines which serve the same stops, only allowing riders to transfer as select locations along the overlapping section of both paths. In this figure, there are 6 transfers created using the second algorithm presented here, compared to the 10 that would be created using the first approach, and potentially many more which would be created using a large buffer and the original method proposed by
Reiser and still used in the standard MATSIM framework. Furthermore, this method allows for the explicit identification of routes that overlap but in the reverse direction. Previously, these reverse overlap stops were accounted for by not allowing transfers between two routes on the same line. This more explicit approach will identify other interline cases of overlap and not allow transfers between them. Furthermore, the explicit identification of overlap also allows for transfers between branches of routes in the same line that serve the same direction.

Figure 3.4 A Depiction Of Transfer Rules Using Overlapping Approach Given That Route 1 And Route 2 Are Traveling In The Same Direction

While conceptually fairly straightforward to implement this procedure using visual inspection, because of the large number of potential transfers, an automated procedure was developed. To identify sections of overlap between all of the potential route pairs, the procedure is as follows:

1. Identify an ordered list of all nodes corresponding to each stop that each route visits. Duplicate nodes are permitted, as there may be loops within the route where the same stop or node is visited twice.
2. For every pair of routes with at least one common node in them, compare each node in route 1 to all other nodes in route 2.

3. If the two nodes are the same, that represents a section of overlap. To determine if the overlap is an isolated case, or is part of a larger sequence of overlapping nodes, check to see if the previous nodes in both routes are both equal. If they are, then it represents a sequence of nodes, which overlap, and they should be stored together. Rather than storing the nodes themselves, the index of each node’s position in the list was stored. Using node indices rather then nodes is important, as many routes would visit the same node or stop multiple times; making it difficult to determine which stop was being referred to.

4. All overlap sections with only a single node were then stored separately as potentially representing nodes/stops along two routes traveling in the opposite direction. To identify if these were simply routes with a single point of overlap or represented sections of reverse overlap the next node in the first route was compared to the previous node in the second route and vice versa. If these two nodes were found to be equal in both cases, then no transfer was created between those two routes at that stop. A depiction of such a case can be seen in figure 3-6, where the stops at node ‘i’ will not have any transfers.

5. The first and last stop-pairs were then identified for all remaining sections that overlap and transfers were created between all identified stop pairs.

6. As with the first approach described above, the pre-coded transfer links were also parsed and used to create transfers between all stops incident at each node with a corresponding transfer.

This approach resulted in a significant reduction in the number of links in the transit router-network for the GTHA; producing a network with 13689 pre-coded transfer links, 216627 new transfer links, and 48537 actual road or rail links (the same as the first two methods) resulting in a transit router-network with 278453 links and 49974 nodes/stops (note that the number of stops remains consistent throughout all three approaches outlined here).
Figure 3.5 A Depiction of Reverse Overlap Conditions for Transfer Rules; Stop “I” would have no transfers created between route 1 and route 2

As discussed above, the worst case performance for the Dijkstra shortest path algorithm is equal to $O(|L| \times |n| \times \log(|n|))$ for the network in question. Therefore the worst case performance based on the original MATSIM transit router-network implementation algorithm is equal to $O(692980.0)$ and the final approach used for this work having a worst case performance equal to $O(513268.0)$. The third method should therefore result in a 25% improvement relative to the standard routing performance computation time, while still upholding a realistic transit network with the majority of longer distance connections maintained through the use of precoded links. This is particularly relevant due to the high routing time typically required to run the MATSIM assignment and replanning procedures as discussed by the Singapore MATSIM team. Even on the servers used to perform the multimodal assignment in Singapore (which result in far faster run times compared to the single computer used for the assignment in this case), a single iteration might take as long as an hour. It should be noted that the Singapore model is a full-fledged activity model, which updates mode choice, departure time choice and secondary activity choice along with assignment, which therefore requires a larger number of iterations to reach completion and may drastically increase
the time within each iteration dedicated to replanning, particularly in cases where score maximization approaches are employed. In the case of the GTHA multimodal MATSIM assignment procedure, route choice is the only component that is updated so fewer iterations are required and the replanning phase of the simulation takes considerably less time, however as a result of not being able to perform the assignment on a server, the computation time per iteration is still not insignificant. For further details regarding the runtime of the simulation, please refer to chapter 5.

The following section will present a discussion regarding the integration of econometric choice models with traffic assignment models for the purpose of policy analysis as well the use of activity based frameworks such as MATSIM for policy analysis. The discussion of the actual assignment results can be found in chapter 5.
4. INTEGRATION THEORY

The main purpose of the models discussed in section 2 and 3 is for policy testing. As separate entities, a mode choice model and a traffic assignment model are limited in terms of their ability of testing behavioral response to policies. The choice model requires updated level of service parameters to output changes in behaviour and the assignment model needs to understand how traveler behaviour will respond to changes in the level of service attributes on the network that are the result of a policy or infrastructure improvement. This section will first examine an existing framework for a similar integration of a joint mode and departure time choice model within the GTHA using the EMME assignment model and then review the existing procedure within the MATSIM framework to perform mode choice updating (as well as departure time updating, activity location updating and activity duration updating).

4.1. EXISTING GTHA/EMME FRAMEWORK

The proposed framework was developed by Day (Day 2008; Day, Habib & Miller 2010; Habib, Day & Miller 2010) as a potential application for a joint mode choice/departure time choice model for commuting trips within the GTHA. The model employs a discrete continuous specification with a standard MNL formulation for the mode choice and an accelerated time hazard model for trip timing, which allows for the consideration of the correlation between these two choices. The proposed framework for using the model for policy analysis was constrained by the departure period of a traveler influencing the travel time for each mode a traveler can take. These travel times in turn influences both the mode choice of the individual as well as the departure time choice of that individual. This was not a problem during the estimation procedure as recorded departure times were used to select the appropriate hour-long travel period estimated using the EMME traffic assignment procedure. However, this is much more problematic for policy scenario testing where an agent’s travel time for each mode may change as a result of the policy being tested. This travel time change will force a change in departure time based on the
mode, which will hypothetically result in different mode specific travel times during each of the 24 hour periods, again causing agents to update their mode choice an departure time, and so on. Day identified this cyclic feedback loop and suggested an iterative updating procedure for both mode and departure time to achieve convergence. In this procedure, should the previous iteration’s departure time result in different travel times for any mode within an agents choice set, that individual will have their mode and departure time re-estimated based on the updated travel times associated with their new departure time. These updated departure times and mode choice requires the traffic simulation to be rerun again to obtain new auto travel times for each departure period. For a full description of this approach, it is recommended that the reader refer to the thesis of Nicolas Day.

This approach is analogous to an activity-based feedback approach such as the one used in MATSIM or the TASHA implementation used by Hao, where travel times from traffic assignment are fed back into a behavioral model. The formulation presented by Day is constrained by a number of different factors related to the use of the EMME traffic assignment framework, which is used to obtain the auto travel times used in the process. The first and less serious constraint relates to zone-to-zone auto travel times being used. This approach will aggregates travel times for all agents traveling between the same zonal pair. This becomes a much smaller issue if the zones are relatively small, however is still a minor limitation of this approach as per the discussion in section 3.1 as the travel time accuracy decreases as a result of this aggregation.

The second and more significant limitation of this approach relates to the use of hour-long time bins for travel time. For full day activity modelling, it is necessary to know the travel time at any point during a full 24-hour period. The approach used to determine these travel times involves the use of travel period; with each period being typically an hour long. Travel times in these periods are calculated by performing the assignment for all departures during each period separately. Each travel period traffic assignment produces a unique zone-to-zone travel time matrix. The implications of this are that travel times on all links are held constant over the period, and there is no interaction between
travelers traveling in different periods. In reality, travel time along a link will fluctuate drastically over the course of an hour, based on the number of travelers travelling along the link at any particular instant. This consistency across the hour-long time bin is the result of all agents departing from their origins simultaneously, which results in huge temporal aggregation and negatively impacts the validity of the link travel times simulated. Furthermore, the separation of travellers into different time periods is unrealistic as travelers with longer trips or travelers who in reality depart near the end of their time period would spill over into the traffic generated by the subsequent time period. The failure to consider the interaction between different artificial time periods adversely impacts the realism of the simulation. Both of these considerations are particularly important given the relationship between level of service attributes and travel behaviour, as inaccurate predictions of one value could result in inaccurate predictions of the other value and vice versa.

The simple solution to this concern is to use smaller travel periods, potentially 15-minute travel periods rather than an hour, though for networks the size of the GTHA, this can be problematic as that would result in 96 relatively large travel time matrices. This also aggravates the concerns regarding inter period interaction, as most trips should in theory be affecting the travel times of travellers in subsequent bins.

The use of an activity based traffic assignment model, such as the MATSIM toolkit presented in section 3, alleviates some of these concerns as the assignment approach allows for unique departure times and full 24-hour simulation. This will create dynamic link travel times and removes the artificial traffic separation caused by time periods. As such, it is prudent to examine in more detail the plan updating procedures that are native to the MATSIM framework. The following section will provide a brief overview of the basic MATSIM plan updating procedure built into MATSIM and discuss some of the limitations associated with this procedure.

4.2. MATSIM UPDATING PROCEDURE

The MATSIM updating procedure for departure time, activity duration, and activity location is very similar to the procedure for updating route choice outlined and discussed
in section 3.2.1. This procedure performs a single iteration of full day travel assignment based on existing travel demand provided by the modeler and then randomly draws a percentage of the assigned travelers to perform plan updating. From the sample drawn for plan revision, subsamples are drawn to determine what aspect of their plan is to be updated. Replanning is done through the use of replanning modules, each of which will update some aspect of a plan based on an algorithm. Each of the replanning modules acts independently from each other, which allows for the modeler to specify which replanning modules are used during the simulation, and also allows the modeller to potentially create their own replanning module. Each agent within MATSIM is permitted to have multiple plans, which are selected based on a standard MNL formulation where the utility of each plan is equal to the plan score. Plan selection is done using a random number generator and the probability of selecting each plan. To add additional plans to an agent’s plan choice set, the plan selected from the previous iteration is copied and then modified based on the module selected. In cases where an agent already has a full set of plans (the maximum number of plans an agent can have is specified in the configuration), the plan with the lowest score is deleted to make room for the incoming modified plan (Raney & Nagel 2004). The modules can either mutate the plan randomly or use some form of intelligence to maximize the utility score of the plan itself. There exists a tradeoff between using a less realistic and more random replanning module which is less taxing on the processing power of the computer, and a more realistic utility maximization approach which can be more computationally intensive and therefore time consuming. For example in the case of mode choice, the replanning module can either randomly select a mode from all possible modes that were not used by the traveler during the previous iteration (with each candidate mode having an equal probability of being selected), or can use a probabilistic/econometric choice based model inputted by the model developer to determine the choice of mode conditional on the travel times of the previous iteration (Erath et al. 2012). The computation of the mode probabilities requires the level of service attributes (most notably travel time and travel cost) for all potential modes for each agent being replanned. Because these level of service attributes are updated after every iteration, the replanning module must first query the MATSIM events to determine those values and then compute the probabilities for each mode, for each
agent, which is significantly more complex than simply selecting a new mode at random for each agent.

Two unique options for joint departure time and activity duration exist within the MATSIM core, both following the same practical approach as the theory described for mode choice above. The first module, named the *Time Allocation Mutator* randomly mutates the departure time and corresponding impacts on activity duration with a range around the existing departure time. The second approach, the *Planomat Module* (Meister et al. 2005), uses a genetic algorithm to update the list of activity in sequence to be performed that day, the location of each activity, and activity durations such that the MATSIM score is maximized. This module is highly flexible and allows for soft constraints (with corresponding penalties should these constraints not be met). Example constraints include agents must arrive at work before work officially begins, or all agents within a household must be home in time for supper. While theoretically quite powerful and flexible, genetic algorithms such as the one used in the *Planomat Module*, are often quite time consuming, which becomes more problematic for large regions and populations.

While both the random updating modules and the utility score maximizing modules described above have their own strengths and weakness (computation time versus behavioral realism with respect to utility maximization), neither approach considers full equilibrium travel times when performing the updating. This is clearly the case for random updating modules, as they do not use any level of service attributes in their replanning procedure, however, it is less apparent for utility score maximizing approaches. The Planomat module and the econometric choice model modules described above utilize the results of the previous mobility simulation when determining how to modify an agent’s plan. While not an issue if the simulation is close to convergence, the use of non-equilibrium travel patterns can create some odd plans during early iterations, where the simulation is not close to a stable equilibrium state. These odd plans may result in the simulation taking much longer to converge or potentially cause the simulation to reach a drastically different convergence point relative to what would be achieved if
equilibrium level of service attributes were used. Although outside of the scope of this work, a potentially interesting investigation would be to examine the implications on both the accuracy of the final output and the convergence time of both the within-iteration-replanning and the method proposed below in section 4.3 for both mode choice and potentially a full activity based econometric modelling framework. As a reference point, updating departure time and activity duration typically takes 100 iterations to reach equilibrium (Raney & Nagel 2006), while updating mode choice in conjunction with these values can result in the number of iterations required to reach convergence increasing to 200 (Erath et al. 2012). It should be noted that these convergence numbers are for specific scenarios, and as such it is possible that a scenario may take more or fewer iterations to reach convergence. Despite this, investigating replanning methods, which allow for quicker convergence methods, is relevant given the more lengthy run times of these assignment procedures.

4.3. PROCEDURE

The procedure presented here addresses the concerns raised in the previous section regarding both the accuracy and the convergence time of utilizing non-user equilibrium network level of service attributes in the replanning procedure. This replanning procedure was initially developed for the PLC model outlined in section two, however it could easily be applied to the joint model proposed by Day (REF), or any other econometric choice or activity based model. The initial scope of this thesis had included the implementation of the framework presented here, however complications associated with the development of the multimodal assignment within MATSIM resulted in this portion of this project being shifted to future work. As it stands there are also plans to implement the departure time choice model proposed by Sasic (2012), and to develop a full day econometric activity model for use within either the entire GTHA or a sub region. Both of these choice models will be implemented in a similar fashion to the proposed integration between the PLC model and the MATSIM assignment presented here, with potentially different level of service attributes being selected.

The procedure for performing updating as follows:
1. Given the initial travel demand, an initial plans file is generated and used to perform the MATSIM traffic assignment until equilibrium is reached. The updating procedure of this assignment will only use a route-choice replanning module, holding all other aspects of each plan constant.

2. Based on the equilibrium traffic assignment result, network level of service attributes are drawn from the MATSIM events file, which are fed into the PLC model. In this case travel time and travel cost for all possible modes for each agent commuting during the AM peak are collected and used in the model.

3. Updated modal probabilities are calculated for each agent; agents will update their mode choice based on these probabilities through a random number generation procedure. These new mode selections are used to create a new plans file with updated mode choice.

4. The plans file from the previous iteration (or the initial plans file if this is the first iteration) is compared to the new plans file. If the percentage of total agents with different modes between two iterations is within a reasonable tolerance (for example 1% or fewer agents with different modes), then it is assumed that the simulation has reached a state of equilibrium and the iterative updating procedure is complete. If the difference in modal share is not within a reasonable tolerance, then steps 1 through 4 are repeated with the new plans file replacing the plans file from the previous iteration in step 1.

The set of iterative procedures utilized in this work are succinctly outlined in the graphic below.
In cases where the modification to the plan is not a discrete value such as mode, but a continuous value such as departure time or activity duration, the absolute percent different between the two values within two different iterations can be calculated for all agents and then averaged to determine if the simulation has reached a state of equilibrium. It should also be noted that because of the modular nature of the MATSIM framework, it is possible to specify this post-equilibrium replanning module within the MATSIM configuration such that the simulation will perform the plan updating automatically. This means that instead of having to perform manual calculations and updating after each network equilibrium is reached, the modeler can simply run the simulation a single time, and let the simulation run until plan convergence is reached.
While feasible in the long run, as discussed in section 3, there are still some significant limitations to the MATSIM software package in general as well as specific issues related to the assignment scenario within the GTHA that prevent this framework from being implemented in the short term. These factors will be elaborated on in section 5 and 6.

The following section will provide a review of the validation procedure for the MATSIM run and a brief discussion of a potential application case study for this replanning procedure.
5. MATSIM MULTIMODAL SIMULATION AND VALIDATION

This section presents a discussion of changes to the run parameters from previous MATSIM GTHA applications as well as a discussion of the results of the MATSIM multimodal assignment for the GTHA. The discussion expands on the approaches of Gao and of Kucirek, using evaluation methods similar to those used in their works. The evaluation will examine traffic screenline counts relative to simulated traffic counts, the average straight-line trip distance recorded from TTS compared to simulated trip distance, and reported TTS transit line boardings against simulated transit boardings. The discussion regarding both the parameter changes and the validation procedures will focus on the implications for the integration framework described in section 4. Where major discrepancies in the validation exist, a discussion regarding the potential reasons for those inconsistencies is provided.

5.1. POPULATION CONSIDERATIONS

As discussed in the literature review of the MATSIM framework of both the work by the Singapore MATSIM team and the work of Kucirek within the GTHA, the traditional smaller population sample size of 5 to 10% creates issues related to both side street auto traffic and bus capacity. As such, based on the experience and recommendations of the Singapore MATSIM team, it was decided to generate and use a 25% sample of both auto and transit travellers throughout the day for the assignment presented in this work. The procedure to generate this 25% sample population involved expanding the TTS data using expansion factors, which are calculated by the Data Management Group (2008) and provided alongside the TTS records. These expansion factors are determined based on census data (Statistics Canada, 2008) and used to expand the sample up to a representation of the entire GTHA population. To scale up the initial sample to represent a larger sample, the expansion factors must be adjusted as follows:

\[
Scaled \ Exp \ Factor = Original \ Exp \ Factor \times \frac{New \ Sample \ Size}{100}\% \tag{13}
\]

Both the original expansion factor and the new expansion factor are not integer numbers and because it is impossible to have a fraction of an agent, the expansion factors are
rounded down to the nearest integer during the expansion procedure. This rounding down procedure was selected over rounding to the nearest integer in order to manage the population sample size and therefore the overall simulation runtime, while still expanding the population. This results in a sample population that is smaller than the predefined sample size rather than exactly equal to it.

This expansion procedure resulted in a total of 756047 travellers and 140572 trips taken from the 2006 TTS survey data for a full 24-hour trip, representing 22.5 percent of the GTHA population. These trips include trips for work, shopping, school, entertainment and returning home. The initial dataset had a number of trips or travelers with inconsistent trip chains (no home activity/location, departure from an activity occurs before the arrival, etc). These inconsistent travellers were removed from the data set. Intermodal trips involving cars and another mode of travel were removed from the dataset as MATSIM does natively support these trips. These intermodal trips were predominately in the form of transit trips with car access and egress for commuting to and from work. A more detailed discussion regarding this limitation and plans to correct it can be found below. Also removed were all transit trips in the regions of Durham and Halton. The only transit service simulated in those regions is the regional GO rail/bus service, which is predominantly accessed by car. As the simulation does not include the local transit agencies operating in the area, transit trip records with origins or destinations in those regions were removed from the TTS data set before generating the plans file.

5.1.1. MULTIMODAL LEGS
As discussed above, activities and the travel legs between those activities are used to define the plan of an agent within MATSIM. Travel legs may only consist of a single simulated mode, therefore true multimodal trips, such as transit with park and ride are not by default supported within the MATSIM framework. This oversight is quite considerable as over 10% of all (24 hour) transit trips use some sort of automobile for access or egress as can be seen in figure 5.1. To circumvent this limitation, it is possible to define placeholder activities at transit stations with parking, such that agents will travel by car to the placeholder activities, and then from those activities, travel by transit to their final actual destination. Unfortunately due to time constraints, this could not be
implemented within this work, however the implementation of access and egress modes within the MATSIM framework will be required for the development and execution of the framework discussed in section 4 given that a number of the modes in the PLC model involve transit with auto access.

![Transit Access and Egress Mode](image)

Data from TTS for transit trips occurring into or out of GTHA

**Figure 5.1 Percentage of all transit Types by Access and Egress Mode**

Data extracted from TTS trip records and chart prepared by Peter Kucirek

### 5.2. CHANGES TO GTHA PARAMETERS AND PROCEDURES

As discussed in section 3.2.1, the majority of the MATSIM calibration parameters have been unaltered from the work of Gao and Kucirek as their results were generally found to be reasonable. Despite this, because of the nature of this work, changes to some of the parameters were necessary.

#### 5.2.1. STUCK VEHICLES

The first parameter changed from previous GTHA applications of MATSIM accounts for vehicles that become “stuck” in the vehicle queue simulation. Depending on how the
simulation is configured, it is possible that vehicles will be unable to exit their links. MATSIM is capable of identifying these cases in which a vehicle is at the front of a queue to exit a link and for whatever reason does not exit that link for a configurable amount of time. MATSIM typically handles these cases through an approach initially proposed by Charypar as a method to handle gridlock (2008) whereby vehicles which have been identified as stuck are moved onto the next link, bypassing the capacity constraint on that link, thereby slowly clearing the traffic congestion caused by the vehicle. The alternative approach utilized by Gao, Hao and Kucirek in their work with MATSIM in the GTHA simply removed those vehicles from the simulation. This latter approach results in a different number of simulated vehicles between two iterations within the same run and between different simulations runs with the same population. This creates problems in terms of comparing the outputs of different simulations using the same travel demand data, which is problematic for policy analysis as the level of service attributes of a scenario where a policy is implemented is often compared to the base scenario’s level of service attributes. In cases where the total number of trips differs between scenarios the comparison between the two scenarios does not provide a level playing field. Therefore for this work it was decided to utilize the approach proposed by Charypar as the end goal of the simulation is for use in policy testing.

5.2.2. FLOW, SCALE AND TRANSIT CAPACITY FACTORS
Aside from the changes associated with the handling of stuck vehicles, a number of other parameters needed to be modified within the MATSIM configuration files in order to accurately depict the simulation. The first parameters that needed to be changed were the transit vehicle capacities, which define the number of passengers that are permitted on each vehicle type. In order to ensure that all agents are able to board their selected transit vehicle, each transit vehicle was scaled to approximately 40% of the capacity. This will ensure that no transit vehicles accidently leave an agent behind due to the hard capacity constraints imposed by MATSIM. Subsequent simulation runs may modify the transit vehicle capacity to be more reflective of reality, which may reduce the number of simulated boardings.
The second and third parameters to be modified were the flow capacity factor (FCF) and the storage capacity factor (SCF). The flow capacity factor is a scaling factor, which regulates the number of vehicles permitted to exit a link per unit time; typically vehicles per hour. The scale capacity factor is related to the length of each link and the corresponding number of vehicles that each link can contain. Vehicles take up virtual space on the link and vehicles cannot enter full links. Both of these factors will scale down both the storage and the flow of vehicles per hour of each link to reflect the population sample size. In the works of Gao, Hao, and Kucirek, these values were held at 0.06 and 0.2. The selection of 0.06 for the FCF in previous works relates to the 2001 TTS sample size being approximately 5.8% of the full population. Conversely, the use of a much larger SCF in previous work relative to the sample population size is done to ensure that the queue simulation on each link does not fill up causing stuck vehicles to occur. This is particularly important in cases where links are small as a small scaling factor will result in only a few or even a single vehicle being permitted on each link at any given time. As such a much larger value than what is realistic is used to allow for more vehicles to travel on the link, reducing the number of stuck vehicle incidents. In the case of this work establishing the scaling parameters is not quite as straightforward. It is first necessary to determine the exact size of the population that was simulated. The population expansion procedure results in a population that is only an approximation of the target sample size due to the rounding technique used to create an integer number of agents. As such, the sum of the rounded scaled expansion factors discussed in section 5.1 is divided by the sum of the original expansion factors to obtain the new population sample size. In the case of the 2006 TTS data, the true size of the approximate sample population used for simulation is 22.5%. Despite this, it was decided to use a value of 0.25 for the FCF for these initial tests in order to limit the incidents of congestion and stuck vehicles. By conservatively setting the FCF to be higher than the sample, the simulated transportation system will output less gridlock and will therefore result in fewer stuck vehicles. The SCF was similarly set to the conservative value of 1. This value was selected because the link splitting procedure used during the transit stop map matching outlined in section 3 resulted in a number of relatively short links. These short links do not function well with queue-based simulations, as short links will result in an
increase in the incidence of full links, creating artificial traffic jams. To account for this limitation, a larger than regular SCF was used, permitting more vehicles to enter onto these shorter links. A long-term solution might be to develop an algorithm to join links together and aggregate transit stops where applicable.

5.3. SIMULTANEOUS VS. SEQUENTIAL MULTIMODAL ASSIGNMENT

The multimodal MATSIM assignment procedure follows a standard queue based approach for auto assignment and operates transit vehicles based on the provided schedule file. Because the assignment specified for the GTHA in this work is fully multimodal with both assignments happening simultaneously, mixed traffic transit vehicles do not necessarily follow the schedule. Instead, these transit vehicles will travel on the road network, and experience congestion or free flow travel speeds, depending on the other agents traveling on the road, which may result in either earlier or later stop departures relative to the scheduled values. It is also possible to perform the assignment sequentially however it is important to consider feedback (in terms of link travel time) from the auto assignment when performing the subsequent transit assignment. Failure to consider this feedback assumes the transit operates on empty links and therefore travels at free flow speeds, which is unrealistic and will result in higher than observed cycle times.

In cases where sequential assignment is performed, it is feasible to consider different population sample sizes for both automobile users and transit users. This allows the modeller to reduce the sample size of the automobile users to the size that would be conventionally run for an auto-only assignment, while increasing the size of the sample of the transit riders. Kucirek (2012) employed this approach to reduce the run time of his version of the GTHA multimodal simulation. The use of separate populations has the benefit of allowing a lower overall population to be simulated, cutting down on the runtime of the simulation, which is imperative for a region as large as the GTHA. While more efficient, this approach also has the downside of not allowing any mode switching for agents as the auto driver and transit rider sample populations are two separate populations that cannot be combined. It is impossible to scale down a single agent in a 50% sample transit rider population to fit into a 5% auto driver sample population, as that would involve combining 10 distinct transit rider agents into a single auto driver agent. It
is therefore important to ensure that one full population with a single sample size is used if there is any interest in examining mode-switching behavior. Furthermore, the use of sequential assignment will not allow dynamic interaction between the transit vehicles during the same iteration. While not a huge concern, there are no discernable benefits to performing sequential assignment with equal population sample sizes, therefore simultaneous assignment was selected.

5.4. GENERAL RUN RESULTS
The MATSIM multimodal run was performed on an Intel® Core™ i7-3770 CPU @ 3.40GHz with 16.0GB with a 64-bit operating system and was set to do 30 iterations of multimodal assignment. This resulted in a run time of just less than 24 hours, which is quite reasonable for a network and population as large as those being used. The output configuration file, which outlines all of the parameters specified during the run, is located in appendix B. Despite allowing for overflow capacity on links to account for stuck vehicles, there were still a number of vehicles which were unable to complete their trip over the course of the 30 hours (24 hour plus 6 hours overflow time to account for late agents) assignment period. There were a total of 19000 stuck agents (at 22.5% of a full population, representing approximately 80 thousand travelers), with the vast majority of them being drivers. This issue likely relates to the network not being properly specified (short links and incorrect speeds, vehicle flow throughputs or lengths). Because of the relatively small percentage of stuck travellers (only 2.5% of all travelers simulated), the overall implication of these stuck agents was deemed to only have a marginal effect on the validity of the results presented herein. In terms of the short-term repercussions of the stuck vehicles, the validation results presented here should be treated as preliminary and should see some small improvements once the simulation has been properly calibrated to minimize the incidents of stuck vehicles.

5.5. TRIP DISTANCE VS STRAIGHT LINE TTS DISTANCE
Average hourly trip distances for the auto drive mode were extracted from the MATSIM output and compared to the TTS straight-line mode specific trip distances and Manhattan
distance (1.41 times straight line distance). As can be seen in figure 5.2, the simulated distance typically fell in between the straight line and Manhattan distances. A number of discrepancies did occur whereby the average travel distance was either above the Manhattan distance or below the straight-line distance, which can be partially explained by the TTS straight-line distances being calculated on a zone-to-zone basis. This aggregation error may result in the simulated trip distance going above or below the expected value as these distances are calculated on a point-to-point basis. Furthermore, as the cleaning procedure used on the simulated trip plans was not used on the TTS data used for trip collection, there should be a number of small variations between the TTS records and the simulated values.

![Average Trip Distance by Time of Day](image)

**Figure 5.2 A Plot of Average Trip Distance by Time of Day**

### 5.6. SCREENLINE COUNTS

As with the work of Gao, the validation of the traffic counts will focus on screenlines at GTHA sub-region borders and will aggregate flows over AM peak, PM peak and midday
traffic flow. Due to the 24-hour nature of the simulation, full day counts across the screenlines were also considered. Due to the large number of local screenlines within each sub region only 7 regional boundary screenlines were examined.

In order to compare the simulated screenline counts to the collected data, it is necessary to scale the counts up to match a full 100% population. As screenlines are collected based on post processing the MATSIM events output file, each event that is processed is scaled up to a full 100 percent population based on the agent Id associated with the event. Each link leave event that intersects a screenline counts as a single simulated vehicle crossing the screenline. To scale up these vehicles, each count is multiplied by the original expansion factor and then divided by the rounded scaled to 25% expansion factor. This will not produce integer counts, however in this case it is acceptable as this validation procedure examines general goodness of fit, rather than individual activity. The regional border screenlines used were:

- Toronto-Peel boundary
- Toronto-York boundary
- Toronto-Durham boundary
- Peel-Halton boundary
- York-Durham boundary
- York-Peel boundary

The outputs of the screenline analysis are presented in figures 5.3 through to 5.14.
Figure 5.3 Westbound Toronto Peel Boundary Screenlines

Figure 5.4 Eastbound Toronto Peel Boundary Screenlines
Figure 5.5 Northbound Toronto York Boundary Screenlines

Figure 5.6 Southbound Toronto York Boundary Screenlines
Figure 5.7 Eastbound Toronto Durham Boundary Screenlines

Figure 5.8 Westbound Toronto Durham Boundary Screenlines
Figure 5.9 Eastbound Peel Halton Boundary Screenlines

Figure 5.10 Westbound Peel Halton Boundary Screenlines
Figure 5.11 Eastbound York Durham Boundary Screenlines

Figure 5.12 Westbound York Durham Boundary Screenlines
Figure 5.13 Eastbound York Peel Boundary Screenlines

Figure 5.14 Westbound York Peel Boundary Screenlines
As with the work of Gao, this work found that the simulated screenline counts produced results that were almost always below the cordon counts collected on the actual roadways. The rational for this stems from a number of different factors:

1. The TTS data used to simulate did not include trips originating from outside the region or commercial vehicle trips made in personal automobiles. These missed vehicles will result in reduced counts. In particular during the midday off-peak the simulated counts are significantly lower than the observed counts, which supports the theory that commercial vehicle trips in personal automobiles (which will be predominantly made during standard business hours) make up a significant portion of the collected counts, explaining the lower simulated counts.

2. The exclusion of transit with drive access trips will reduce the number of vehicles on the road overall and the screenline counts that are created by these vehicles should they cross over a screenline. While a small portion of the total auto traffic that occurs is a result of these trips, it is not insignificant, particularly for screenlines in more rural regions.

3. The use of full day simulation will result in events occurring dynamically throughout the day. Due to poorly calibrated network speeds or throughput, vehicles may become delayed, shifting the time that they cross the screenline. To test this hypothesis, the full day simulated screenline counts were compared, which resulted in a much closer match. This suggests that further calibration of the network to increase travel speeds and reduce incidents of vehicles becoming stuck is required. Once this has been completed, the screenline counts for specific times of day should normalize.

The one exception to the lower simulated counts relative to actual counts was on the Durham York boundary, which had significantly higher counts relative to the simulated traffic. These higher counts coincide with much lower counts on the Toronto to Durham boundary, which potentially suggests that MATSIM is routing the travel demand between Toronto and Durham through York (which also explains the higher observed counts on the Toronto York boundary). This is quite reasonable given that the 407 express toll highway passes through the York Durham boundary and because the current
The implementation of MATSIM for the GTHA does not include tolling, the 407 is considered to be a toll free highway in the simulation. Given that the MATSIM routing procedure is based on generalized route score (which is a function of both travel time and travel cost), implementing tolling within the MATSIM framework should, in theory, partially explain the discrepancy between the simulation and the traffic counts.

The simulated traffic counts over the full 24-hour period generally matched up in both directions across each screenline. This suggests that the simulation outputs complete trip chains that are reasonable. There were some minor discrepancies in terms of the collected screenline counts not matching up over the full 24 hour period, which could be the result of travellers crossing the screenline either before the counts started or after they had finished.

5.7. LINE BOARDING AND STATION COUNTS

5.7.1. LINE BOARDINGS RELATIVE TO TTS DATA

The transit analysis section of this report examines simulated transit line boarding against recorded TTS boarding along different transit lines. Both the simulated and the TTS values were scaled up to represent a full 100 percent population. In many cases, there was no direct match between the 2006 TTS route record and the 2012 GTFS transit line and as such it was decided to examine aggregate totals by mode (subway, rail, streetcar, bus) and by agency. The trip records were examined based on full day records rather than time of day as TTS route boarding did not contain the specific time the boarding occurred. The only information available from the TTS was the start time of the trip and as such there will be temporal variability between the simulation boarding time and the TTS trip time, making the comparison between the two values unfavorable. As with the MATSIM population, the TTS records had all trip records with drive access and egress modes removed in order to provide a direct comparison between the TTS records and the simulation output.
The results of this analysis seen in figures 5.15 found that the MATSIM simulation provides a fairly good overall match in terms of boarding numbers, with MATSIM over predicting overall boarding by 9%. A closer examination shows that there is a mismatch between predicted and observed ridership for different agencies and modes. In particular, TTC bus, TTC streetcar, GO bus, and GO train boardings are all over predicted, while York and Hamilton services are vastly under predicted. The rational behind this is that York TTS riders are selecting both GO and TTC bus services for their trips in the MATSIM simulation while Hamilton TTS riders are selecting GO services in the MATSIM simulation. A depiction of this can be seen in figure 5.16. Although not reflective of reality, this simulation output is expected given the standardized fare penalty used for all transit in MATSIM. In reality, the price of GO transit services are considerably higher than all of the local transit providers in the region and therefore the overall score for using the regional transit service should reflect the price difference. This important drawback may also provide an explanation for the over prediction of boardings on many of the local transit services. Because the regional GO transit service does not provide the same service coverage area as the local transit agencies, travelers that make use of the GO service will likely transfer to a regional transit operator to access either the regional transit service or their final destination.

TTC streetcar and boardings were slightly over predicted, and TTC subway boardings were slightly under predicted, suggesting that further calibration between the competing transit modes is required. This also holds true for both the Mississauga and Brampton transit agencies, which also have moderate over predictions of boardings. These over predictions are likely caused in part by the low simulated transit cycle time. For this initial trial run it was assumed that transit vehicles experienced only a single second of delay for each passenger boarding or alighting (or the maximum of the two if boarding and alighting may occur in parallel) and as such the simulated transit vehicle travel time will be faster than the real world equivalent.
Figure 5.15 Boardings Over 24 Hour Period for all Agencies by Mode

Figure 5.16 Boardings Over 24 Hour Period Where Boarding Discrepancies Occur
It should also be noted that the results analysis have some minor limitations as the simulation is integrating two temporally different services; namely the 2006 TTS data and the 2012 GTFS transit schedule and a close approximation of a 2012 multimodal network. Given that transit service improvements have been implemented between 2006 and 2011, the 13% can be partially explained by the higher frequency along surface transit bus routes. The increase in service will make transfers more attractive (lower wait times), which may potentially increase the overall simulated boarding counts.

5.7.2. Station Counts

Despite the issues with the validation procedures used in section 5.7.1, the under prediction of TTC subway usage is quite substantial, suggesting that there may be a problem with the individual subway boarding counts on a station by station level. Therefore a comparison of recorded TTC station entrances (generously provided by the TTC) (TTC, 2007) to the MATSIM simulated station boardings was performed over the full 24-hour simulation period. The comparisons are presented in figures 5.17 through to 5.20.

![24 Hour Station Boardings from Finch to Union](image)

Figure 5.17 Station Boardings Over 24 Hour Period Finch to Union
Figure 5.18 Station Boardings Over 24 Hour Period Union to Downsview

Figure 5.19 Station Boardings Over 24 Hour Period Kipling to Bay
There were a number of discrepancies between the two data sets, most notably an under prediction of boardings at the St. George and Yonge–Bloor stations. These two stations are two of the largest transfer points within the TTC, each connecting the Yonge University Spadina subway line, with Bloor Danforth line. The under prediction of station entrances at these locations specifically suggest that the transfer penalty at these stop locations is significantly too high. Alternatively, there is also a possibility that no transfers were occurring at these locations due to a problem associated with the hard coded transfer links used in the transit router-network. If the cause of the lack of transfers relates to poorly calibrated transfer penalties an improvement to the transfer penalty system may be required. One potential solution would to develop mode-to-mode specific transfer penalties, whereby the penalty for transferring from a subway to a subway would be significantly lower than transferring from a bus to a bus. This solution would encourage subway transfers and discourage bus transfer, thereby addressing the problems found in section 5.7.1. If the lack of transfers is a result of the precoded transfer links at these station not functioning properly, an investigation into properly coding these transfer penalties would be necessary.
links (and all other precoded transfer links) must be undertaken. This investigation falls outside of the time frame of this thesis and therefore was not undertaken here.

The second overall issue with the station boardings relates to the TTC counts not distinguishing between different access-modes, which results in the TTC counts including transit with auto access trips. As such stations with large parking facilities as well as Union Station (which obtains a significant number of GO transit transfers where the initial access mode was auto) will result in MATSIM under predicting boardings relative to the TTC counts. As soon as transit with auto access has been implemented, this source of discrepancy should be resolved.

Despite these limitations, the overall outcome of this work compares favorably to the work of Kucerik (2012) and address part of the concerns related to cycle time misspecification due to insufficient auto traffic on side roads. Despite these advances, there are a number of limitations associated with the work presented here, which will be addressed in the subsequent section.
6. FUTURE WORK AND OUTLOOK

This section presents the limitations associated with this thesis and potential solutions to these limitations, which fell outside of the scope of this work. In particular, the applications of the multimodal assignment model and the integration with the PLC model or other activity based travel behavior models will be examined as these are the areas that require the most work to improve, or in the case of the integration method, to perform initial implementation. Future graduate students or senior undergraduate students can likely address these limitations in the coming years.

6.1. LIMITATIONS ASSOCIATED WITH PLC CHOICE MODEL

The PLC presented in this work addresses an often-overlooked component of choice modelling: choice set generation. The model presented is fairly robust, however in order for it to be truly applicable for policy analysis procedures it must be integrated into a 4-stage model or an activity based model such as MATSIM. The model formulation was limited by the data set that was used for estimation, in particular the lack of income parameters in the MNL portion of the model. An interesting application of the model formulation would be to apply it to a different data set for commuting trips that includes income values and compare the results to the existing model presented in this work. A further limitation addressed in chapter two is the reliance on indicators of the underlying reasoning behind latent captivity rather than the actual factors that influence it. The use of both home and work location characteristics, such as proximity to rapid transit stops or bicycle storage infrastructure would improve the overall model fit and allow for the modeller to test the impact of transit or non-motorized infrastructure projects and their impact on overall mode split and modal captivity using a feedback tool similar to the one described in section 4. As the model was estimated using 2006 data, it would also likely be prudent to estimate a 2011 model using the 2011 TTS data once it has been made available, in particular given problems associated with temporal misalignment, as discussed in section 6.2.1
6.2. LIMITATIONS ASSOCIATED WITH MATSIM ASSIGNMENT

As the majority of the work performed in this thesis was focused on the implementation of the traffic assignment model, there are a significantly larger number of limitations associated with this work relative to the PLC model. These limitations are addressed here.

6.2.1. 2006 POPULATION DATA VERSUS 2011 TRANSIT SCHEDULE DATA

The assignment procedure presented here utilizes a modified version of the 2006 EMME auto network. This modified version of the network is the basis for the 2011 GTHA network currently being developed by the travel-modelling group at the University of Toronto. As such the modification created to the network, in particular with respect to the creation of new roads to deal with type 4 stops discussed in section 3, results in the hybrid network being a closer depiction of the 2011 road network than the 2006 road network. Furthermore, this approach utilizes 2011/2012 transit schedule data extracted from GTFS files provided by the different transit agencies within the GTHA. Conversely, the travel demand population data used for both the PLC and for the travel assignment was taken from the 2006 TTS survey. This temporal mismatch in data creates problems in terms of testing the validity of the assignment procedure and limits the applicability of using the framework presented in section 4 for policy analysis. Fortunately the procedure to process the 2011 TTS data is identical to the 2006 TTS data processing procedure. This will allow a new PLC or other econometric model to be estimated and a new 2011 plan file to be generated. Once this has been performed, the assignment can be further validated using the procedures presented in chapter 5.

6.2.2. GENERALIZED UTILITY ROUTING/PLAN SCORING PARAMETER CALIBRATION

The initial parameters developed by Gao for the GTHA MATSIM auto assignment plan scoring were assigned based on the predefined values used in standard MATSIM tutorials. An attempt to further calibrate these parameters to match the intricacies of the GTHA region would likely improve the overall routing procedure. This is particularly
important for generalized cost routing for transit as travellers may decide to select a set of transit services that do not have an additional fare associated with them. This is also important for cases such as the 407 express toll highway, which operates within the GTHA. Toll highways and other tolling systems will influence choice and may cause travelers to not select certain routes to avoid extra cost. In order to accomplish the task of considering all the factors that influence route choice, a full-scale econometric model of route choice should be developed and implemented within the MATSIM framework for the GTHA. This will not only improve the accuracy of the simulation in terms of behavioral realism and network calibration, but also allow for further testing of road pricing scenarios.

In order to accomplish this, a fare calculator would need to be developed to accurately determine the fares incurred by an agent who is transferring. Such a fare calculator is currently under development within the research group by undergraduate student Ankit Bhardwaj and is near ready for implementation. Because of the complexity of the different transit fare systems within the region, and their potential interaction with each other, this fare calculator must be sufficiently robust to distinguish between different fare collection models (flat fare, zone based fares, distance based fares, routes with special extra fares) and the potential fare interactions between different agencies (discounts for transferring from one agency to another, etc.). Furthermore, this transit fare calculator should allow modellers to modify the fare regimes of the different transit agencies in order to test the impacts of changes to fare policy or fare coordination amongst different agencies. The integration of this fare calculator with the routing such that instead of using a base flat fare, each agent using transit will be assigned a unique fare based on the services that they use, will result in a superior split of transit riders between all agencies and all modes within each agency relative to the boarding results presented in section 5.3.

### 6.2.3. SUMMARY OF OTHER AREAS FOR IMPROVEMENT

Other potential areas for improvement within the MATSIM GTHA assignment tool include:
• Adding in intermodal trips such as transit with auto access/egress. Their inclusion will impact overall network realism (as there are more vehicles on the road) and will allow for more mode choices in replanning modules or procedures than simply public transit and private automobile.

• The current lack of schedule data for three transit agencies needs to be addressed. The procedure for this work is established, however the work to do so is non trivial and requires a significant amount of manual editing and network/transit system knowledge. There is also the question of GTFS format adoption, as the three missing transit agencies have not yet adopted the GTFS data format to store their schedule information.

• During the schedule generation process, a number of routes were removed due to one or more of their stops being outside the study boundary and therefore not considered within the simulation. Many of these routes were commuter rail lines, which terminate at Union station, and therefore represent a significant transit service. Efforts to reintegrate these services back into the assignment model should be made to better represent the transit service that operates within the region.

• Further configuration of the transit operations in terms of the transit wait time at stops is required. The stop dwell time should be a function of the number of passengers boarding and alighting from the vehicle at each stop and should not be linear as in the case now (one passenger alighting takes more time per person than two or three alighting passengers).

• The comparison of travel times to traditional assignment software fell outside of the scope of this work. This analysis should use an approach similar to that employed by Kucerik (2012) and Gao (2009) which extracted average hourly travel times from the EMME 2006 assignment procedure for transit and auto respectively. This analysis should likely be performed for the 2012 TTS data once it becomes available as the temporal mismatch between the transportation supply and demand in this thesis hinders the comparison.

• Due to time constraints, the number of runs of the assignment that could be completed was minimal. As such, an investigation into the effect of modifying the run
parameters (in particular the flow and scale capacity factors) would potentially improve the accuracy of the simulation.

- Although only mentioned briefly, the inclusion of a freight travel assignment model within the MATSIM GTHA assignment package would provide further room for policy development. This work falls well outside the scope of the thesis presented here, however the inclusion of freight should be considered a long-term goal for the MATSIM GTHA project.

- Much of the code written to process the data should be shared with the rest of the MATSIM community. Many of the advances, particularly in transit routing and map matching could have large-scale implications with respect to the development of multimodal traffic assignment with the MATSIM toolkit. This code should be organized and compiled in order to allow other research teams to benefit from the advances made on this project.

### 6.2.4. FUTURE WORK ASSOCIATED WITH MODE CHOICE SWITCHING FRAMEWORK

The conceptual framework for mode choice switching is currently in the preliminary stages of investigation and cannot be implemented until the problems associated with both the PLC and the multimodal assignment models are addressed. As such, the medium to long term goals associated with this work aside from the eventual implementation of the framework for policy analysis include:

- Performing a similar equilibrium updating procedure using a departure time choice model and a mode choice model simultaneously through either the use of a joint model, or two separate econometric models.

- Investigating the performance of such an updating framework relative to the Planomat feature developed by the MATSIM team.

- Developing a fully econometric activity based model, which could be integrated into the framework to update the full itinerary of the agent rather than simply the agent’s mode choice for commuting trips.
Although there is still a considerable amount to be done to achieve a fully operational framework, the work presented in this thesis lays a solid foundation for activity based modelling and policy analysis.

7. CONCLUSIONS
This thesis presents a parameterized logit captivity model for the GTHA and discusses potential applications for such a model in terms of predicting mode switching. A PLC model for the GTHA was estimated and found to be reasonable as the model parameters were typically significant and were of the correct sign and reasonable magnitude. PLC models address an aspect of choice modelling which is often overlooked: latent captivity to a mode of travel. As a result of this characteristic, the PLC model has strong policy applications, particularly in the realm of tracking mode switching behaviour. Because the PLC model asserts that there is a probability each individual is captive to a mode based on longer-term decisions or personal characteristics (automobile ownership, transit pass ownership and age), this model formulation provides a stronger predictor of mode switching behaviour based on short-term changes to level of service attributes. Captive users will not change modes, even if other modes become more attractive or their selected mode becomes less attractive; behaviour which is often overlooked in traditional model formulations. The consideration of captivity will allow transportation modellers to predict the short and medium term impacts of their policies on mode switching behaviour which can result in modellers providing more accurate information to decision makers.

This work then addresses improvements to the dynamic agent based 24-hour multimodal assignment model using the MATSIM assignment tool for the GTHA. The agent based approach utilized by MATSIM results in a drastic improvement over the conventional aggregate assignment approach. By looking at travel as unique departures from unique origins to unique destinations, with each traveller leaving at a unique time, this approach addresses many of the limitations associated with the conventional approaches. These approaches aggregate both spatially and temporally, which reduces the overall accuracy and realism of the simulation. The assignment procedure for transit is also addressed through a simplification of the routing network, which results in reasonable run times,
even for large populations. The implication of this is that more subtle policies such as road pricing scenarios or the selection of a new transit stop location can now be tested with a greater degree of precision within a reasonable time frame.

A framework that integrates the PLC model, or any other travel behaviour model, with the traffic assignment framework, is presented and compared to existing models from a theoretical standpoint. This integrated modelling structure for mode switching is ideal as it combines the benefits of both the PLC model to account for mode switching resistance and the MATSIM assignment model for increased accuracy with respect to travel time and travel cost. The model structure also improves on existing activity based models by performing the assignment and mode switching sequentially rather than simultaneously, resulting in the mode switching behaviour being the result of equilibrium level of service attributes, unlike the standard approach utilized within the MATSIM framework. This framework has wider implications and can be used to test models such as departure time choice, activity duration, and activity location choice and of course mode choice, as was suggested by this thesis. This full-based activity tool can be used to test how travellers are likely to respond to policies and changes, rather than the conventional methods which force travelers to respond by only changing a single aspect of their trip.

This work builds upon the work of others in the modeling of GTHA transportation analysis. This work provides an improved mode choice model relative to conventional methods, and further development of the public transit and automobile travel assignment modeling in MATSIM for both the GTHA and the overall MATSIM software package. Finally, this work suggests a novel framework for integration of the mode choice model with the traffic assignment in a way that will allow changes in one to affect the functioning or behavior of the other. There are huge potential benefits in terms of the level of detail available for policy makers.
8. REFERENCES


Gao, W., Balmer, M., & Miller, E. J. (2010). Comparison of MATSIM and EMME/2 on greater Toronto and Hamilton area network, Canada. Transportation Research Record: Journal of the Transportation Research Board, 2197(1), 118-128.


Metrolinx. (2008). *The big move: Transforming transportation in the greater toronto and hamilton area (GTHA)*. ()


Nagel, K., & Axhausen, K. (2013). *MATSIM. MATSIM.org*


APENDIX A

Parameterized Logit Captivity Model Parameter Table

<table>
<thead>
<tr>
<th>Captivity Parameters</th>
<th>Estimates</th>
<th>Est./s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Specific Constants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automobile Drive Mode</td>
<td>-1.7095</td>
<td>-155.761</td>
</tr>
<tr>
<td>Automobile Passenger Mode</td>
<td>-6.0315</td>
<td>-59.236</td>
</tr>
<tr>
<td>Local Transit with Non-Motorized Access Mode</td>
<td>-3.6184</td>
<td>-360.766</td>
</tr>
<tr>
<td>Local Transit with Automobile Access Mode</td>
<td>-10.3431</td>
<td>-154.173</td>
</tr>
<tr>
<td>Regional Transit with Local Transit Access Mode</td>
<td>-0.6278</td>
<td>-105.797</td>
</tr>
<tr>
<td>Regional Transit with Automobile Access Mode</td>
<td>-5.019</td>
<td>-75.658</td>
</tr>
<tr>
<td>Non-Motorized Transit</td>
<td>-4.1271</td>
<td>-317.697</td>
</tr>
<tr>
<td>Modal Specific Parameters for Non Motorized Travel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger than 26 years</td>
<td>-0.0236</td>
<td>-1.134</td>
</tr>
<tr>
<td>26 to 35 years</td>
<td>0.5889</td>
<td>41.023</td>
</tr>
<tr>
<td>36 to 45 years</td>
<td>0.8023</td>
<td>59.199</td>
</tr>
<tr>
<td>46 to 55 years</td>
<td>0.5291</td>
<td>37.847</td>
</tr>
<tr>
<td>56 years or older</td>
<td>0</td>
<td>*</td>
</tr>
<tr>
<td>Modal Specific Parameters for all Local and regional Transit Modes with Automobile Access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Vehicle in Household</td>
<td>5.4607</td>
<td>82.758</td>
</tr>
<tr>
<td>2 or more Vehicles in Household</td>
<td>7.5132</td>
<td>113.851</td>
</tr>
<tr>
<td>Modal Specific Parameters for Local Transit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Transit Metro Pass ownership</td>
<td>5.2994</td>
<td>542.109</td>
</tr>
<tr>
<td>Modal Specific Parameters for Regional Transit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional Transit Metro Pass ownership</td>
<td>0.8266</td>
<td>129.006</td>
</tr>
<tr>
<td>Modal Specific Parameters for Automobile Driver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Vehicles in Household</td>
<td>2.6142</td>
<td>293.169</td>
</tr>
<tr>
<td>3 or more Vehicles In Household</td>
<td>3.057</td>
<td>328.997</td>
</tr>
<tr>
<td>Modal Specific Parameters for Passenger Driver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Vehicles in Household</td>
<td>3.2228</td>
<td>32.313</td>
</tr>
<tr>
<td>2 or more Vehicles In Household</td>
<td>3.6621</td>
<td>36.543</td>
</tr>
</tbody>
</table>
### Systematic Utility Parameters

<table>
<thead>
<tr>
<th>Alternative Specific Constants</th>
<th>Estimates</th>
<th>Est./s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automobile Drive Mode</td>
<td>8.2121</td>
<td>17.642</td>
</tr>
<tr>
<td>Automobile Passenger Mode</td>
<td>5.6898</td>
<td>12.23</td>
</tr>
<tr>
<td>Local Transit with Non-Motorized Access Mode</td>
<td>8.5685</td>
<td>18.404</td>
</tr>
<tr>
<td>Local Transit with Automobile Access Mode</td>
<td>7.6126</td>
<td>16.353</td>
</tr>
<tr>
<td>Regional Transit with Local Transit Access Mode</td>
<td>0</td>
<td>*</td>
</tr>
<tr>
<td>Regional Transit with Automobile Access Mode</td>
<td>5.5215</td>
<td>11.815</td>
</tr>
<tr>
<td>Non-Motorized Transit</td>
<td>1.3309</td>
<td>2.785</td>
</tr>
</tbody>
</table>

**Generic Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Travel Time</td>
<td>-0.023</td>
<td>-192.999</td>
</tr>
<tr>
<td>Cost perception for Professional Employment Class</td>
<td>-0.1051</td>
<td>-175.119</td>
</tr>
<tr>
<td>Cost Perception for General Office Employment Class</td>
<td>-0.1012</td>
<td>-133.226</td>
</tr>
<tr>
<td>Cost Perception for Service Industry Employment Class</td>
<td>-0.1295</td>
<td>-165.28</td>
</tr>
<tr>
<td>Cost Perception for Manufacturing Employment Class</td>
<td>-0.3953</td>
<td>-230.182</td>
</tr>
</tbody>
</table>

**Modal Specific Parameters for all Public Transit Travel**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walk Time</td>
<td>-0.0269</td>
<td>-182.814</td>
</tr>
<tr>
<td>Wait Time</td>
<td>-0.1159</td>
<td>-258.375</td>
</tr>
</tbody>
</table>

**Modal Specific Parameters for Non Motorized Travel**

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.1846</td>
<td>50.504</td>
</tr>
<tr>
<td>Younger than 26 years</td>
<td>1.0719</td>
<td>138.505</td>
</tr>
<tr>
<td>26 to 30 years</td>
<td>0.6401</td>
<td>101.431</td>
</tr>
<tr>
<td>55 years or older</td>
<td>-0.2241</td>
<td>-42.484</td>
</tr>
</tbody>
</table>

**Modal Specific Parameters for Regional Transit with Automobile Access Travel**

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-0.0877</td>
<td>-1.725</td>
</tr>
<tr>
<td>Younger than 26 years</td>
<td>-0.0818</td>
<td>-0.605</td>
</tr>
<tr>
<td>26 to 30 years</td>
<td>-0.8504</td>
<td>-6.698</td>
</tr>
<tr>
<td>55 years or older</td>
<td>-0.3545</td>
<td>-3.841</td>
</tr>
</tbody>
</table>

**Modal Specific Parameters for Regional Transit with Local Transit Access Travel**

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>5.487</td>
<td>10.559</td>
</tr>
<tr>
<td>Younger than 26 years</td>
<td>0.8207</td>
<td>0.42</td>
</tr>
<tr>
<td>26 to 30 years</td>
<td>0.0882</td>
<td>0.044</td>
</tr>
<tr>
<td>55 years or older</td>
<td>-0.0404</td>
<td>-0.028</td>
</tr>
</tbody>
</table>

**Modal Specific Parameters for Local Transit with Auto Access Travel**

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.6483</td>
<td>98.341</td>
</tr>
<tr>
<td>Younger than 26 years</td>
<td>1.3833</td>
<td>116.984</td>
</tr>
<tr>
<td>26 to 30 years</td>
<td>0.5188</td>
<td>52.671</td>
</tr>
<tr>
<td>55 years or older</td>
<td>-0.0333</td>
<td>-3.606</td>
</tr>
<tr>
<td>Modal Specific Parameters for Local Transit with Walk Access Travel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>Male</td>
<td>0.5616</td>
<td>217.949</td>
</tr>
<tr>
<td>Younger than 26 years</td>
<td>1.3833</td>
<td>116.984</td>
</tr>
<tr>
<td>26 to 30 years</td>
<td>0.5188</td>
<td>52.671</td>
</tr>
<tr>
<td>55 years or older</td>
<td>-0.0333</td>
<td>-3.606</td>
</tr>
</tbody>
</table>

| Modal Specific Parameters for Automobile Passenger Travel |
|---------------------------------|-----|-----------|
| Male                            | 1.2391 | 274.563 |
| Younger than 26 years           | 1.8114 | 272.387 |
| 26 to 30 years                  | 0.3635 | 87.401  |
| 55 years or older               | -0.1253 | -31.882 |
APENDIX B
Configuration File Containing MATSIM Parameter Specification

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE config SYSTEM "http://www.matsim.org/files/dtd/config_v1.dtd">
<config>

<module name="TimeAllocationMutator" >
    <!-- Default:1800.0; Defines how many seconds a time mutation can maximally shift a time. -->
    <param name="mutationRange" value="1800.0" />
</module>

<module name="controller" >
    <!-- Default=false; -->
    <param name="enableLinkToLinkRouting" value="false" />
    <!-- Default=xml; Specifies the file format for writing events. Currently supported: txt, xml. Multiple values can be specified separated by commas (','). -->
    <param name="eventsFileFormat" value="xml" />
    <!-- Default=0; -->
    <param name="firstIteration" value="0" />
    <!-- Default=1000; -->
    <param name="lastIteration" value="30" />
    <!-- Defines which mobility simulation will be used. Currently supported: queueSimulation qsim JDEQSim

    Depending on the chosen mobsim, you'll have to add additional config modules to configure the corresponding mobsim.

    For 'qsim', add a module 'qsim' to the config."
For 'queueSimulation', add a module 'simulation' to the config. -->
<param name="mobsim" value="null" />

<param name="outputDirectory" value="C:\Users\Adam Weiss\Documents\Run Matsim\output23" />

<!-- The type of routing (least cost path) algorithm used, may have the values: Dijkstra, FastDijkstra, AStarLandmarks or FastAStarLandmarks -->
<param name="routingAlgorithmType" value="Dijkstra" />

<!-- An identifier for the current run which is used as prefix for output files and mentioned in output xml files etc. -->
<param name="runId" value="null" />

<!-- Comma-separated list of visualizer output file formats. `transims', `googleearth', and `otfvis'. -->
<param name="snapshotFormat" value="" />

<!-- iterationNumber % writeEventsInterval == 0 defines in which iterations events are written to a file. `0' disables events writing completely. -->
<param name="writeEventsInterval" value="30" />

<!-- iterationNumber % writePlansInterval == 0 defines (hopefully) in which iterations plans are written to a file. `0' disables plans writing completely. Some plans in early iterations are always written -->
<param name="writePlansInterval" value="30" />

<!-- iterationNumber % writeSnapshotsInterval == 0 defines in which iterations snapshots are written to a file. `0' disables snapshots writing completely -->
<param name="writeSnapshotsInterval" value="30" />

</module>

</module>

<!-- Specifies over how many iterations the link volumes should be averaged that are used for the counts comparison. Use 1 or 0 to only use the link volumes of a single iteration. This values cannot be larger than the value specified for writeCountsInterval -->
<param name="averageCountsOverIterations" value="5" />

<!-- factor by which to re-scale the simulated values. necessary when simulation runs with something different from 100%. needs to be adapted manually -->
<param name="countsScaleFactor" value="1.0" />

<!-- distance to distanceFilterCenterNode to include counting stations. The unit of distance is the Euclidean distance implied by the coordinate system -->
<param name="distanceFilter" value="null" />

<!-- node id for center node of distance filter -->
<param name="distanceFilterCenterNode" value="null" />

<!-- input file name to counts package -->
<param name="inputCountsFile" value="null" />

<!-- possible values: `html`, `kml`, `txt`, `all` -->
<param name="outputformat" value="txt" />

<!-- Specifies how often the counts comparison should be calculated and written. -->
<param name="writeCountsInterval" value="10" />

</module>

<!--
================================================================================================

<module name="facilities">
  <param name="inputFacilitiesFile" value="null" />
  <param name="inputFacilityAttributesFile" value="null" />
</module>

<!--
================================================================================================

<module name="global">
  <param name="coordinateSystem" value="NAD83_UTM17N" />
  <param name="numberOfThreads" value="8" />
  <param name="randomSeed" value="4711" />
</module>

<!--
================================================================================================

<module name="households">
  <param name="inputFile" value="null" />
</module>

<!--
================================================================================================


<module name="linkStats">

<!-- Specifies over how many iterations the link volumes should be averaged that are used for the link statistics. Use 1 or 0 to only use the link volumes of a single iteration. This values cannot be larger than the value specified for writeLinkStatsInterval -->
<param name="averageLinkStatsOverIterations" value="5" />

<!-- Specifies how often the link stats should be calculated and written. Use 0 to disable the generation of link stats. -->
<param name="writeLinkStatsInterval" value="10" />
</module>

<module name="locationchoice">

<param name="algorithm" value="random" />
<param name="analysisBinSize" value="20000.0" />
<param name="analysisBoundary" value="200000.0" />
<param name="centerNode" value="null" />
<param name="destinationSamplePercent" value="100.0" />
<param name="epsilonDistribution" value="gumbel" />
<param name="epsilonScaleFactors" value="null" />
<param name="fAttributesFileName" value="null" />
<param name="fkValuesFile" value="null" />
<param name="flexible_types" value="null" />
<param name="idExclusion" value="2147483647" />
<param name="maxDCScoreFile" value="null" />
<param name="maxDistanceDCScore" value="-1.0" />
<param name="maxRecursions" value="1" />
<param name="pBetasFileName" value="null" />
<param name="pkValuesFile" value="null" />
<param name="planSelector" value="SelectExpBeta" />
<param name="prefsFile" value="null" />
<param name="probChoiceSetSize" value="10" />
<param name="radius" value="null" />
<param name="recursionTravelSpeedChange" value="0.1" />
<param name="restraintFcnExp" value="0.0" />
</module>
<param name="restraintFcnFactor" value="0.0" />
<param name="scaleFactor" value="1" />
<param name="travelSpeed_car" value="8.5" />
<param name="travelSpeed_pt" value="5.0" />
<param name="tt_approximationLevel" value="0" />

</module>

<!--
===============================================================
=======
-->

<module name="network">
  <param name="inputChangeEventsFile" value="null" />
  <param name="inputNetworkFile" value="C:\Users\Adam Weiss\Documents\Run MatSim\Network Data\2012NetworkWithTransitFixedAndTypesCapSpeedFixedLinkLengthsFixed.xml" />
  <param name="laneDefinitionsFile" value="null" />
  <param name="timeVariantNetwork" value="false" />
</module>

<!--
===============================================================
=======
-->

<module name="otfvis">
  <param name="agentSize" value="120.0" />
  <param name="coloringScheme" value="standard" />
  <param name="drawNonMovingItems" value="false" />
  <param name="drawTransitFacilities" value="true" />
  <param name="drawTransitFacilityIds" value="true" />
  <param name="leftMouseFunc" value="Zoom" />

  <param name="linkWidth" value="30.0" />

  <!-- The (initial) size of the agents. Only a range of numbers is allowed, otherwise otfvis aborts rather ungracefully, or displays no agents at all. -->
  <param name="agentSize" value="120.0" />

  <!-- coloring scheme for otfvis. Currently (2012) allowed values: standard bvg bvg2 byId gtfs taxicab -->
  <param name="coloringScheme" value="standard" />

  <!-- If non-moving items (e.g. agents at activities, at bus stops, etc.) should be showed. May affect all non-moving items. -->
  <param name="drawNonMovingItems" value="false" />

  <param name="drawTransitFacilities" value="true" />
  <param name="drawTransitFacilityIds" value="true" />
  <param name="leftMouseFunc" value="Zoom" />

  <!-- The (initial) width of the links of the network. Use positive floating point values. -->
  <param name="linkWidth" value="30.0" />

</module>
<!-- Link width is proportional to `numberOfLanes' or to `capacity'. -->
<param name="linkwidthIsProportionalTo" value="numberOfLanes" />

<!-- URL to get WMS tiles from. For a local GeoServer instance, use http://localhost:8080/geoserver/wms?service=WMS& -->
<param name="mapBaseUrl" value="" />

<!-- The WMS layer to display. For GeoServer and a layer called clipped in workspace mz, use mz:clipped -->
<param name="mapLayer" value="" />

<!-- Render everything on top of map tiles. Default: From tiles.openstreetmap.org -->
<param name="mapOverlayMode" value="false" />
<param name="middleMouseFunc" value="Pan" />
<param name="rightMouseFunc" value="Select" />
<param name="showTeleportation" value="false" />
</module>

<--
==================================================================
=======
-->

<module name="parallelEventHandling">

<!-- estimated number of events during mobsim run, useful for configuration -->
<param name="estimatedNumberOfEvents" value="null" />

<!-- number of threads for parallel events handler. 0 or null means parallel events handler is disabled -->
<param name="numberOfThreads" value="null" />
</module>

<--
==================================================================
=======
-->

<module name="planCalcScore">

<!-- new_score = (1-learningRate)*old_score + learningRate * score_from_mobsim. learning rates close to zero emulate score averaging, but slow down initial convergence -->
<param name="learningRate" value="1.0" />

<param name="BrainExpBeta" value="2.0" />
<param name="PathSizeLogitBeta" value="1.0" />
</!
-- [utils/hr] utility for arriving late (i.e. after the latest start time). Normally negative -->
<param name="lateArrival" value="-18.0" />
</!
-- [utils/hr] utility for departing early (i.e. before the earliest end time). Normally negative. Probably implemented correctly, but not tested. -->
<param name="earlyDeparture" value="-0.0" />
</!
-- [utils/hr] marginal utility of doing an activity. normally positive. also the opportunity cost of time if agent is doing nothing. -->
<param name="performing" value="6.0" />
</!
-- [utils/hr] additional marginal utility of traveling by car. normally negative. this comes on top of the opportunity cost of time -->
<param name="traveling" value="-6.0" />
</!
-- [utils/hr] additional marginal utility offset of traveling by pt. normally negative. this comes on top of the opportunity cost of time -->
<param name="travelingPt" value="-6.0" />
</!
-- [utils/hr] additional marginal utility offset of traveling by foot. normally negative. this comes on top of the opportunity cost of time. also see marginalUtlOfDistanceWalk -->
<param name="travelingWalk" value="-6.0" />
<param name="travelingOther" value="-6.0" />
<param name="travelingBike" value="-6.0" />
</!
-- [utils/hr] additional marginal utility for waiting. normally negative. this comes on top of the opportunity cost of time. Probably implemented correctly, but not tested. -->
<param name="waiting" value="-0.0" />
</!
-- [utils/hr] additional marginal utility for waiting for a pt vehicle. normally negative. this comes on top of the opportunity cost of time. Default: if not set explicitly, it is equal to traveling_pt!!! -->
<param name="waitingPt" value="-6.0" />
</!
-- [utils/m] utility of walking per m, normally negative. this is on top of the time (dis)utility. -->
<param name="marginalUtlOfDistanceWalk" value="0.0" />

<param name="marginalUtlOfDistanceOther" value="0.0" />

<!-- [utils/unit_of_money] conversion of money (e.g. toll, distance cost) into utils. Normall positive (i.e. toll/cost/fare are processed as negative amounts of money). -->
<param name="marginalUtilityOfMoney" value="1.0" />

<param name="monetaryDistanceCostRateCar" value="0.0" />

<param name="monetaryDistanceCostRatePt" value="0.0" />

<param name="utilityOfLineSwitch" value="-1.0" />

<param name="constantCar" value="0.0" />

<param name="constantWalk" value="0.0" />

<param name="constantOther" value="0.0" />

<param name="constantBike" value="0.0" />

<param name="constantPt" value="0.0" />

<param name="writeExperiencedPlans" value="false" />

<param name="activityType_0" value="H" />

<!-- write a plans file in each iteration directory which contains what each agent actually did, and the score it received. -->
<param name="writeExperiencedPlans" value="false" />

<param name="activityType_0" value="H" />
<param name="activityPriority_0" value="1.0" />
<param name="activityTypicalDuration_0" value="12:00:00" />
<param name="activityMinimalDuration_0" value="undefined" />
<param name="activityOpeningTime_0" value="undefined" />
<param name="activityLatestStartTime_0" value="undefined" />
<param name="activityEarliestEndTime_0" value="undefined" />
<param name="activityClosingTime_0" value="undefined" />

<!-- **************************** -->
<param name="activityType_1" value="W" />

<param name="activityPriority_1" value="1.0" />
<param name="activityTypicalDuration_1" value="08:00:00" />
<param name="activityMinimalDuration_1" value="undefined" />
<param name="activityOpeningTime_1" value="06:00:00" />
<param name="activityLatestStartTime_1" value="undefined" />
<param name="activityEarliestEndTime_1" value="undefined" />
<param name="activityClosingTime_1" value="18:00:00" />

<!-- **************************** -->
<param name="activityType_2" value="D" />

<param name="activityPriority_2" value="2.0" />
<param name="activityTypicalDuration_2" value="01:00:00" />
<param name="activityMinimalDuration_2" value="undefined" />
<param name="activityOpeningTime_2" value="undefined" />
<param name="activityLatestStartTime_2" value="undefined" />
<param name="activityEarliestEndTime_2" value="undefined" />
<param name="activityClosingTime_2" value="undefined" />

<!-- **************************** -->
<param name="activityType_3" value="F" />

<param name="activityPriority_3" value="3.0" />
<param name="activityTypicalDuration_3" value="08:00:00" />
<param name="activityMinimalDuration_3" value="undefined" />
<param name="activityOpeningTime_3" value="undefined" />
<param name="activityLatestStartTime_3" value="undefined" />
<param name="activityEarliestEndTime_3" value="undefined" />
<param name="activityClosingTime_3" value="undefined" />

<!-- **************************** -->
<param name="activityType_4" value="M" />

<param name="activityPriority_4" value="1.0" />
<param name="activityTypicalDuration_4" value="08:00:00" />
<param name="activityMinimalDuration_4" value="undefined" />  
<param name="activityOpeningTime_4" value="undefined" />  
<param name="activityLatestStartTime_4" value="undefined" />  
<param name="activityEarliestEndTime_4" value="undefined" />  
<param name="activityClosingTime_4" value="undefined" />

<!-- **************************** -->
<param name="activityType_5" value="O" />

<param name="activityPriority_5" value="4.0" />  
<param name="activityTypicalDuration_5" value="12:00:00" />  
<param name="activityMinimalDuration_5" value="undefined" />  
<param name="activityOpeningTime_5" value="undefined" />  
<param name="activityLatestStartTime_5" value="undefined" />  
<param name="activityEarliestEndTime_5" value="undefined" />  
<param name="activityClosingTime_5" value="undefined" />

<!-- **************************** -->
<param name="activityType_6" value="S" />

<param name="activityPriority_6" value="1.0" />  
<param name="activityTypicalDuration_6" value="08:00:00" />  
<param name="activityMinimalDuration_6" value="undefined" />  
<param name="activityOpeningTime_6" value="09:00:00" />  
<param name="activityLatestStartTime_6" value="undefined" />  
<param name="activityEarliestEndTime_6" value="undefined" />  
<param name="activityClosingTime_6" value="15:00:00" />

<!-- **************************** -->
<param name="activityType_7" value="9" />

<param name="activityPriority_7" value="4.0" />  
<param name="activityTypicalDuration_7" value="08:00:00" />  
<param name="activityMinimalDuration_7" value="undefined" />  
<param name="activityOpeningTime_7" value="undefined" />  
<param name="activityLatestStartTime_7" value="undefined" />  
<param name="activityEarliestEndTime_7" value="undefined" />  
<param name="activityClosingTime_7" value="undefined" />

<!-- **************************** -->
<param name="activityType_8" value="pt interaction" />

<param name="activityPriority_8" value="1.0" />  
<param name="activityTypicalDuration_8" value="00:02:00" />  
<param name="activityMinimalDuration_8" value="undefined" />  
<param name="activityOpeningTime_8" value="00:00:00" />  
<param name="activityOpeningTime_8" value="00:00:00"/>

115
<param name="activityLatestStartTime_8" value="undefined" />
<param name="activityEarliestEndTime_8" value="undefined" />
<param name="activityClosingTime_8" value="00:00:00" />
</module>

<!--
====================================================================================================
=======
-->

<module name="plans">

<!-- Path to a file containing person attributes (required file format: ObjectAttributes). -->
<param name="inputPersonAttributesFile" value="null" />

<param name="inputPlansFile" value="C:\Users\Adam Weiss\Documents\Run Matsim\noAutoAccessPlans23pecentnotIncripted.xml" />

<!-- Defines how routes are stored in memory. Currently supported: LinkNetworkRoute, CompressedNetworkRoute. -->
<param name="networkRouteType" value="LinkNetworkRoute" />
</module>

<!--
====================================================================================================
=======
-->

<module name="planscalcroute">

<!-- factor with which beeline distances (and therefore times) are multiplied in order to obtain an estimate of the network distances/times. Default is something like 1.3 -->
<param name="beelineDistanceFactor" value="1.3" />

<!-- All the modes for which the router is supposed to generate network routes (like car) -->
<param name="networkModes" value="car,ride" />

<!-- Free-speed factor for a teleported mode based on freespeed: freespeedFactor \* <freespeed car travel time>. Insert a line like this for every such mode. freespeedFactor wins over teleportedModeSpeed, if both are set (says michaz). -->
<param name="teleportedModeFreespeedFactor_pt" value="2.0" />

<!-- Speed for a teleported mode based on beeline-distance: (<beeline distance> \* beelineDistanceFactor) / speed. Insert a line like this for every such mode. -->

<param name="teleportedModeSpeed_bike" value="4.166666666666667" />

<!-- Speed for a teleported mode based on beeline-distance: (<beeline distance> * beelineDistanceFactor) / speed. Insert a line like this for every such mode. -->
<param name="teleportedModeSpeed_undefined" value="13.88888888888889" />

<!-- Speed for a teleported mode based on beeline-distance: (<beeline distance> * beelineDistanceFactor) / speed. Insert a line like this for every such mode. -->
<param name="teleportedModeSpeed_walk" value="0.8333333333333333" />
</module>

<!--
==============================================================
=======

<module name="ptCounts" >

<!-- factor by which to re-scale the simulated values. necessary when simulation runs with something different from 100% needs to be adapted manually -->
<param name="countsScaleFactor" value="1.0" />

<!-- distance to distanceFilterCenterNode to include counting stations. The unit of distance is the Euclidean distance implied by the coordinate system -->
<param name="distanceFilter" value="null" />

<!-- node id for center node of distance filter -->
<param name="distanceFilterCenterNode" value="null" />

<!-- input file containing the alighting (getting off) counts for pt -->
<param name="inputAlightCountsFile" value="null" />

<!-- input file containing the boarding (getting on) counts for pt -->
<param name="inputBoardCountsFile" value="null" />

<!-- input file containing the occupancy counts for pt -->
<param name="inputOccupancyCountsFile" value="null" />

<!-- possible values: `html', `kml', `txt', `all' -->
<param name="outputFormat" value="null" />

<!-- every how many iterations (starting with 0) counts comparisons are generated -->
<param name="ptCountsInterval" value="10" />
<param name="endTime" value="30:00:00" />
<param name="flowCapacityFactor" value="0.25" />

<!-- decides if waiting vehicles enter the network after or before the already driving vehicles were moved. Default: false -->
<param name="insertingWaitingVehiclesBeforeDrivingVehicles" value="false" />

<!-- Defines which mode should be the qsim `main' (=congested) mode. Technically, this is the mode that the departure handler of the netsimengine handles. Effective cell size, effective lane width, flow capacity factor, and storage capacity factor need to be set with diligence. Needs to be a vehicular mode to make sense. -->
<param name="mainMode" value="car" />

<!-- Shortens a link in the visualization, i.e. its start and end point are moved into towards the center. Does not affect traffic flow. -->
<param name="nodeOffset" value="0.0" />

<!-- Use number of threads > 1 for parallel version using the specified number of threads -->
<param name="numberOfThreads" value="1" />

<!-- Boolean. `true': stuck vehicles are removed, aborting the plan; `false': stuck vehicles are forced into the next link. `false' is probably the better choice. -->
<param name="removeStuckVehicles" value="false" />

<!-- `maxOfStarttimeAndEarliestActivityEnd' (default behavior) or `onlyUseStarttime' -->
<param name="simStarttimeInterpretation" value="maxOfStarttimeAndEarliestActivityEnd" />

<!-- snapshotStyle: `equiDist' (vehicles equidistant on link) or `queue' (vehicles queued at end of link) or `withHolesExperimental' (experimental!!) -->
<param name="snapshotStyle" value="equiDist" />

<param name="snapshotperiod" value="00:00:00" />
<param name="startTime" value="03:30:00" />
<param name="storageCapacityFactor" value="1.5" />

</module>
time in seconds. Time after which the frontmost vehicle on a link is called `stuck' if it does not move. -->
<param name="stuckTime" value="100.0" />

<param name="timeStepSize" value="00:00:01" />

`queue' for the standard queue model, `withHolesExperimental' (experimental!!) for the queue model with holes -->
<param name="trafficDynamics" value="queue" />

Defines what happens if an agent wants to depart, but the specified vehicle is not available. One of: teleport, wait, exception -->
<param name="vehicleBehavior" value="teleport" />
</module>

============================================================================

<module name="roadpricing" >
</module>

============================================================================

<module name="scenario" >

Set this parameter to true if households should be used, false if not. -->
<param name="useHouseholds" value="false" />

Set this parameter to true if knowledge should be used, false if not. -->
<param name="useKnowledge" value="true" />

Set this parameter to true if lanes should be used, false if not. -->
<param name="useLanes" value="false" />

Set this parameter to true if roadpricing should be used, false if not. -->
<param name="useRoadpricing" value="false" />

Set this parameter to true if signal systems should be used, false if not. -->
<param name="useSignalsystems" value="false" />
<param name="useTransit" value="true" />

<param name="useVehicles" value="true" />

<module name="signalsystems">
    <param name="ambertimes" value="null" />
    <param name="intergreentimes" value="null" />
    <param name="signalcontrol" value="null" />
    <param name="signalgroups" value="null" />
    <param name="signalsystems" value="null" />
    <param name="useAmbertimes" value="false" />
    <param name="useIntergreentimes" value="false" />
</module>

<module name="strategy">
    <!-- maximum number of plans per agent. `0` means ``infinity``. Currently (2010), `5` is a good number -->
    <param name="maxAgentPlanMemorySize" value="4" />

    <!-- name of strategy (if not full class name, resolved in StrategyManagerConfigLoader) -->
    <param name="Module_1" value="SelectExpBeta" />

    <!-- probability that a strategy is applied to a given a person. despite its name, this really is a ``weight`` -->
    <param name="ModuleProbability_1" value="0.8" />

    <!-- iteration after which module will be disabled. most useful for ``innovative`` strategies (new routes, new times, ...) -->
    <param name="ModuleDisableAfterIteration_1" value="null" />

    <!-- path to external executable (if applicable) -->

</module>
<param name="ModuleExePath_1" value="null" />

<!-- ******************************************* -->
<param name="Module_2" value="ReRoute" />

<param name="ModuleProbability_2" value="0.2" />
<param name="ModuleDisableAfterIteration_2" value="null" />
<param name="ModuleExePath_2" value="null" />
<param name="ExternalExeConfigTemplate" value="null" />
<param name="ExternalExeTmpFileRootDir" value="null" />
<param name="ExternalExeTimeOut" value="3600" />

<!-- name of PlanSelector for plans removal. If not full class name, resolved in StrategyManagerConfigLoader. default is `null`, which eventually calls SelectWorstPlan. This is not a good choice from a discrete choice theoretical perspective. Alternatives, however, have not been systematically tested. kai, feb'12 -->
<param name="planSelectorForRemoval" value="null" />

<!-- fraction of iterations where innovative strategies are switched off. Something link 0.8 should be good. E.g. if you run from iteration 400 to iteration 500, innovation is switched off at iteration 480 -->
<param name="fractionOfIterationsToDisableInnovation" value="Infinity" />

</module>

<!--
==================================================================================================

<module name="subtourModeChoice" >

<!-- Defines the chain-based modes, seperated by commas -->
<param name="chainBasedModes" value="car,bike" />

<!-- Defines whether car availability must be considered or not. A agent has no car only if it has no license, or never access to a car -->
<param name="considerCarAvailability" value="false" />

<!-- Defines all the modes available, including chain-based modes, seperated by commas -->
<param name="modes" value="car,pt,bike,walk" />

</module>

<!--
==================================================================================================

121
<module name="transit">

  <!-- Comma-separated list of transportation modes that are handled as transit. Defaults to 'pt'. -->
  <param name="transitModes" value="pt" />

  <!-- Input file containing the transit schedule to be simulated. -->
  <param name="transitScheduleFile" value="C:\Users\Adam Weiss\Documents\Run Matsim\Network Data\schedule_LowDeparturesMergedorRemoved_July24.xml" />

  <!-- Input file containing the vehicles used by the departures in the transit schedule. -->
  <param name="vehiclesFile" value="C:\Users\Adam Weiss\Documents\Run Matsim\Network Data\routeVehicleOutputjuly18ScaledCapacity.xml" />
</module>

<!--====================================================================================================

<module name="transitRouter">

  <!-- additional time the router allocates when a line switch happens. Can be interpreted as a 'safety' time that agents need to safely transfer from one line to another -->
  <param name="additionalTransferTime" value="0.0" />

  <!-- step size to increase searchRadius if no stops are found -->
  <param name="extensionRadius" value="200.0" />

  <!-- maximum beeline distance between stops that agents could transfer to by walking -->
  <param name="maxBeelineWalkConnectionDistance" value="100.0" />

  <!-- the radius in which stop locations are searched, given a start or target coordinate -->
  <param name="searchRadius" value="1000.0" />
</module>

<!--====================================================================================================

122
<module name="travelTimeCalculator" >

<!-- Transport modes that will be respected by the travel time collector. 'car' is default, which includes also bussed from the pt simulation module. Use this parameter in combination with 'filterModes' = true! -->
<param name="analyzedModes" value="car" />

<param name="calculateLinkToLinkTravelTimes" value="false" />
<param name="calculateLinkTravelTimes" value="true" />

<!-- If true, link travel times from legs performed on modes not included in the 'analyzedModes' parameter are ignored. -->
<param name="filterModes" value="false" />

<!-- How to deal with congested time bins that have no link entry events. 'optimistic' assumes free speed (too optimistic); 'experimental_LastMile' is experimental and probably too pessimistic. -->
<param name="travelTimeAggregator" value="optimistic" />

<!-- The size of the time bin (in sec) into which the link travel times are aggregated for the router -->
<param name="travelTimeBinSize" value="900" />

<param name="travelTimeCalculator" value="TravelTimeCalculatorArray" />

<!-- How to deal with link entry times at different positions during the time bin. Currently supported: average, linearinterpolation -->
<param name="travelTimeGetter" value="average" />

</module>

<!--
========================================
=======
-->

<module name="vspExperimental" >

<!-- String: minOfDurationAndEndTime tryEndTimeThenDuration endTimeOnly. Anything besides minOfDurationAndEndTime will internally use a different (simpler) version of the TimeAllocationMutator. -->
<param name="activityDurationInterpretation" value="minOfDurationAndEndTime" />

<!-- REQUIRED: file with HBEFA 3.1 fleet average cold emission factors -->
<param name="averageFleetColdEmissionFactorsFile" value="null" />

123
<!-- REQUIRED: file with HBEFA 3.1 fleet average warm emission factors -->
<param name="averageFleetWarmEmissionFactorsFile" value="null"/>

<param name="chainBasedModes" value="car"/>

<param name="detailedColdEmissionFactorsFile" value="null"/>

<param name="detailedWarmEmissionFactorsFile" value="null"/>

<param name="emissionRoadTypeMappingFile" value="null"/>

<param name="emissionVehicleFile" value="null"/>

<param name="inputMZ05File" value=""/>

<param name="isAbleToOverwritePtInteractionParams" value="false"/>

<param name="isGeneratingBoardingDeniedEvent" value="false"/>

<param name="isUsingOpportunityCostOfTimeForLocationChoice" value="true"/>

<-- (do not use) Set this filename of MZ05 daily analysis -->
<param name="inputMZ05File" value=""/>

<-- (do not use except of you have to) There was a problem with pt interaction scoring. Some people solved it by overwriting the parameters of the pt interaction activity type. Doing this now throws an Exception. If you still insist on doing this, set the following to true. -->
<param name="isAbleToOverwritePtInteractionParams" value="false"/>

<param name="isGeneratingBoardingDeniedEvent" value="false"/>

<-- if an approximation of the opportunity cost of time is included into the radius calculation for location choice. 'true' will be faster, but it is an approximation. Default is 'true'; 'false' is available for backwards compatibility. -->
<param name="isUsingOpportunityCostOfTimeForLocationChoice" value="true"/>
<param name="logitScaleParamForPlansRemoval" value="1.0" />

<!-- changes MATSim's global time format used in output files. Can be used to enforce writing fractional seconds e.g. in output_plans. default is 'hh:mm:ss' (because of backwards compatibility). see Time.java for possible formats -->
<param name="matsimGlobalTimeformat" value="HH:mm:ss" />

<!-- (do not use) set the traffic mode option for subTourModeChoice by Yu -->
<param name="modes" value="car, pt" />

<!-- (not tested) will remove plan attributes that are presumably not used, such as activityStartTime. default=false -->
<param name="removingUnnecessaryPlanAttributes" value="false" />

<!-- first iteration of MSA score averaging. The matsim theory department suggests to use this together with switching of choice set innovation, but it has not been tested yet. -->
<param name="scoreMSAStartsAtIteration" value="null" />

<!-- if true then detailed emission factor files must be provided! -->
<param name="usingDetailedEmissionCalculation" value="false" />

<!-- indicates if, for routing, the opportunity cost of time should be added to the mode-specific marginal utilities of time. Default is true; false is possible only for backwards compatibility. This is only a suggestion since there is (by matsim design) no way to enforce that mental modules obey this. -->
<param name="usingOpportunityCostOfTimeForPtRouting" value="true" />

<!-- Options: 'ignore', 'warn', 'abort'. Default: either 'ignore' or 'warn'. When violating VSP defaults, this results in nothing, warnings, or aborts. Members of VSP should use 'abort' or talk to kai. -->
<param name="vspDefaultsCheckingLevel" value="ignore" />

<!-- if true then writes output_events in output directory. default is 'false'. Will only work when lastIteration is multiple of events writing interval -->
<param name="writingOutputEvents" value="false" />

</module>

<!--
===============================================================
======= -->

</config>