An Integrated Modelling and Analytics Platform for Service Planning of Bus Stops Using Archived AVL, APC, and Schedule Data

by

Alexander Richard Knowles

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 Alexander R. Knowles, 2015
An Integrated Modelling and Analytics Platform for Service Planning of Bus Stops Using Archived AVL, APC and Schedule Data

Alexander Richard Knowles

Master of Applied Science, 2015

Department of Civil Engineering
University of Toronto

2015

Abstract

Decision-making by transit agencies in response to dynamic system and external conditions has significant impact on transit system performance. Analysis of Automatic Data Collection (ADC) system data has meaningful potential to improve transit agencies’ decision-making related to system planning and operations.

The objective of this research was twofold. First, a set of analytic models was defined to estimate changes in bus performance and passenger travel times, given user-specified stop layout changes to an existing transit trip pattern. The models drew upon archived Automatic Vehicle Location (AVL), Automatic Passenger Counter (APC), and schedule data, and were constructed to accommodate bus stop removal, stop addition, and introduction of limited-stop service. Second, a data analytics platform was created to present ten Key Performance Indicators (KPIs) which drew upon the analytic models. Software was created to integrate analytic modelling with a frontend graphical user interface for defining stop layout changes, and returning resulting KPIs.
Acknowledgments

Many thanks (in no particular order) to the many individuals from IBI Group, the LTC, the TTC, Brampton Transit, YRT, ITS Canada, ITE, the CAA, as well as independent consultants and any organizations I’ve forgotten from whom I’ve received help over the past two years.

I’m grateful to everyone at U of T who has enriched my time here; in particular, MDA & GLP, for the help, the distractions, and for putting up with me, and MJH & ECT, without whom I may well have never gotten myself into this mess.

To Paul Luong, who provided programming labour and wisdom from the outset of this research, thanks for putting your true calling as an eBay tycoon on ice to see this through to the end.

To my supervisor, Amer Shalaby, without whom much of my success over the past two years would not have materialized, thanks for providing the opportunity for me to do this.

To my family, thanks for making it possible for me to do this.

To my partner, Alanna, thank you for giving me a reason to do this, as well as a reason for everything else.
I have observed that the world has suffered far less from ignorance than from pretensions to knowledge.

- Daniel Boorstin

If you’ve got a truffle dog, you can go truffling, and you might get lucky now and then.

- Mark Knopfler, “Get Lucky”
# Table of Contents

Acknowledgments .......................................................................................................................... iii

Table of Contents .................................................................................................................................. v

List of Tables ........................................................................................................................................ v

List of Figures ......................................................................................................................................... x

List of Equations .................................................................................................................................... xii

1 Introduction ........................................................................................................................................ 1

1.1 Background ..................................................................................................................................... 1

1.2 Motivation ....................................................................................................................................... 2

1.3 Objective ......................................................................................................................................... 4

1.4 Case Study: London Transit Commission (LTC) Network ............................................................... 5

1.5 Precursory Data Analytics Project ................................................................................................. 8

1.6 Thesis Organization ....................................................................................................................... 8

2 Literature Review .............................................................................................................................. 10

2.1 Data Analytics for Transit Operations and Planning ...................................................................... 10

2.2 Use of Archived AVL and APC Data for Performance Improvement .............................................. 12

2.3 Stop Removal and Addition, and Limited-Stop Operation ............................................................. 14

2.4 Data Analytics for Route and Stop Changes .................................................................................. 17

2.5 Summary of Reviewed Literature and Gaps in Research .............................................................. 19

3 Data Sources, Cleaning, and Limitations ......................................................................................... 22

3.1 Data Sources .................................................................................................................................. 22

3.2 Cleaning of Inaccurate, Missing, and Irrelevant Data .................................................................... 22

3.3 Data Limitations ............................................................................................................................. 24

4 Methodology and Models .................................................................................................................. 26

4.1 Basic Methodology ......................................................................................................................... 26

4.2 Key Performance Indicators (KPIs) ............................................................................................... 29
4.3 Assumptions.................................................................................................................. 35
4.4 Average Boarding and Alighting Approximations ...................................................... 37
4.5 Dwell Time Estimation .................................................................................................. 41
4.6 Passenger Redistribution Estimation .......................................................................... 44
   4.6.1 Stop Addition ........................................................................................................ 44
   4.6.2 Stop Removal & Limited-Stop Operation ................................................................ 46
4.7 Chance of Dwell Estimation .......................................................................................... 47
   4.7.1 Stop Addition ........................................................................................................ 48
   4.7.2 Stop Removal & Limited-stop Operation ................................................................ 49
4.8 Acceleration and Deceleration Time Penalty Estimation ............................................. 53
4.9 Access and Egress Penalty Estimation ........................................................................ 56
4.10 Cycle Time, Headway, Average Speed, Productive Capacity, Maximum Load Section
     Load Factor, and Passenger Wait Time Estimation ....................................................... 60
   4.10.1 Cycle Time Increase/Decrease Estimation ......................................................... 61
   4.10.2 Headway Increase/Decrease Estimation .............................................................. 62
   4.10.3 Average Speed Increase/Decrease Estimation ...................................................... 62
   4.10.4 Productive Capacity Increase/Decrease Estimation ............................................ 63
   4.10.5 Maximum Load Section Load Factor Estimation ................................................ 64
   4.10.6 Wait Time Increase/Decrease Estimation ............................................................ 65
4.11 Limited-Stop Operation Preference Estimation .......................................................... 65
   4.11.1 Passenger-Count Origin-Destination (O-D) Matrix Estimation ......................... 67
   4.11.2 In-Vehicle Time Savings Matrix Estimation ....................................................... 69
   4.11.3 Access/Egress Penalty Matrix Estimation ............................................................ 70
   4.11.4 Passenger Redistribution Matrix Estimation ...................................................... 72
   4.11.5 Limited-Stop Preference Matrix Estimation ......................................................... 75
4.12 Stop Removal Benefit-Cost Analysis .......................................................................... 77
4.13 Summary of Models ................................................................................................. 78
Software Platform .................................................................................................................. 79
5.1 Data Filters and Parameters ............................................................................................. 79
  5.1.1 Date Range, Day(s) of the Week, and Time of Day ................................................. 80
  5.1.2 Route and Trip Pattern ............................................................................................. 81
  5.1.3 Maximum Stop Spacing and Minimum Stop Usage (Limited-Stop) ....................... 81
  5.1.4 Value of Access/Egress Time ................................................................................... 82
5.2 Desktop Software Platform ............................................................................................ 83
Demonstration and Validation of Sample Results ................................................................. 89
6.1 Practical Application Demonstration ............................................................................. 90
  6.1.1 Stop Addition and Removal (Stop Consolidation) .................................................. 92
  6.1.2 Limited-stop Operation Preference Sample Results .............................................. 97
6.2 Validation ....................................................................................................................... 102
  6.2.1 Stop Removal (SR) Sample Results ..................................................................... 103
  6.2.2 Stop Addition (SA) Sample Results ..................................................................... 103
  6.2.3 Validation of Sample Results ................................................................................. 105
Conclusions, Limitations, and Future Work ........................................................................ 110
7.1 Summary ......................................................................................................................... 110
7.2 Conclusions .................................................................................................................... 113
7.3 Limitations ..................................................................................................................... 116
  7.3.1 Methodological Limitations ............................................................................... 116
  7.3.2 Data-Imposed Methodological Limitations ......................................................... 117
  7.3.3 Data-Imposed Limitations ................................................................................... 118
7.4 Future Work .................................................................................................................... 118
References ............................................................................................................................ 120
List of Tables

Table 1: Hypothetical Initial Average Boarding and Alighting Counts Along a Trip Pattern ..... 39
Table 2: Hypothetical $Bj$ Values Along a Trip Pattern, Stage 0 ............................................... 40
Table 3: Adjusted $Bj$, $Bj^*$ Values at Stages 0-3 ....................................................................... 40
Table 4: Hypothetical Adjusted Average Boarding and Alighting Counts Along a Trip Pattern 41
Table 5: Dwell Time Estimator Parameters .................................................................................. 43
Table 6: Example One-Directional Distances Between Stops (metres) ........................................... 68
Table 7: Example Average Boarding and Alighting Counts (passengers) ...................................... 68
Table 8: Example Estimated Passenger-Count O-D Matrix ............................................................. 69
Table 9: Example Time-Saved Matrix ............................................................................................. 70
Table 10: Example Access/Egress Penalty Matrix for Downstream/Downstream Passengers (seconds) .................................................................................................................. 72
Table 11: Example Passenger Redistribution Matrix: Downstream/Downstream Passengers (% of total) ........................................................................................................................................ 74
Table 12: Example Passenger Redistribution Matrix: Downstream/Upstream Passengers (% of total) ........................................................................................................................................ 74
Table 13: Example Passenger Redistribution Matrix: Upstream/Downstream Passengers (% of total) ........................................................................................................................................ 74
Table 14: Example Passenger Redistribution Matrix: Upstream/Upstream Passengers (% of total) ........................................................................................................................................ 75
Table 15: Example Limited-Stop Benefit Matrix .............................................................................. 76
Table 16: Example of Automated Stop Removal Procedure for Limited-stop Operation ........... 82
Table 17: Results of stop removal and addition (stop consolidation) along Pattern 50649 ........ 96

Table 18: Pattern 50649 Limited-Stop Results, 900m spacing (Psngr. Threshold = 4, Walk Value Ratio = 1) .......................................................................................................................... 99

Table 19: Pattern 50649 Limited-Stop Results, 800m spacing (Psngr. Threshold = 4, Walk Value Ratio = 1) .......................................................................................................................... 101

Table 20: Pattern 50649 Limited-Stop Results, 800m spacing (Psngr. Threshold = 4, Walk Value Ratio = 2) .......................................................................................................................... 101

Table 21: Test (Control) Trip Patterns Examined For Sample Results and Validation........... 102

Table 22: Trip Pattern 50649 Stop Removal Sample Results................................................. 103

Table 23: Trip Pattern 65414 Stop Addition Sample Results................................................. 104

Table 24: Passenger Distribution Along Patterns 50649 and 65414 ...................................... 104

Table 25: Test and Validation Trip Pattern Characteristics.................................................... 106

Table 26: Linear regression of passenger load and number of stops vs. actual observed run time ........................................................................................................................................................................... 106
List of Figures

Figure 1: Location of London, Ontario, Canada (Google Maps, 2015) ........................................... 6

Figure 2: Map of the London Transit Commission bus network (London Transit Commission, 2013) ................................................................................................................................. 7

Figure 3: Flowchart showing high-level methodology ........................................................................ 27

Figure 4: Catchment distance for the first stop along a trip pattern ..................................................... 36

Figure 5: Distances between new (Ω) and existing (a, b, c) stops along a trip pattern ......................... 44

Figure 6: Stop b catchment distance, and catchment distance within midpoint (θ) from Stop a to Stop c ............................................................................................................................... 46

Figure 7: Deleted stop b catchment distance, and midpoint (θ) from Stop a to Stop c ...................... 57

Figure 8: The intersection of filter domains was used to determine the subset of data for any analysis ......................................................................................................................................................... 79

Figure 9: Sample screenshot of overall software graphical user interface ........................................ 84

Figure 10: Sample screenshot of filters (with the date and time filters open) ................................. 85

Figure 11: Sample screenshot of Google Maps plot of stops along selected trip pattern (Google Maps, 2015) ........................................................................................................................................ 86

Figure 12: Sample screenshot of stops ordered by removal benefit-cost (BC) ratio ...................... 87

Figure 13: Sample screenshot of list of stops on selected trip pattern ........................................... 87

Figure 14: Sample screenshot of results of stop removal function ................................................. 88

Figure 15: Route 13 in the LTC Network (London Transit Commission, 2015) ............................... 89

Figure 16: Initial stop layout of Route 13, Pattern 50649 ............................................................... 91

Figure 17: BC Ratios for the Monday-Friday evening period along Pattern 50649 .................... 92
Figure 18: Interim results of stop removal process, and addition of new stops between stops 14 and 20, and stops 21 and 25, along pattern 50649................................................................. 94

Figure 19: Final results of Pattern 50649 hypothetical stop removal and addition .................. 95

Figure 20: Pattern 50649 limited-stop analysis, with maximum 900m spacing ...................... 98

Figure 21: Pattern 50649 limited-stop analysis, with maximum 800m spacing ..................... 100
List of Equations

[1] $\Delta T_{bus} = T_{busf} - T_{busi}$ ................................................................. 29

[2] $H_f = T_{ci} - \Delta T_{bus}$................................................................. 30

[3] $\Delta H = H_f - H_i$ ................................................................. 30

[4] $T_{cf} = T_{ci} - tti + ttf - \Delta T_{bus}$ ................................................................. 30

[5] $\Delta T_c = T_{ci} - T_{cf}$ ................................................................. 30

[6] $v_i = D\alpha\Omega T_{busi}$ ................................................................. 31

[7] $v_f = D\alpha\Omega T_{busf}$ ................................................................. 31

[8] $\Delta v = v_f - v_i$ ................................................................. 31

[9] $CP_i = v_i(C_{bus})$ ................................................................. 32

[10] $CP_f = v_f(C_{bus})$ ................................................................. 32

[11] $\Delta CP = CP_f - CP_i$ ................................................................. 32

[12] $\alpha_f = (\alpha_i)(H_fH_i)$ ................................................................. 32

[13] $\Delta \alpha = \alpha_f - \alpha_i$ ................................................................. 32

[14] $TP_j = P_j(\Delta T_{bus})$ ................................................................. 33

[15] $TP = jkTP_j$ ................................................................. 33

[16] $t_{inc} = X_{incw}$ ................................................................. 33

[17] $TW_j = (H_fH_i^2)B_j$ ................................................................. 34

[18] $L = N_{LN} + N_{local}$ ................................................................. 34

[19] $B_j = \text{inbin}$ ................................................................. 38
\[ 20 \] TB = jkBj .................................................................................................. 38

\[ 21 \] DAB = TA-TB .................................................................................................. 39

\[ 22 \] Bj* = (BjTB) (DAB) .................................................................................................. 39

\[ 23 \] bΩb = (bb)(dΩ)dac .................................................................................................. 45

\[ 24 \] bΩc = (bc)(dbΩ)dbd .................................................................................................. 45

\[ 25 \] BΩ = bΩb + bΩc .................................................................................................. 45

\[ 26 \] = 2(daθ- dab2dab + dbc) .................................................................................................. 46

\[ 27 \] Rbc = 1-Rba .................................................................................................. 46

\[ 28 \] Ba* = Ba + (Rba)(Bb) .................................................................................................. 46

\[ 29 \] Bc* = Bc + (Rbc)(Bb) .................................................................................................. 46

\[ 30 \] Cj = Sjnj .................................................................................................. 48

\[ 31 \] CΩ = CbTΩb + CcTΩcTΩb + TΩc .................................................................................................. 48

\[ 32 \] Cbf = 1-1-Cbi .................................................................................................. 49

\[ 33 \] Cbf = 1-(1-Cbi)(1-Cai) .................................................................................................. 49

\[ 34 \] Cbf = 1-(1-Cbi)(1-Cai)(1-Cci) .................................................................................................. 49

\[ 35a \] Abf = Cbiabi + Cciabci .................................................................................................. 52

\[ 36 \] = vmax2 .................................................................................................. 55

\[ 37 \] tlost = (1-uvgumax)(taccel + tdecel) .................................................................................................. 55

\[ 38 \] Dw = Dn2 + Df2 + 4DnDf-2DnDt2Dt .................................................................................................. 56

\[ 39 \] Xi = (li)(x0i) + (lii)(x0ii) + (liii)(x0iii)li + lii + lii .................................................................................................. 58
\[ X_f = (li)(x_fi) + (lii)(x_fii) + (liii)(x_fiii)(li + lii + liii) \ldots \]

\[ T_{bus} = jkDj + top \]

\[ T_c = T_{bus} + tt \]

\[ \alpha = PC = P(freq)(Cb) = HP60Cbus \]

\[ \alpha_f = HfPf60C(hiPi60Cb = HfHi \]

\[ T_{cup} = DbcDbc + DcdTC \]

\[ T_{down} = DcdDbc + DcdTC \]

\[ \text{Benefit to Cost ratio at Stop } j = \text{BenefitjCostj} \]

\[ r = VAETAETV \]

\[ \text{LODr = 0 for } r > 1 \]
1 Introduction

1.1 Background

Within any urban centre, decision-making by transit agencies in response to dynamic system and external conditions has significant impact on overall transit system performance. Such decision-making has always been a massive technical and logistical challenge, and formally or not, transit agencies have always collected system performance data in order to inform decisions related to operations and planning. Historically, this collection, curation, and analysis of data was completed manually: passenger travel histories and perceptions were solicited via surveys, and headway management, estimation of on-time performance, and passenger counts were conducted by on-site supervisors. Data was then analyzed by hand, or using calculators or early spreadsheet software. Understandably, this method for producing actionable intelligence was slow, labour-intensive, and had high marginal cost. Transit agencies, which were and still tend to be chronically under-resourced, were obliged to place reliance upon incomplete and infrequent analysis of sparse datasets. Where assessed at all, factors such as time-of-day, day of the week, and seasonal weather were often not examined for weeks or months following, resulting in unresponsive system changes and belated reactive measures.

Even today, many transit agencies gather and analyze data manually, especially data related to passenger activities and perceptions. However, the advent of Automatic Data Collection (ADC) systems has created an opportunity to generate, at very low marginal cost, large quantities of precise, disaggregate passenger trip data and transit operations data, which can be pooled into data warehouses (Wilson, 2013). Analysis of these warehouses in order to produce Key Performance Indicators (KPIs), for instance, has immense potential to improve transit agencies’ decision-making related to system planning and operations. Such analysis presents transit agencies with an opportunity to take advantage of recent developments in the world of data science (often referred to somewhat ambiguously and insipidly as “big data”).

Making effective use of data warehouses requires strong data mining, management, and analysis tools. Further complicating the situation, few comprehensive ADC systems have been designed
with long-term data storage or analysis in mind, and the data generated is often unreliable, incomplete, or unusable for system planning or improvement purposes. For example, most Automatic Vehicle Location (AVL) systems are designed primarily to feed real-time data to a complementary Computer Aided Dispatch (CAD) system, with little or no effort put towards storing and processing data such that post-operation analysis is feasible. Additionally, as Zhao noted (2004), many transit agencies lack the necessary data processing tools and personnel to compile and make use of ADC-generated data for functions other than those for which the systems were explicitly designed. This is as true now as it was at the time of Zhao’s writing.

Other obstacles exist, or rather persist, particularly in regards to the quality of automatically generated data. While Automatic Passenger Counter (APC) equipment and, consequently, APC data quality, has improved greatly in the last decade, limitations posed by the (lack of) granularity and accuracy of AVL data continue. One of the ongoing ironies of ADC systems is that although APC equipment tends to be very expensive and is now capable of producing highly accurate data, the APC data is far less valuable when it is not accompanied by accurate and complete AVL data with which it can be oriented.

1.2 Motivation

Every time a transit vehicle stops along a route, its average speed decreases, and its average run time and run time variability increase. For a transit agency – saddled with obligations to pay hourly wages and vehicle and station upkeep costs – increased run time denotes increased operational cost. From a passenger’s point of view, increased run time means increased average in-vehicle travel time and less frequent service, all other things being equal. The costs of stopping are clear and readily observable for transit agencies and passengers alike, and have been described in many studies (Schwarcz, 2004; El-Geneidy & Surprenant-Legault, 2010; Vuchic, 2005). Numerous studies, including those by Furth & Rahbee (2000) and Li & Bertini (2008), have found that surface transit lines suffer materially from overly tight spacing.

Conversely, adding stops along a route increases its connectivity to other routes in the network. It reduces the average distance that users must travel, by other modes, in order to access the
transit service, saving access time for those users, and makes the service viable for some individuals who would otherwise ignore it. Obviously, appropriate stop spacings must be found for each route. In recent years, transit agencies across the world have begun to review stop spacing with an eye to meeting budgetary constraints (Mamun & Lownes, 2014). This is especially true in the United States, where surface transit stop spacing tends to be in the order of 160–230m (Furth & Rahbee, 2000), and Canada, which has similar average stop spacing, but tends to be less of a concern in Europe, where stop spacing is closer to 320-400m on average (Furth & Rahbee, 2000). A study by Li & Bertini (2008) determined an optimal stop spacing for local bus operation (as opposed to limited-stop operation) of 1,222 feet (373 m) in the city of Portland.

A means of assessing the effects of Stop Removal (SR) – the process of deleting stops along an existing route – and the reverse process, Stop Addition (SA), is thus a critical tool for any transit agency. (The term Stop Consolidation generally refers to some combination of SR and SA; use of the term was therefore avoided throughout this thesis for the sake of clarity). While microscopic, and to a lesser extent, macroscopic models can estimate with great accuracy and precision the effects of making changes to a route’s stop layout, such efforts tend to be expensive and time-consuming. Many agencies also lack personnel with the extensive training required to capitalize on such complex modelling tools. Yet few site-specific sketch planning tools exist for quickly estimating the effects of stop layout changes, and the result is that SR and SA procedures by transit agencies are often guided by anecdotal evidence and gut feel rather than evidence-based and sound analysis.

Recent developments in the field of data science and the plummeting cost of computer processing power have rendered previously unaffordable levels and types of analysis trivial. Moreover, fuel for this analysis has become increasingly available to transit agencies in the form of large quantities of disaggregate transit operations data coming from ADC systems. Combining this opportunity with the dearth of existing high-level transit planning and operations tools, this thesis’s research attempted to produce an integrated platform for cleaning, analyzing, and presenting warehoused data to transit agencies to facilitate fact-based decision-making. Specifically, automatically generated data was used to produce a number of simple predictive
models yielding KPIs related to both passengers and transit vehicles. These KPIs serve as objective means of comparing existing and potential performance of routes and stops based on SR and SA actions under consideration. Conversion of existing routes to include limited-stop operation in parallel to an existing local service was also assessed. Limited-stop operation is typically defined as operating along an existing local service transit corridor, but with a significantly reduced number of stops.

Changes to total ridership and the effect of seasonal weather were considered exogenous in this study. Additionally, this study was not intended to provide detailed statistical analysis of the benefits of SR or SA or limited-stop operation. Previous studies using archived transit data to estimate the effects of stop removal on bus run times and passenger activity produced one-time analyses of individual routes (El-Geneidy, Strathman, Kimpel, & Crout, 2006; Tetreault & El-Geneidy, 2010). Instead, this thesis aimed to provide a robust vehicle for high-level sketch planning for minor route changes, regardless of location or ridership. This work is intended to provide a flexible, long-term solution which can be applied to any standard fixed route, and can be rerun as desired as new data is generated. A high level of precision was neither sought nor achievable using such widely-applicable models. Paratransit operation was not considered.

1.3 Objective

The first objective of this research was to define and, where necessary, create models to predict changes in bus performance and passenger travel times, given user-specified stop layout changes to an existing transit trip pattern. The second objective of this thesis was to build a data analytics platform, powered by the prediction models, in order to calculate and present ten Key Performance Indicators (KPIs). Ultimately, this integration of predictive models into a data analytics framework was intended to facilitate fact-based decision-making for bus stop service planning. Specifically, the models were designed to make use of archived AVL, APC, and schedule data to automatically calculate the following KPIs as a means of measuring the effects of hypothetical SR, SA, and introduction of limited-stop operation. These KPIs are described in greater detail in Chapter 0:
i. Total bus run time saved/added, broken down by stop;
ii. Bus headway increase/decrease;
iii. Bus cycle time increase/decrease;
iv. Average bus run speed increase/decrease;
v. Productive capacity increase/decrease;
vi. Maximum Load Section load factor increase/decrease;
vii. Total passenger in-vehicle time saved/added, broken down by stop;
viii. Total passenger access time saved/added, broken down by stop;
ix. Total passenger wait time saved/added based on changes to vehicle headways, broken down by stop;
x. Number of existing passengers expected to prefer limited-stop over local operation, given the particular stop layout the user has selected.

1.4 Case Study: London Transit Commission (LTC) Network

While the methodology used herein is generally applicable for any transit agency, the research was completed using data from the archived databases of the London (Ontario, Canada) Transit Commission (LTC). Unfortunately, the transit vehicle sensor equipment industry has not yet matured to a point where data formats, units, and storage are standardized using a universal ontology akin to the GTFS standard for schedule data. As a result, the data cleaning processes and some analytical procedures contained herein were tailored to accommodate the idiosyncrasies of LTC’s dataset, and may not be directly applicable or optimal for analysis of automatically generated data from other transit agencies.

Founded in 1826, London, Ontario, Canada is a city of 366,000 people with a land area of roughly 420 km², resulting in a relatively low urban population density of 870/km². Like many midsize cities in Canada, population growth and associated development has been sporadic throughout its history, and annexations of numerous nearby towns has resulted in a relatively large urban area with low density. As a regional education centre for southwestern Ontario, the city has a disproportionately large post-secondary student population of roughly 50,000 – 80,000 full and part-time students, or 14-22 percent of the total population.
The London Transit Commission (LTC) was founded in 1875. As of 2014, the LTC’s network included 42 routes, of which three are community bus routes and two are express routes. The agency served 23.8 million passengers in 2013 (London Transit Commission, 2014), and has roughly 244 buses in active service, all of which are outfitted with a GPS-based CAD/AVL system. Sixty-six vehicles, or roughly 25% of the active fleet, are equipped with APCs. As of 2014, the LTC is in the process of installing a smart card fare media system, and is considering conversion of two routes to Bus Rapid Transit operation (London Transit Commission, 2014).

Figure 1: Location of London, Ontario, Canada (Google Maps, 2015)
The LTC’s mode share compared to other forms of transportation in London is relatively low at roughly 7% as of 2011 (Statistics Canada, 2011). Ridership is thought to consist largely of “captive” riders, including many students at the post-secondary institutions, which are relatively well served by LTC service. For the sake of this study, owing to the low ridership and “captive” nature of most riders, significant changes to ridership as a result of minor changes to service quality was considered unlikely.

The LTC also provides paratransit service. Operation of paratransit service is not considered in this thesis.
1.5 Precursory Data Analytics Project

Citing frustration with the cost and delays associated with manual system performance KPI generation, the LTC engaged IBI Group and the University of Toronto jointly under an NSERC ENGAGE partnership in October 2013 to develop near-real-time Business Intelligence (BI) dashboards for descriptive reporting of KPIs. The LTC provided copies of their raw APC and AVL databases to function as the source data for the BI tools.

Key Performance Indicators, with respect to system performance, were defined by the LTC: schedule adherence, percent seated capacity, overload events, percent scheduled run time served, and running time adequacy. The dashboards allowed the user to view the KPIs across several dimensions (specifically system, route, trip, and stop), so information could be displayed anywhere from a high level (e.g., system-wide average occupancy of buses) to a very detailed level (e.g., occupancy of a given bus on a given route on a given day). The BI tools were created with the goal of providing the LTC with better data, regularly updated, with which to make timely decisions.

Because the BI tools were strictly intended to be descriptive reporting tools, no diagnostic, predictive, or prescriptive analytic capabilities were incorporated (refer to Chapter 2.1 for further discussion of data analytics and associated terminology). This limited the user to making observations related to past performance.

With this in mind, given the rich data source that the LTC had made available for the project, the University of Toronto was eager to further the research, particularly into the realm of predictive analytics: given historical performance and hypothetical changes to a route’s stop layout, how would such changes affect the route’s future performance?

1.6 Thesis Organization

This balance of this thesis is organized into six chapters. Chapter 2 comprises a review of literature related to data analytics, analysis of archived AVL and APC data, and SR, SA, and limited-stop operation. Chapter 3 describes the sources of data used to develop the KPI models, as well as the general data cleaning process and limitations of the available AVL and APC
databases provided by the LTC. Chapter 4 discusses the analysis methodology and defines the models used to compute the KPIs. Chapter 5 discusses the software platform created to display and run the models in order to return feedback to the user. Chapter 6 contains sample results and a discussion of model validation. Finally, Chapter 0 summarizes the project, and provides limitations, concluding remarks, recommendations, and suggestions for further research.

Throughout this paper, the term “boardings” will be used interchangeably with “ons”, as will “alightings” with “offs”.
2 Literature Review

Broadly speaking, this chapter attempts to describe three areas of ongoing research: implementation of data analytics (and, by extension, fact-based decision-making); analysis of archived (i.e., automatically generated) AVL and APC data; and performance improvements resulting from transit route adjustments (specifically, stop removal, stop addition, and partial conversion of existing local transit operation to limited-stop operation). In keeping with the theme of this thesis, it is the high-level synthesis of these research areas in particular which is most relevant, but this synthesis is also scarce in existing literature. For brevity, data analytics will be defined only in the context of archived transit data (Chapter 2.1), and discussion of archived data analysis will imply use of data analytics (Chapter 2.2). The motivation and processes for route adjustment will be examined briefly in their own right (Chapter 2.3), followed by a discussion of the fusion of the three research areas – data analytics, archived transit data, and transit route adjustments to improve performance – in Chapter 2.4. Chapter 2.5 will summarize the literature review, and discuss gaps in existing research that this thesis attempts to fill.

2.1 Data Analytics for Transit Operations and Planning

Data analytics refers to the process of analyzing typically raw data for the purpose of making informed decisions. Similarly, as the name implies, fact-based decision-making refers to the practice of basing decisions upon facts revealed through data analytics, rather than on gut instinct or political expediency. As Provost and Fawcett note in their admirable introductory work on the broader topic of data science (2013), fact-based decision-making need not be used in an all-or-nothing manner, but where used, its benefits are tangible and have been demonstrated decisively. The authors summarize the conclusions of a 2011 study by Brynjolfsson, Hitt, and Kim, in which the researchers developed a means of objectively (statistically) quantifying the rates at which firms engage in fact-based decision-making in comparison to their productivity. In short, moving one standard deviation higher on the study’s “Data-Driven [fact-based] Decision-Making” scale was found to be associated with a 4-6% increase in a firm’s productivity, even after controlling for many possible confounders.
For the purposes of this thesis, the term “data analytics” will be categorized as described by Simchi-Levi (2014), and by Kothari (2015) in brackets. The four basic levels of data analytics are described as follows:

i. **Descriptive analytics (Profile):** What happened?

ii. **Diagnostic analytics (Segment):** Why did it happen?

iii. **Predictive analytics (Predict):** What will happen?

iv. **Prescriptive analytics (Optimize):** How can one make x happen?

Transit planning and operations can benefit from every level of data analytics. Descriptive analytics, as exemplified by the set of tools created for the LTC as described in Chapter Error! Reference source not found., can be used by a transit agency to better understand its network and performance. As in the case of the LTC, descriptive tools represent the obvious starting point for any transit agency looking to build a suite of data analytics tools. Descriptive tools allow a transit agency to efficiently explore its archived data, and answer such simple questions as, “Was vehicle no. 2532 indeed at the intersection of Richmond and Oxford at the time of the alleged incident?” With little effort, a transit agency can use descriptive tools to aid manual diagnostic analysis of its data using the human brain as the diagnostic analytic tool. Alternatively, a transit agency looking to streamline its diagnostic analysis could add a second layer of sophistication by employing data mining tools to automate the process, supplementing and perhaps one day supplanting altogether this manual diagnostic work. Diagnostic tools, whether human or machine, can be used to determine which routes, vehicles, operators, times of day, days of the week, months, etc. are associated with a higher incidence of incidents, inferior schedule adherence (on-time performance), or lower percent-seated capacity, for example.

By predicting specific results given a controlled set of inputs, the third layer of data analytics (predictive analytic tools) enables a transit agency to answer “what if” questions such as “If bus Stop 962 is deleted, how will this affect total run time?” Predictive analytics also facilitate automated reporting, informing the transit agency of impending overload events or impending equipment failures, for instance. The fourth and final layer of tools, prescriptive analytic tools, are in a sense the converse of predictive tools. The use of prescriptive analytics allows the transit
agency to ask questions such as, “How should stop placement along Route 88A be modified to decrease passenger access/egress distance to an average of 400m?”

As introduced in Chapter 1.3, the primary objective of this thesis is to provide predictive analytic tools which can estimate KPIs as a means of answering “what if” questions for transit operations and planning analyses at the stop level.

2.2 Use of Archived AVL and APC Data for Performance Improvement

Chapter 2.1 describes the general processes by which data analytics can be employed to aid transit planning and operations. Chapter 2.2 delves into strategies, benefits, and challenges of using archived AVL and APC data to improve transit performance.

Transit Cooperative Research Program (TCRP) Report 113 (2006) provides an excellent, high-level overview of use of archived AVL and APC data for transit performance improvement. Similar topics are covered in TCRP Web Document 23 (2003), and Hammerle, Haynes, and McNeil (2005). Historically, AVL systems were installed with the intention of providing the real-time data required to operate Computer Aided Dispatch (CAD) systems. The authors of TCRP 113 report that at the time of writing (2006), AVL data was typically not stored and processed in such a way that post-operation analysis would be feasible, in contrast to data from APC systems. However, the benefits of storing AVL data with post-operation analysis in mind have been discussed in literature since at least 2004 (Furth, Muller, Strathman, & Hemily, 2004), and have been discussed in industry for more than two decades. While transit agencies have since begun to make use of archived AVL data for performance measurement and improvement purposes, limitations posed by the (lack of) granularity and accuracy of AVL data persist. The report goes on to describe existing practice and recommended strategies for improving current practice, and though nearly a decade old, virtually all of its recommendations remain valid today. This is reflective of the transit industry’s technological lethargy in relation to that of more dynamic, less infrastructure-heavy industries. That said, the ability to process and review large
quantities of data affordably has improved vastly over the past decade, and many transit agencies across the world are starting to take full advantage therein.

The recommendations within TCRP 113 include frequent (granular) recording of data; a few basic strategies for cleaning archived data, such as balancing passenger counts at points of known zero passenger loads to minimize the effects of error propagation from one trip (or trip segment) to the next; the benefits of integrating on-board sensors (such as APC, AVL, and Automatic Fare Card systems); and a variety of other suggestions. Crucially, this report includes a discussion of the value of “time-at-location” data – informative and often crucial for post-operation analysis – and the pitfalls of attempting to rely upon “location-at-time” polling data. Furth, Hemily, Muller & Strathman (2003, 2004) also provide a comprehensive description of the type and granularity of descriptive analytic reporting that can be achieved with various types and granularities of AVL data.

The benefits of gathering and analyzing archived AVL and APC data are nearly limitless, but virtually all can be categorized as contributing to one or both of two primary end goals: improvement of operational efficiency, and improvement of service quality. Automatic (rather than manual) generation and gathering of data vastly increases the available sample size, which allows for more granular analysis of data (e.g., ability to analyze on-time performance at a trip or route level, rather than at the system level); more rapid analysis of data (where a minimum sample size is required); and more reliable estimation of values with high variability (Furth, Hemily, Muller, & Strathman, 2006). Vehicle reliability analysis is improved. For example, with sufficient and sufficiently reliable AVL data, approximation and short-term prediction of vehicle dwell times is achievable. One such analysis (Prashidi & Ranjitkar, 2013) found that a Wakeby distribution was a better approximation of vehicle dwell times than either normal or lognormal distributions – a potentially useful finding for transit system modellers, and one that is unlikely to have been gleaned without large-scale archived data available. Mandelzys and Hellinga (2010) found that performance measures which appear satisfactory when relying upon manually gathered data no longer appear adequate when large volumes of AVL and APC data are used instead, indicating that a transit agency’s increased ability to measure its performance generally
spurs a desire for proportional improvements in performance. Under-performing stops, routes, operators, vehicles, schedules, etc. can be identified far more easily when performance statistics are put under a microscope which allows its operator to drill up and down through the data at will.

It is important to note the risks of using data analytics to assess archived data. While such techniques tend to greatly reduce human error, machine error (such as sensor miscalibration) is introduced. Additionally, scaling up performance analysis to include millions or billions of data points makes it impossible for the human eye to review every calculation. Systemic mistakes in performance analysis tend to camouflage and can be difficult to root out, especially where they fail to cause remarkable results. One inevitable solution is to create algorithms to automate the quality assurance process, employing strategies and screening procedures such as those described by Hamad and Quiroga (2014) for assessment of archived traffic sensor data. Simple spot-checking to ensure correct calculations and plausible results is also necessary, though not sufficient.

Though not directly applicable to this thesis, a paper by Gittens and Shalaby (2015), which proposed a new objective measure of transit reliability, also made use of the LTC’s archived AVL and APC dataset. The paper sought to provide a greater degree of accuracy than industry-standard measures by using context-specific values for model parameters in place of system average or typical industry values. This thesis took a similar approach by relying upon route- and stop-specific data in place of industry standard values wherever possible.

2.3 Stop Removal and Addition, and Limited-Stop Operation

The term Stop Removal (SR) refers to the process of deleting an existing stop along a route, or deleting a number of existing stops and replacing them with some lesser number of new stops in alternate locations. Stop Addition (SA) is the reverse process. SR and SA can have a significant impact on operational efficiency and ridership, but making modifications to existing transit routes is also often politically charged, so the process by which stops are selected for SR and SA is not simple. Strategies for determination of optimal stop spacing, such as the methodologies
described by Medina-Tapia, Giesen, & Munoz (2013) and Li & Bertini (2008), rely on many assumptions, such as known passenger demand, high-frequency service, no congestion, and that buses have sufficient capacity to load all users present at a stop. But as Wagner and Bertini (2014) point out, though many strategies for determining optimal stop spacing have been proposed, strategies for performing SR and SA are sparse. Wagner and Bertini provide a straightforward benefit-cost analysis method, which also relies upon a number of assumptions (including no changes to ridership, all stops along a route are served, and access to stops provided by contiguous streets in a grid pattern). Their work bleeds into the realms of prescriptive analytics and optimization, but only scratches the surface.

Using dynamic programming to estimate the effects of changing bus stop spacing, (Furth & Rahbee) found that bus run time savings following an optimized SR program would be in the order of 14-15 seconds per stop, of which nine seconds would be saved due to elimination of acceleration and deceleration time. When estimating SR effects during a study in Seattle, Washington, de Vries Kehoe (2004) used an initial “rule of thumb” of 20 seconds per deleted stop. Later, de Vries Kehoe produced a regression model to estimate effects of SR observed during the study’s test case, simplified for situations where ridership change was considered to be negligible:

\[
\Delta t = (-0.0015 \Delta s) \times (d)
\]

*Where:*

\(\Delta t\) = change in average travel time (minutes)

\(\Delta s\) = change in bus stop spacing (feet)

\(d\) = length of project segment (miles)

One reason for a transit agency to undergo a program of SR is the introduction of limited-stop operation. Limited-stop operation, in contrast to local operation, refers to the practice of deleting or bypassing the majority of stops along an existing route, or creating a new route with wider-than-average stop spacing of 450 – 1,600m (Conlon, Foote, O’Malley, & Stuart, 2001; Tetreault & El-Geneidy, 2010). This differs from express, skip-stop, short-turn, and zonal operation, none
of which are considered expressly in the context of this thesis. Limited-stop operation has been shown to be an effective means of decreasing bus run time and improving customer perceptions of service quality (Schwarcz, 2004; El-Geneidy & Surprenant-Legault, 2010; Vuchic, 2005; Conlon, Foote, O’Malley, & Stuart, 2001). However, passenger perceptions of time savings have been shown to overestimate the benefits of limited-stop operation for passengers (Schwarcz, Service Design for Heavy Demand Corridors: Limited-Stop Bus Service (Master of Science in Transportation thesis), 2004; El-Geneidy & Surprenant-Legault, Limited-stop Bus Service: an Evaluation of an Implementation Strategy, 2010). Recommended stop spacing for limited-stop operation has been examined and proposed by many studies (Conlon, Foote, O’Malley, & Stuart, 2001; Tetreault & El-Geneidy, 2010; Vuchic, 2005). Schwarcz (2004) provides a methodology for determining what proportion of a route’s bus fleet should be converted to limited-stop operation based on estimated passenger origin-destination matrices. Schwarcz’s methodology builds on Navick and Furth’s method for estimating origin-destination matrices (1994), using one-directional downstream distance as the impedance function for a gravity-based model formulation.

TCRP Synthesis 110 (Boyle, 2013) offers a wide-scale survey of transit agencies’ approaches to improving transit bus speeds. The report found that three-quarters of transit agencies engage in route adjustment procedures with an eye to improving bus speeds, yet only six of 39 responding transit agencies reported measuring the specific impacts of such route adjustments. Of these six, four agencies realized small increases to average speed, one agency found small decreases to average speed, and one agency stated that measuring the effects of route adjustments was too complex to provide a single overall outcome. Similarly, while 16 of 39 responding agencies reported introducing at least one limited-stop service as a means of improving bus speeds, only four agencies reported measuring the impact of introducing the limited-stop service, of which two reported minor increases in average speeds, and two reported moderate increases. The report also found that two-thirds of responding agencies made changes to stops in order to improve bus speeds, of which 79% increased stop spacing (i.e., consolidated stops). Stop removal was considered to be the most effective means of increasing average bus speeds. Overarching strategies for implementing SR included setting higher bus stop spacing standards, and aiming
for increased average stop spacing. Only two agencies reported measuring changes to average bus speeds as a result of SR, with one agency reporting an increase of average speed from 8.8 to 9.8 km/h along a one-mile segment of a route. Another common stop adjustment strategy undertaken by responding agencies included replacing near-side stops with far-side stops.

TCRP Report 165, better known as the Transit Capacity and Quality of Service Manual, 3rd edition, or TCQSM (Kittelson & Associates Inc., Parsons Brinckerhoff, KFH Group Inc., Texas A&M Transportation Institute, & Arup, 2013), provides guidance for evaluating and measuring quality of service; capacity, speed and reliability performance; and sizing of transit stops, platforms, and stations. It serves as a useful reference for SR, SA, and limited-stop analysis. Information specifically related to standard scheduling procedures can be found in TCRP Report 135 (Boyle, et al., 2009), though limited-stop operation is discussed only very briefly.

2.4 Data Analytics for Route and Stop Changes

The application of data analytics to process and visualize archived transit data to assess SR, SA, and limited-stop operation is a relatively recent development, and the majority of published examples describe research studies of limited scope intended to gauge feasibility of wide-scale implementation. A pioneering example of descriptive analytic work in this field by El-Geneidy, Strathman, Kimpel, & Crout (2006) used data from Portland’s TriMet’s AVL and APC systems to examine changes in passenger activity, reliability, and bus run times of buses as a result of SR along a segment of a route. Three months of passenger counts and bus running times between consecutive timepoints were gathered for the route of study for the time period roughly six months prior to SR, and an additional three months of data was gathered roughly six months following consolidation. Before-and-after differences between the two datasets were computed and compared to before-and-after differences observed on nearby segments of the same route which did not undergo SR. The authors found that while bus running times along the test segment were reduced by approximately six percent (or 5-6 seconds per stop), their regression model estimated a run time reduction of 42.2 seconds per consolidated stop, suggesting that the vast majority of potential time savings was not realized. No significant effect on passenger activity was observed; a similar result was noted following a SR study by de Vries Kehoe
(2004). Notably, El-Geneidy et al. found no evidence of reliability improvement (running time or headway variability), though further improvements to bus running times were considered achievable with appropriate schedule adjustments. Schedule adjustments to take advantage of improved running times would also tend to decrease variability.

Tetreault & El-Geneidy (2010) used archived AVL and APC data from Société de transport de Montréal (STM) in conjunction with the 2003 Montréal Origin-Destination survey to predict bus run times for a soon-to-be-implemented limited-stop service (467 Express Saint-Michel) running parallel to an existing, heavily used bus route (67 Saint-Michel). The study found that implementing the limited-stop service would significantly reduce the run times below existing run times for both the new and the existing service. The 467 route is in operation to this day, indicating that the STM has been satisfied with its performance. To the authors’ knowledge, as of the time of writing, the transit industry had no well-defined criteria for selecting which stops should be served by a new limited-stop service based on existing passenger demand and vehicle performance. Seemingly, this has not changed in the years since.

In a follow-up descriptive and diagnostic analytic study using archived data from STM’s newly established route 467, Diab & El-Geneidy (2012) evaluated a collection of strategies for improving bus service on their effects on bus run time and passenger perception. The strategies included introduction of fare cards, limited-stop operation, bus-only lanes, articulated buses, and introduction of transit signal priority. The combination of strategies was found to reduce bus run time by about 10.5 percent along the route with limited-stop operation. Passengers were found to be satisfied, and indeed overestimated the benefits of the service improvements, as has been found in other studies (Schwarcz, 2004; El-Geneidy & Surprenant-Legault, 2010). Passengers were also found to be willing to walk further to access the limited-stop service (Route 467) instead of the local operation service (Route 67), though this increase in walk distance was not described quantitatively in the paper.

A paper by El-Geneidy, Horning, & Krizek (2011) used predictive analytics to generate visual and tabular (text-based) predictions of vehicle run time, schedule adherence, and reliability for Minnesota’s Metro Transit. Analysis was conducted at the timepoint and route levels, using
statistical models to study the effects of route length, number of stops, and passenger activity on bus run time and schedule adherence. The paper concludes that changes to existing schedules are required in order to improve bus run time and schedule adherence. Each scheduled stop was found to add 0.9 percent to schedule deviation, which translated to only about three seconds of additional run time. The authors note that not all stops were well utilized, and recommended SR accordingly. Further, this paper noted the impracticality of trying to assess schedule adherence and variation at the route level, concluding that analysis must be conducted at the trip pattern or trip segment level, subcategorized by time of day, in order to avoid measurement error.

Schwarcz & Wyss (2013) compared historical bus run times from existing local and limited-stop services operated by New York City Transit to resulting bus run times following a pilot program which converted all timepoints to regular stops. The goal of the pilot was to decrease run times by eliminating bus holding at timepoints. One critical function of using a timepoint system is increasing reliability, building in buffer time to allow transit vehicles to maintain their schedules. However, the study found that while timepoint re-designation had the desired effect of decreasing average bus run times, and reducing delays for passengers, there was no significant impact on bus reliability. Passenger perceptions of the services also generally improved. It is worth noting that run times and reliability were only measured at the trip level.

Reilly (2014) used archived AVL data to develop tools to estimate optimal running times between timepoints for Albany, New York’s Capital District Transportation Authority. New scheduled times were assigned such that buses would have a probability of 95 percent of being ready to depart the following timepoint on schedule. Though theoretically possible to reduce a route’s fleet size as a result of increasing average trip speed, the study found that this did not occur in practice.

2.5 Summary of Reviewed Literature and Gaps in Research

Fact-based decision-making is growing in popularity across virtually every major industry, and studies have shown statistically the associated benefits to productivity. Various types of data
analytics, including descriptive, diagnostic, predictive, and prescriptive analytics, are employed to facilitate fact-based decision-making.

Transit agencies are no exception. Agencies have been reviewing archived data from ADC systems for decades, with the aim of gaining insights to tackle inefficiencies and reoccurrence of preventable mistakes. Specifically, AVL data, and more recently, APC data, have begun to be assessed post-operatively to improve future decision-making. Many challenges persist in the transit industry, however, and much of the data gathered by ADC systems is either erroneous, discarded altogether, stored without sufficient granularity, or recorded in a way more relevant to real-time than post-operative use of data.

A common use for archived ADC data is in assessing stop layouts along existing routes. Average local bus stop spacing in North America tends to be in the order of 160-230m, while studies have shown optimal spacing is typically significantly wider (Li & Bertini, 2008). Considerable research has thus been conducted using archived ADC data to justify implementation of stop consolidation programs, under which a number of stops are removed, and a lesser number of stops are added. Estimates have been produced for vehicle run time saved per deleted stop, both as a general rule, and for individual routes at individual agencies.

Published transit data analytics research is relatively thin. One (Portland, Oregon) study’s analytic analysis showed an estimated bus run time savings of about 6 seconds per deleted stop, contrasting sharply with the study’s linear regression estimate of roughly 42 seconds saved per deleted stop for the same route subsection (El-Geneidy, Strathman, Kimpel, & Crout, 2006). Other studies have been conducted to determine the benefit to passengers and transit agencies of introducing limited-stop operation in parallel to an existing local route (Tetreault & El-Geneidy, 2010; Diab & El-Geneidy, 2012; Schwarcz & Wyss, 2013). One paper used predictive analytics to generate visual and text-based predictions of vehicle run time, schedule adherence, and reliability for Minnesota’s Metro Transit (El-Geneidy, Horning, & Krizek, 2011).

Thus, while a handful of studies have used ADC-based data analytics to assess transit stop layout changes, the majority have focused on calculation of only one or two KPIs (e.g., bus run time saved), and tend to represent one-time analyses for a very limited number of routes, for a single
transit agency. In contrast, this thesis seeks to provide analysis for every route in the LTC network. Rather than providing a one-time analysis, the tools produced as part of this thesis allow for analysis of data gathered over any time period in the dataset. This thesis also calculates 10 KPIs, of which the following have not been presented in literature related to transit analytics to date:

- Bus headway increase/decrease;
- Bus cycle time increase/decrease;
- Productive capacity increase/decrease;
- Maximum Load Section load factor increase/decrease.

This thesis also contains new models for estimating the following metrics:

- Average boarding and alighting volumes;
- Likelihood of vehicle dwell;
- The number of passengers expected to prefer a hypothetical limited-stop service over an existing local service.

Finally, while this thesis uses the LTC network as a case study to showcase the tools produced, the analysis has been completed and tools have been designed to accommodate data from any transit agency, provided the data has been cleaned and formatted appropriately.
3 Data Sources, Cleaning, and Limitations

3.1 Data Sources

The primary data sources used in this analysis were the LTC’s archived AVL and APC databases. Trapeze was the vendor responsible for creating the database platform used by the LTC (TransitMaster™ Intelligent Transportation System). The data was housed and managed using Microsoft SQL Server 2012.

Pertinent scheduling information was incorporated into the raw AVL database by the LTC (and/or Trapeze) as a matter of course prior to data handover. Though data from 2008 through to September 2013 was made available by the LTC to facilitate the research, only data from 1 September 2011 through 31 August 2013 was analyzed in an effort to minimize processing times required by the software produced in pursuit of this thesis. After basic cleaning processes were implemented, the unified dataset used in this analysis consisted of approximately 10.7 million rows of data, with roughly 25 fields.

Google Maps was used to estimate geographic coordinates for stops in the LTC’s network where no coordinates were contained in the LTC database. The geographic coordinate imputation process is described in Chapter 3.2.

3.2 Cleaning of Inaccurate, Missing, and Irrelevant Data

Data cleaning algorithms, and models used to calculate the effects of SR and SA, were designed to address the peculiarities of the LTC’s archived data. To date, data from every transit agency has peculiarities specific to the agency’s unique combination of equipment and databases. Though the general data cleaning and modelling concepts described below are applicable from one agency to the next, the specific processes must be tailored to fit an agency’s data.

The archived data did not suffer materially from syntactic inaccuracies (such as the presence of integer values in fields where strings were expected). Where observed, these errors were easily corrected or bypassed through use of other data sources.
Semantic accuracy is violated where a data point belongs to the correct domain, but is incorrect – for example, where the AVL system records time of departure as 9:04, but the actual time of departure is 11:15. Semantic accuracy of the LTC’s dataset was challenging to verify due to magnitude of the database, and the impossibility of identifying one-off semantic inaccuracies. However, two critical systemic semantic inaccuracies were noted. First, calculated dwell times (and door open/close times) were determined to be suspect. The AVL system recorded actual arrival and departure times, and initially, the dwell time for each stop was calculated as follows:

\[
\text{Dwell Time} = \text{Actual Departure Time} - \text{Actual Arrival Time}
\]

However, in many cases, dwell time was found to be as low as zero seconds, even when the APC system recorded a nonzero number of ons and offs (boardings and alightings). The implications of this are discussed in Chapter 3.3.

Second, recorded passenger ons and offs were found to be unreliable at stops where the AVL system failed to confirm its location by GPS: rather than recording null (i.e., no) information for the number of ons and offs, or identifying the location based on known stop sequence and thus correctly recording the number of ons and offs, the APC system recorded zero ons and zero offs except at terminus stops. Oddly, in all cases where stop geographic coordinates were successfully identified and recorded by the AVL system, the sum of ons and offs was never equal to zero, indicating that the AVL system may not have recorded geographic coordinates at stops where no passengers boarded or alighted from buses. Implications are discussed in Chapter 3.3.

Finally, a third and less significant semantic inaccuracy was noted. For a small number of stops in the AVL database, multiple values for distance from previous stop were recorded within a single trip pattern. Clearly erroneous, this problem was addressed in two ways: where the stop in question was the first stop along a trip pattern, its value was set to zero, and in all other cases, distance to next stop values were retrieved from a static scheduling table (where values were not recorded by the ADC systems, but rather input manually).
For the remaining fields, as no other systemic inaccuracies were detected, it was deemed reasonable (and more practical) to assume that undetected semantic inaccuracy was minimal and would not significantly affect the validity of results.

All stops defined in the AVL database had a location description, and the vast majority (> 95%) had associated geographic coordinates stored in adjacent fields. In cases where geographic coordinates were not recorded in the database, values were inferred by entering the stop description (e.g., Bandbury & Pond Mills EB) into Google Maps, and confirming the approximate coordinates using Google Maps Street View.

Trips deemed not relevant were removed from the dataset. For example, non-revenue trips were removed to avoid skewing average passenger on and off counts. Similarly, trips lacking critical data (such as a trip pattern identification number) were disregarded as a matter of course.

3.3 Data Limitations

Given this study’s overarching goal of estimating the effects of SR and SA as a function of historical observations, the benefit of using recorded (observed) dwell time data from individual stops is difficult to overemphasize. Unfortunately, while the LTC’s AVL system recorded time-at-location data – which is preferable to location-at-time data for post-hoc analysis of archived data (Furth, Muller, Strathman, & Hemily, 2004) – data required to calculate dwell times was only recorded at timepoints. This lack of granularity guaranteed that, at best, dwell times could be used directly to estimate the effects of SR and SA for a small minority of stops. Effort was thus spent attempting to produce a regression model to estimate dwell time at non-timepoints as a function of passenger ons and offs, bus departure load, and a number of other explanatory variables. These attempts included linear and nonlinear regression.

However, dwell times suffered from more than lack of granularity. As discussed in Chapter 3.2, in many cases at stops where the APC system recorded a nonzero number of ons and offs, calculated dwell time was found to be as low as zero seconds. Clearly, such dwell times could not be relied upon, given an assumption of correct APC readings. Discussions with LTC staff confirmed that APC readings were calibrated at the time of installation, and that APC equipment
has since undergone calibration checks and readings were found to be quite accurate (+/- 3%). Moreover, LTC staff stated that various conditions (including shutting down and restarting a bus at a layover) would cause the AVL system to record incorrect arrival and/or departure times. As a result, no attempted regression model was able to estimate dwell time with an $R^2$ value greater than 0.26. Goodness-of-fit statistics described in literature are typically much higher; Tirachini (2013) produced values ranging from 0.89 to 0.92, while Rajbhandari, Chien, & Daniel (2003) produced values ranging from 0.57 to 0.91. Given that dwell times calculated from recorded bus arrival and departure time data were considered unreliable, average boarding and alighting counts were instead calculated and fed into the dwell time estimator provided with the TCQSM, 3rd edition (2013); this process is discussed in detail in Chapters 4.4 and 4.5.

As discussed in detail in Chapter 3.2Error! Reference source not found., recorded values of passenger ons and offs were considered to be unreliable at mid-route stops where the AVL system failed to identify its location using the GPS; it may also be the case that the AVL system failed to record its geographic coordinates at stops where no passengers boarded or alighted, except at terminus stops. Rows containing unreliable data in these crucial fields were treated as missing and excluded from the dataset. One regrettable corollary to this exclusion is that, for all rows (stops) which remained in the cleaned dataset, the sum of passenger ons and offs was nonzero, indicating that the bus dwelled at every stop, in all cases. A bus’s likelihood of dwelling at any given stop was therefore calculated as 100%. The various models described in Chapter 4 were initially designed to take likelihood of dwell into account, but the cleaned dataset functionally imposed the assumption that buses stopped at every stop along a route.

A final noteworthy limitation to the dataset was the lack of Automatic Fare Card (AFC) data with which origin-destination matrices could have been estimated. While the LTC’s AFC system is expected to be up and running in the near future, no data was available at the time of this study. The methodology used to estimate origin-destination matrices is described in Chapter 4.11.
4 Methodology and Models

Chapter 2 examined existing literature related to SR, SA, and limited-stop analysis using archived AVL and APC data, and Chapter 3 described the limitations to and preparation of the LTC’s archived data for the sake of this study. This chapter will explain this study’s methodology for estimating the effects of SR and SA along the LTC’s bus routes, and describe the models upon which the methodology is built.

4.1 Basic Methodology

While previous studies have focused SR or limited-stop assessments on a small number of important or representative routes, the intention of this study was to produce an automated, standardized approach that could be applied to any route, given the necessary raw data. Processes were automated such that the user was only required to input a trip pattern of interest, changes to the existing stop layout along that trip pattern, and filter parameters of interest (refer to Chapter 5.1 for further discussion of filters). The effects of SR, SA, and limited-stop operation were measured using Key Performance Indicators (KPIs) as defined in Chapter 0. A more detailed breakdown of the calculations involved in computing the KPIs is found in Chapters 4.4 through 4.12. Chapter 4.13 provides a summary of the chapter.

Figure 3 shows a high-level view of the models and KPIs contained in this thesis, as well as the filters available to the user (refer to Chapter 5.1 for discussion of filters).
Trip patterns, rather than routes, are the basic building blocks for trip and run organization by a transit agency. As such, this software was designed to accommodate modifications to existing bus trip patterns rather than existing bus routes. To illustrate the logic behind this, consider that each bus route in the LTC’s system had anywhere from 10-30 patterns, all varying by time of day, day of the week, etc. (Similar volumes of trip patterns per route were found in Minneapolis’s Metro Transit system by El-Geneidy, Horning, & Krizek (2011)). This made

Figure 3: Flowchart showing high-level methodology
analysis of modifications to a route only feasible at the pattern level, as passenger count aggregations averaged across many patterns would produce skewed results. For example, averaging passenger counts from a short-turn pattern with a full-route pattern would be likely to return a large number of alightings (and few boardings) at the last stop on the short-turn route. This information would be misleading for planners looking to modify a specific trip on a specific day.

Algorithms were created to clean the LTC’s archived data; the cleaned data would be aggregated and extracted for the user’s trip pattern of interest, according to filters applied by the user. The extracted data was then used to power a number of simple models estimating passenger relocation, dwell times savings/penalties, etc. (described in Chapters 4.4 through 4.12), based on the user’s desired stop layout changes. The models were used in combination to compute the desired KPIs.

It is worth reiterating that this automated analysis could be applied to any trip pattern for any route in the LTC’s archived dataset. One challenge of creating an automated tool which can be applied to literally billions of filter/trip pattern/stop layout change combinations is the difficulty in assessing the statistical significance of the results. To avoid providing detailed, automatically generated statistical inferences which could not be validated by the user of the software, a simple method was devised to provide the user with a basic indication of the representativeness of the raw data upon which the results were based. The number of rows (samples) of data aggregated to compute the average passenger on and off count for each stop was returned to the user alongside the results of the automated analysis. In other words, the software reported the number of bus trips which were averaged to produce the mean number of ons and offs at each stop.

For simplicity, formulae which relied upon, or calculated adjustments to, passenger offs will not be described except where deemed necessary. Formulae which relied upon, or calculated adjustments to, passenger ons are shown throughout. Calculations based on passenger offs were carried out using the same methodology as passenger ons, except where noted.
4.2 Key Performance Indicators (KPIs)

The effects of SR, SA, and limited-stop operation were measured using the following KPIs, of which the first nine apply to SR analysis, SA analysis, and limited-stop analysis; result (29x) applies only to limited-stop analysis. KPIs (i) through (iii) are relevant to the transit agency; KPIs (vii) through (x) are relevant to passengers; and KPIs (iv) through (vi) are relevant to both the transit agency and passengers, and will thus be referred to as system-based KPIs. More detailed calculations and explanations can be found in Chapters 4.4 through 4.11.

i. Total bus run time saved/added, broken down by stop;
ii. Bus headway increase/decrease;
iii. Bus cycle time increase/decrease;
iv. Average bus run speed increase/decrease;
v. Productive capacity increase/decrease;
vi. Maximum Load Section load factor increase/decrease;
vii. Total passenger in-vehicle time saved/added, broken down by stop;
viii. Total passenger access time saved/added, broken down by stop;
ix. Total passenger wait time saved/added based on changes to vehicle headways, broken down by stop;
x. Number of existing passengers expected to prefer limited-stop over local operation, given the particular stop layout the user has selected.

**KPI i (Total Bus Run time Savings/Penalty)** was defined as the estimated net total bus run time savings or penalty resulting from SR or SA changes to a trip pattern. The KPI was also broken down into savings/penalty per stop. This KPI was calculated as follows:

\[ \Delta \hat{T}_{bus} = \hat{T}_{bus_f} - \hat{T}_{bus_i} \]  

*Where:*

\( \hat{T}_{bus_i} = \text{initial average bus run time, excluding terminal time (s [seconds])} \)

\( \hat{T}_{bus_f} = \text{estimated final average bus run time, excluding terminal time (s)} \)
\( \Delta \hat{T}_{bus} = \text{estimated run time change resulting from changes to trip pattern (s)} \)

**KPI ii (Average Bus Headway Increase/Decrease)** was defined as the estimated net headway change resulting from SR or SA changes to a trip pattern, assuming constant before-and-after fleet size. This KPI was calculated as follows:

\[
\hat{H}_f = \frac{T_{ci} - \Delta \hat{T}_{bus}}{N_{bus}} \quad [2]
\]

\[\Delta \hat{H} = \hat{H}_f - H_i \quad [3]\]

Where:
- \( T_{ci} = \text{initial scheduled cycle time (s)} \)
- \( \Delta \hat{T}_{bus} = \text{time savings or penalty due to stop layout changes (s)} \)
- \( N_{bus} = \text{fleet size operating on pattern of interest (vehicles)} \)
- \( H_i = \text{initial scheduled headway (s)} \)
- \( H_f = \text{estimated final headway (s)} \)
- \( \Delta \hat{H} = \text{estimated change in headway due to changes to trip pattern (s)} \)

**KPI iii (Average Bus Cycle Time Increase/Decrease)** was defined as the estimated net cycle time change resulting from SR or SA changes to a trip pattern, assuming constant before-and-after fleet size. It was assumed that changes to terminal time would be proportionate to any change to run time caused by SR or SA. This KPI was calculated as follows:

\[ t_{tf} = \left( \frac{T_{bus_f}}{T_{bus_i}} \right) t_{ti} \]

\[ \hat{T}_{cf} = T_{ci} - t_{ti} + \hat{t}_{tf} - \Delta \hat{T}_{bus} \quad [4] \]

\[ \Delta \hat{T}_c = T_{ci} - \hat{T}_{cf} \quad [5] \]
Where:

\[ T_{ci} = \text{initial (scheduled) cycle time (s)} \]
\[ \hat{T}_{cf} = \text{estimated final cycle time (s)} \]
\[ \Delta \hat{T}_{bus} = \text{estimated time savings or penalty due to stop layout changes (s)} \]
\[ t_{ti} = \text{total initial terminal time (s)} \]
\[ \hat{t}_{tf} = \text{estimated final terminal time (s)} \]

KPI iv (Average Bus Run Speed Increase/Decrease) was defined as the estimated change in average speed resulting from SR or SA changes to a trip pattern. Average run speed represents the constant speed at which a bus would have to travel from departure at the first stop until arrival at the final stop along the trip pattern, such that the vehicle’s run time would equal the observed average run time. This KPI was calculated as follows:

\[ \bar{v}_i = \frac{D_{\alpha \Omega}}{\overline{T}_{busi}} \] \hspace{1cm} [6]

\[ \hat{v}_f = \frac{D_{\alpha \Omega}}{\hat{T}_{busf}} \] \hspace{1cm} [7]

\[ \Delta \hat{v} = \hat{v}_f - \bar{v}_i \] \hspace{1cm} [8]

Where:

\[ \bar{v}_i = \text{initial average run speed (km/h)} \]
\[ \hat{v}_f = \text{estimated final average run speed (km/h)} \]
\[ \Delta \hat{v} = \text{speed increase/decrease due to stop layout changes (km/h)} \]
\[ D_{\alpha \Omega} = \text{total trip distance from first stop (\(\alpha\)) to last stop (\(\Omega\)) (km)} \]
\[ \overline{T}_{busi} = \text{initial average bus run time (h)} \]
\[ \hat{T}_{busf} = \text{estimated final bus run time (h)} \]

KPI v (Productive Capacity Increase/Decrease) incorporated measures of a trip pattern’s initial and the final productive capacity. Vuchic (2007) defined productive capacity as “the best representative of mode performance: it is the product of speed, affecting primarily passengers,
and capacity, important for the operator”. Vuchic’s definition was used in this thesis, and the KPI was calculated as follows:

\[ \bar{C}_{Pi} = \bar{v}_i(C_{bus}) \]  
\[ \bar{C}_{Pf} = \bar{v}_f(C_{bus}) \]  
\[ \Delta \bar{C}_P = \bar{C}_{Pf} - \bar{C}_{Pi} \]  

Where:
- \( \bar{C}_{Pi} \) = initial (existing) productive capacity (psgr. km/h)
- \( \bar{C}_{Pf} \) = final (adjusted) productive capacity (psgr. km/h)
- \( \Delta \bar{C}_P \) = productive capacity change due to stop layout changes (psgr. km/h)
- \( \bar{v}_i \) = initial (existing) average run speed (km/h)
- \( \bar{v}_f \) = final (adjusted) average run speed (km/h)
- \( C_{bus} \) = vehicle capacity (passengers)

**KPI vi (Maximum Load Section Load Factor Increase/Decrease)** was defined as the estimated net change to average passenger load factor resulting from SR or SA changes to a trip pattern, assuming constant before-and-after fleet size. This KPI was calculated as follows:

\[ \hat{\alpha}_f = (\alpha_i)\left(\frac{\hat{H}_f}{H_i}\right) \]  
\[ \Delta \hat{\alpha} = \hat{\alpha}_f - \alpha_i \]  

Where:
- \( \alpha_i \) = initial Maximum Load Section passenger loading (pass.)
- \( \hat{\alpha}_f \) = estimated final Maximum Load Section passenger loading (pass.)
- \( H_i \) = initial scheduled headway (s)
- \( \hat{H}_f \) = estimated final headway (s)
KPI vii (Total Passenger In-vehicle Time Saved/Added) was defined as the estimated total net time saved/added for on-board passengers as a result of SR or SA changes to a trip pattern. The KPI was also broken down by stop. This KPI was calculated as follows:

\[
\bar{T}_{Pj} = \bar{P}_j(\Delta \bar{T}_{busj}) \tag{14}
\]

\[
\bar{T}_P = \sum_j^k \bar{T}_{Pj} \tag{15}
\]

For \( j = 1, 2, \ldots, k \)

Where:
- \( j \) = sequential stop number along a trip pattern (unitless)
- \( k \) = total number of stops in a trip pattern (unitless)
- \( \bar{T}_{Pj} \) = average cumulative in vehicle trip time savings or penalty for all onboard passengers at Stop \( j \) (s)
- \( \bar{T}_P \) = average cumulative trip time savings or penalty for all onboard passengers along a trip pattern (s)
- \( \bar{P}_j \) = average number of onboard passengers at Stop \( j \) (pass.)
- \( \Delta \bar{T}_{busj} \) = bus time savings or penalty resulting from SC or SA of Stop \( j \) (s)

KPI viii (Total Passenger Access/Egress Time Saved/Added) was defined as the estimated total net access/egress time saved/added for passengers as a result of SR or SA changes to a trip pattern. The KPI was also broken down by stop. This KPI was calculated as follows:

\[
\bar{t}_{inc} = \frac{\bar{x}_{inc}}{w} \tag{16}
\]

Where:
- \( \bar{t}_{inc} \) = incremental average walk time (s)
- \( \bar{x}_{inc} \) = weighted average incremental walk distance (m)
- \( w \) = walk speed (m/s)
KPI ix (Total Passenger Wait Time Saved/Added) was defined as the estimated total net wait time saved/added for passengers, based on changes to vehicle headways as a result of SR or SA changes to a trip pattern. Calculation of the KPI assumes buses run frequently (< 6 minute headways), causing passengers to arrive randomly (rather than timing their arrival to the next bus arrival), and that there is no variation in bus headways. Passengers are therefore expected to wait for an average of half the length of one headway before arrival of the next bus. The total wait time saved/added could also be broken down by stop. This KPI was calculated as follows:

\[
\hat{T}_{W_j} = \left(\frac{H_f - H_i}{2}\right)B_j
\]  

[17]

Where:

\( j = \) sequential stop number along a trip pattern (unitless)

\( \hat{T}_W = \) incremental average passenger wait time (s)

\( H_i = \) initial bus headway (s)

\( H_f = \) estimated final bus headway (s)

\( B_j = \) average boarding count at stop \( j \) (passengers)

In the case of the LTC, since headways are greater than 6 minutes, passengers are assumed to time their arrival, in which case changes to headway are unlikely to affect passenger wait times.

KPI x (Number of Existing Passengers Expected to Prefer Limited-stop Operation) was defined as the estimated average number of passengers benefiting from introduction of a limited-stop service based on the particular stop layout the user has selected. The value was also broken down by stop. This KPI was calculated as follows:

\[
L = \frac{N_L}{N_L + N_{local}}
\]  

[18]

Where:

\( L = \) percentage of passengers preferring limited stop operation (pass.)

\( N_L = \) number of passengers preferring limited stop operation (passengers)

\( N_{local} = \) number of passengers preferring local operation (passengers)
4.3 Assumptions

This section defines a number of key assumptions upon which the models described in Chapters 4.4 through 4.12 rely.

**Assumption 1:** The existing first and last stop along each trip pattern were assumed to remain as the terminus stops. In other words, the first and last stop along patterns could not be deleted, nor could new stops be added outside the original footprint of the pattern.

**Assumption 2:** It was assumed that a maximum of one stop would be added between any existing pair of stops. However, this assumption could be circumvented by running the software iteratively. No such limitation was included for SR and limited-stop operation.

**Assumption 3:** It was assumed that no dwell time savings or penalties would be realized at terminals, even if an adjacent stop was added or consolidated. Any addition to, or reduction in, dwell time required to load or unload passengers at terminals is assumed to be incorporated into the vehicle’s terminal time between trips, and is therefore considered negligible with no bearing on time savings or penalties for passengers or for the vehicle itself.

**Assumption 4:** Passengers who relocate due to SR or SA do not benefit from in-vehicle time savings / penalties at their chosen access and egress stops.

**Assumption 5:** It was assumed that passengers would walk to the nearest stop, though in reality some passengers would choose to walk a longer distance in order to access a downstream stop.

**Assumption 6:** Uniform passenger demand distribution (stemming from an assumption of uniform population density) was assumed within each stop’s catchment area, though the same need not be true between catchment areas along a trip pattern.
Assumption 7: Because SR and SA only affect access distance parallel to the trip pattern, changes to a stop’s catchment distance will be discussed as a proxy for changes to catchment area. The catchment distance for a stop was assumed to extend halfway to immediately adjacent stops in each direction. The catchment distance for the first (and last) stop along a trip pattern was assumed to be twice the length of its catchment distance that falls within the pattern’s limits, as shown below:

![Diagram of catchment distances](image)

**Figure 4: Catchment distance for the first stop along a trip pattern**

Assumption 8: It was assumed that no change in ridership would be triggered by changes to the trip pattern stop layout, given the relatively minor changes being made to the route and network as a whole. Based on existing literature (El-Geneidy, Strathman, Kimpel, & Crout, 2006; de Vries Kehoe, 2004), this assumption seemed safe. For the sake of this thesis’s case study, this assumption seemed doubly safe given that most stops in London have relatively low average passenger ons and offs. For example, at a stop with an average boarding count of two passengers per trip, the expected change in ridership would have to exceed +/- 25 percent (or +/- 0.5 passengers) before this would result in a measurable change to dwell time and access time estimates, given that passenger ons and offs were rounded to the nearest whole number. Due to the LTC’s low mode share compared to other forms of transportation in London (approximately 7 percent in 2011 according to Statistics Canada (2011)), and because the ridership is thought to consist largely of “captive” riders, significant changes to ridership as a result of changes to trip patterns were considered unlikely.
Assumption 9: It was assumed that bus capacities to accommodate demand at any given stop is unchanged from existing conditions as a result of changes to the trip pattern stop layout. For example, in the case of the LTC, bus load factors tend to be low and buses are able to serve all demand present at a given stop; this condition is assumed to continue regardless of stop layout changes. The models also assume there are no changes to existing congestion or other external factors affecting bus running times.

Assumption 10: It was assumed that the road network functioned as a perfect grid for the sake of calculating changes to access/egress distance to/from stops.

Assumption 11: Passenger access/egress walk speed was assumed to be 4.8km/hour, based on walk speeds described by the TCQSM, 3rd edition (2013) for topographically flat cities (gradients < 5%). Moreover, it was assumed that the neighbourhoods being analyzed were conducive to walking to transit, rather than accessing transit by other modes. For the sake of the LTC case study, London was assumed to be topographically flat.

Assumption 12: Due to limitations imposed by the dataset analyzed in this study, buses were assumed to stop at every stop. Refer to Chapter 3.3 for further discussion.

Assumption 13: As this software was built with the intention of analyzing the LTC’s archived dataset, and because headways in the LTC’s network tend to be relatively high, it was assumed that passengers would time their arrival based on expected bus arrival time. As such, no passenger wait time savings were expected as a result of moderate headway changes caused by stop layout changes.

4.4 Average Boarding and Alighting Approximations

Average ons and offs (boardings and alightings) at each stop along a trip pattern were required inputs for most models used in this study. As discussed briefly in Chapter 3.3, given that dwell times calculated from recorded bus arrival and departure times were considered unreliable, average ons and offs were used to drive the dwell time estimator provided with the TCQSM, 3rd edition (2013) (refer to Chapter 4.5). Average ons and offs were also used to estimate passenger redistribution (Chapter 4.6), access/egress penalties (Chapter 4.9), limited-stop operation benefit
including passenger-count origin-destination matrices (Chapter 4.11), and SR benefit-cost ratios (Chapter 4.12).

Average ons (and offs) at a given stop during any user-specified time period were calculated as follows:

\[ \bar{B}_j = \frac{\sum b_i}{n} \]  

\[ \text{For } j = 1, 2, \ldots, k \]

Where:
\( j \) = sequential stop number along the trip pattern (unitless)
\( k \) = total number of stops in the trip pattern (unitless)
\( \bar{B}_j \) = average boarding count at Stop \( j \) (passengers)
\( b_i \) = boarding count at Stop \( j \) during Trip \( i \) (passengers)
\( n \) = number of trips with recorded passenger counts at Stop \( j \) (unitless)

The results of equation [19] above were rounded up to the nearest whole number, to avoid fractional estimates of passenger activities.

Summing the average boarding counts for an entire trip pattern yielded total average ons for a trip:

\[ T_B = \sum_j^{k} \bar{B}_j \]  

\[ \text{Where:} \]
\( T_B \) = Total average boardings for a trip pattern (passengers)

However, the total average ons rarely equaled the total average offs after the rounding process. (To a degree, discrepancies were also likely a result of imperfect measurement of passenger ons and offs by the APC system). Because other models used in this study rely on the departure load of the bus at each stop, it was important to balance total ons with total offs. To address this, and to partially compensate for the slight overestimation of passenger ons and offs caused by...
rounding, the lesser of the two sums was adopted as the “correct” value. The greater of the two sums thus needed to be adjusted down, as follows (for the sake of this example, assume the total boarding count is greater than the total alighting count):

\[
D_{AB} = |T_A - T_B| \quad \text{[21]}
\]

\[
\bar{B}_j^* = \left( \frac{\bar{B}_j}{T_B} \right) (D_{AB}) \quad \text{[22]}
\]

For \(j, j^* = 1, 2, ..., k\)

Where:

\(D_{AB} = \text{Absolute diff. between total avg alightings and boardings (pass.)}\)

\(\bar{B}_j^* = \text{relative chance that loss of one boarding occurs at stop } j \text{ (unitless)}\)

Values of \(\bar{B}_j^*\) were then sorted from greatest to smallest, and a value of one (1) was subtracted from the first (largest) \(\bar{B}_j^*\) and its associated \(\bar{B}_j\). Values of \(\bar{B}_j^*\) were then resorted from greatest to smallest, and the procedure was repeated. This procedure was carried out once for every whole number difference between the values of \(T_A\) and \(T_B\). The following example illustrates this process:

Table 1: Hypothetical Initial Average Boarding and Alighting Counts Along a Trip Pattern

<table>
<thead>
<tr>
<th>Average Boarding Count at Stops a-e</th>
<th>Average Alighting Count at Stops a-e</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{B}_a = 12)</td>
<td>(\bar{A}_a = 0)</td>
</tr>
<tr>
<td>(\bar{B}_b = 2)</td>
<td>(\bar{A}_b = 3)</td>
</tr>
<tr>
<td>(\bar{B}_c = 4)</td>
<td>(\bar{A}_c = 3)</td>
</tr>
<tr>
<td>(\bar{B}_d = 3)</td>
<td>(\bar{A}_d = 5)</td>
</tr>
<tr>
<td>(\bar{B}_e = 0)</td>
<td>(\bar{A}_e = 7)</td>
</tr>
<tr>
<td>(\sum \bar{B} = T_B = 21)</td>
<td>(\sum \bar{A} = T_A = 18)</td>
</tr>
</tbody>
</table>
Using equation [21] above,

\[ D_{AB} = |T_A - T_B| = |18 - 21| = 3 \]

Thus, the total number of average boardings must be reduced by 3 to 18, to match the number of alightings. As the total number of average boardings is to be reduced by a value of 3, this adjustment will be completed in three stages. Plugging equation [21] into [22], the initial condition (stage 0) is shown:

<table>
<thead>
<tr>
<th>( B_j (0) )</th>
<th>( B_j^* (0) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_a = 12 )</td>
<td>( B_a^* = \frac{B_j}{T_B} (D_{AB}) = \frac{12}{21} (3) = 1.714 )</td>
</tr>
<tr>
<td>( B_b = 2 )</td>
<td>( B_b^* = 0.286 )</td>
</tr>
<tr>
<td>( B_c = 4 )</td>
<td>( B_c^* = 0.571 )</td>
</tr>
<tr>
<td>( B_d = 3 )</td>
<td>( B_d^* = 0.429 )</td>
</tr>
<tr>
<td>( B_e = 0 )</td>
<td>( B_e^* = 0 )</td>
</tr>
</tbody>
</table>

The values from Table 2 are sorted from largest to smallest value of \( B_j^* \), and Stages 1-3 are shown:

<table>
<thead>
<tr>
<th>( B_j (0) )</th>
<th>( B_j^* (0) )</th>
<th>( B_j (1) )</th>
<th>( B_j^* (1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_a = 12 )</td>
<td>( B_a^* = 1.714 )</td>
<td>( B_a = 12 - 1 = 11 )</td>
<td>( B_a^* = 1.714 - 1 = 0.714 )</td>
</tr>
<tr>
<td>( B_c = 4 )</td>
<td>( B_c^* = 0.571 )</td>
<td>( B_c = 4 )</td>
<td>( B_c^* = 0.571 )</td>
</tr>
<tr>
<td>( B_d = 3 )</td>
<td>( B_d^* = 0.429 )</td>
<td>( B_d = 3 )</td>
<td>( B_d^* = 0.429 )</td>
</tr>
<tr>
<td>( B_b = 2 )</td>
<td>( B_b^* = 0.286 )</td>
<td>( B_b = 2 )</td>
<td>( B_b^* = 0.286 )</td>
</tr>
<tr>
<td>( B_e = 0 )</td>
<td>( B_e^* = 0 )</td>
<td>( B_e = 0 )</td>
<td>( B_e^* = 0 )</td>
</tr>
</tbody>
</table>
The final boarding and alighting counts are now established:

<table>
<thead>
<tr>
<th>$\bar{B}_j$ (2)</th>
<th>$\bar{B}_j^*$ (2)</th>
<th>$\bar{B}_j$ (3)</th>
<th>$\bar{B}_j^*$ (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{B}_a$ = 11 - 1 = 10</td>
<td>$\bar{B}_a^*$ = 0.714 - 1 = -0.286</td>
<td>$\bar{B}_c$ = 4 - 1 = 3</td>
<td>$\bar{B}_c^*$ = 0.571 - 1 = -0.429</td>
</tr>
<tr>
<td>$\bar{B}_c$ = 4</td>
<td>$\bar{B}_c^*$ = 0.571</td>
<td>$\bar{B}_d$ = 3</td>
<td>$\bar{B}_d^*$ = 0.429</td>
</tr>
<tr>
<td>$\bar{B}_d$ = 3</td>
<td>$\bar{B}_d^*$ = 0.429</td>
<td>$\bar{B}_b$ = 2</td>
<td>$\bar{B}_b^*$ = 0.286</td>
</tr>
<tr>
<td>$\bar{B}_b$ = 2</td>
<td>$\bar{B}_b^*$ = 0.286</td>
<td>$\bar{B}_e$ = 0</td>
<td>$\bar{B}_e^*$ = 0</td>
</tr>
<tr>
<td>$\bar{B}_e$ = 0</td>
<td>$\bar{B}_e^*$ = 0</td>
<td>$\bar{B}_a$ = 10</td>
<td>$\bar{B}_a^*$ = -0.286</td>
</tr>
</tbody>
</table>

Table 4: Hypothetical Adjusted Average Boarding and Alighting Counts Along a Trip Pattern

<table>
<thead>
<tr>
<th>Average Boarding Count at Stops a-e</th>
<th>Average Alighting Count at Stops a-e</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{B}_a$ = 10</td>
<td>$\bar{A}_a$ = 0</td>
</tr>
<tr>
<td>$\bar{B}_b$ = 2</td>
<td>$\bar{A}_b$ = 3</td>
</tr>
<tr>
<td>$\bar{B}_c$ = 3</td>
<td>$\bar{A}_c$ = 3</td>
</tr>
<tr>
<td>$\bar{B}_d$ = 3</td>
<td>$\bar{A}_d$ = 5</td>
</tr>
<tr>
<td>$\bar{B}_e$ = 0</td>
<td>$\bar{A}_e$ = 7</td>
</tr>
</tbody>
</table>

The final product from this model was an average number of passenger ons and offs for each stop along a trip pattern, the sums of which were balanced.

4.5 Dwell Time Estimation

A variety of regression models estimating dwell times based on passenger counts were developed by Rajbhandari, Chien, & Daniel (2003), who determined that, in the context of buses operated by New Jersey Transit Corporation in 2001-2002, roughly 26 percent of a bus’s total travel time is spent dwelling. The study found that passenger demand was the primary determinant of dwell time. Number of doors, number of on-board standees, boarding height (standard vs. low-floor), fare collection methods, all-door vs. front-only boarding, operator habits, number of mobility-impaired passengers, and other factors are also thought to play a role.
The TCQSM, 3rd edition (2013) provides a dwell time estimator which takes a number of these conditions into account. Specifically, the tool allows the user to enter average boarding and alighting counts for each stop; front vs. all-door boarding; fare payment method; boarding height; standees present (yes/no); number of doors per vehicle; number of door channels; percent of boarders using the farebox; door opening and closing time; and the number of loading areas to be served at each stop. Finally, the tool allows the user to customize time spent boarding for boarding and alighting passengers at each channel on a bus.

Because this thesis was intended to provide a high-level, widely applicable tool, the TCQSM’s default settings for passenger boarding and alighting times were adopted. Standard values for the number of doors, channels, percent of boarders using the farebox, door opening and closing time, number of loading areas, and presence of standees (assumed no) were assumed constant for the entire LTC network. (No standees were assumed due to the low load factors experienced across the entire LTC network; agencies in larger cities with higher load factors would not be able to make this assumption). Of course, use of these standardized values was a simplification, but building flexibility to control these minor details in an otherwise high-level modelling tool risked conveying a false sense of precision. At the time of this study, no buses in the LTC’s network accepted fare cards. Two sets of potential dwell times were produced: one set for low-floor buses, and one for buses with stairs. The following standard values were input into the TCQSM dwell time estimator to predict bus dwell times:
Table 5: Dwell Time Estimator Parameters

<table>
<thead>
<tr>
<th>Input</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average boarding volume per bus</td>
<td>[specific to stop of interest and selected filters]</td>
</tr>
<tr>
<td>Average alighting volume per bus</td>
<td>[specific to stop of interest and selected filters]</td>
</tr>
<tr>
<td>Boarding door(s)</td>
<td>Front</td>
</tr>
<tr>
<td>Fare payment method</td>
<td>Single ticket/token</td>
</tr>
<tr>
<td>Boarding height</td>
<td>Case 1: Level; Case 2: Stairs</td>
</tr>
<tr>
<td>Standees present?</td>
<td>No</td>
</tr>
<tr>
<td>Number of doors</td>
<td>2</td>
</tr>
<tr>
<td>Available door channels</td>
<td>2 per door; 4 total</td>
</tr>
<tr>
<td>Percent of boarders using farebox</td>
<td>80% (20% assumed to have monthly passes)</td>
</tr>
<tr>
<td>Door opening and closing time</td>
<td>6 seconds (total)</td>
</tr>
<tr>
<td>Number of loading areas</td>
<td>1</td>
</tr>
</tbody>
</table>

Ultimately, two sets of dwell times were calculated for the user’s trip pattern of interest: initial dwell times estimated for the existing trip pattern, and final dwell times estimated for the adjusted pattern. The difference between the two provided the basis for estimating changes in bus and passenger travel times. It should be noted that dwell times were expected to change not only at added or deleted stops, but also at adjacent stops who gain or lose ridership to the modified stops.

\[
\Delta \bar{T}_{bus} = \hat{T}_{busf} - \bar{T}_{busi}
\]

**Where:**

\( \bar{T}_{busi} = \text{initial (existing) average bus run time (s)} \)

\( \hat{T}_{busf} = \text{estimated final (adjusted) average bus run time (s)} \)

\( \Delta \hat{T}_{bus} = \text{time savings or penalty resulting from changes to trip pattern (s)} \)
4.6 Passenger Redistribution Estimation

Changes to anticipated dwell times represented one of two fundamental factors affecting bus and passenger time savings/penalties as a result of SR and SA. Passenger redistribution provided the basis for estimating changes to anticipated dwell times. Given the assumption of constant total ridership, and uniform passenger demand within each stop’s catchment distance, the model used to predict passenger redistribution consisted of a linear probabilistic rebalancing effort. This process is described in Chapters 4.6.1 and 4.6.2.

The other fundamental factor affecting run times as a result of SR and SA was time savings/penalties due to acceleration and deceleration. Refer to Chapter 4.8 for discussion therein.

4.6.1 Stop Addition

Addition of a new stop along a trip pattern was assumed to cannibalize existing ridership from immediately adjacent stops. As total average ridership was assumed unchanged as a result of trip pattern modifications, no “new” ons were calculated; passengers were merely redistributed. Redistribution was probabilistic, based on linear distances between new and existing stops. Distances are described as shown by sample distances in Figure 5.

![Figure 5: Distances between new (Ω) and existing (a, b, c) stops along a trip pattern](image-url)
From Figure 5, the proportion of passengers boarding at Stop b who may be interested in boarding at Stop Ω (i.e., the passengers boarding Stop b from downstream) can be calculated as follows:

\[ b_{b \text{downstream}} = (b_b) \frac{d_{bc}}{2} \frac{d_{ac}}{2} \]

Of Stop b’s initial downstream boarders, some will find it advantageous to relocate to Stop Ω once Stop Ω has been added.

\[ b_{\Omega b} = (b_b) \frac{d_{bc} - d_{b\Omega}}{2} \frac{d_{ac}}{2} = (b_b) \frac{(d_{bc} - d_{b\Omega})}{d_{ac}} \]

Thus,

\[ b_{\Omega b} = \frac{(b_b)(d_{ac})}{d_{ac}} \quad [23] \]

\[ b_{\Omega c} = \frac{(b_c)(d_{b\Omega})}{d_{bd}} \quad [24] \]

\[ \bar{B}_{\Omega} = b_{\Omega b} + b_{\Omega c} \quad [25] \]

Where:

- \( a, b, c, d \) = existing stops along a trip pattern
- \( \Omega \) = new (added) stop
- \( b_n \) = initial boarding count at any existing Stop n (passengers)
- \( b_{\Omega n} \) = boardings redistributed to new Stop Ω from any existing Stop n (pass) 
- \( d_{ni} \) = distance between any existing Point n and any existing Point i (m)
- \( \bar{B}_{\Omega} \) = total final boarding count at Stop Ω (passengers)

Final boarding counts at existing stops were adjusted to compensate for any passengers lost to the new stop. All figures were rounded to the nearest whole number. The final result was a linear probabilistic redistribution of existing ridership along the trip pattern.
4.6.2 Stop Removal & Limited-Stop Operation

The logic used to calculate ridership cannibalization during the process of SA was applied in reverse for SR. Any passengers directly affected by SR were redistributed to immediately adjacent stops.

For clarity, only the catchment distance for Stop b (which is being consolidated in the example below) is shown in Figure 6.

![Diagram of stops a, b, and c with catchment distances]

Legend
- Catchment distance
- Trip pattern extent

Figure 6: Stop b catchment distance, and catchment distance within midpoint (θ) from Stop a to Stop c

\[
R_{ba} = \frac{d_{a0} - \frac{d_{ab}}{2}}{\frac{d_{ab} + d_{bc}}{2}}
\]

\[= 2\left(\frac{d_{a0} - \frac{d_{ab}}{2}}{d_{ab} + d_{bc}}\right)\]  \[\text{[26]}\]

\[R_{bc} = 1 - R_{ba}\]  \[\text{[27]}\]

\[\bar{B}_{a'} = \bar{B}_a + (R_{ba})(\bar{B}_b)\]  \[\text{[28]}\]

\[\bar{B}_{c'} = \bar{B}_c + (R_{bc})(\bar{B}_b)\]  \[\text{[29]}\]

Where:
\[a, c = existing \text{ stops remaining in place}\]
\[ b = \text{existing stop deleted by user} \]
\[ \theta = \text{midpoint between Stops } a \text{ and } c \]
\[ R_{ba} = \% \text{ of passengers from Stop } b \text{ redistributed to Stop } a \text{ (unitless)} \]
\[ R_{bc} = \% \text{ of passengers from Stop } b \text{ redistributed to Stop } c \text{ (unitless)} \]
\[ d_{ni} = \text{distance between any existing Point } n \text{ and any existing Point } i \text{ (m)} \]
\[ \bar{B}_a = \text{total final boarding count at Stop } a \text{ (passengers)} \]
\[ \bar{B}_c = \text{total final boarding count at Stop } c \text{ (passengers)} \]

Final boarding counts at existing stops were adjusted to compensate for any passengers gained from any immediately adjacent deleted stops. All figures were rounded to the nearest whole number. The final result was a linear probabilistic redistribution of existing ridership along the trip pattern. If the assumption that all buses dwell at all stops were to be removed (refer to Assumption 12 in Chapter 0), this model could be adjusted to account for the likelihood of dwell at individual stops. Refer to Chapter 4.7.2 for further discussion.

4.7 Chance of Dwell Estimation

Ideally, to predict the average time savings/penalties experienced by buses and passengers as a result of adding or removing stops from a trip pattern, models would be used to calculate anticipated time savings/penalties given that a bus stops, as well as the likelihood of stopping at each stop on the route. However, one of the limitations to this research posed by the available data was the inability to estimate the likelihood that a bus will dwell at a given stop (this limitation is described in detail in Chapter 3.3). As such, and given that headways in the LTC’s network tend to be long, an assumption was made that buses stop at every stop along a trip pattern.

Prior to the discovery of this limitation of the dataset, however, a model was produced to estimate the likelihood of a random bus dwelling for each stop along a pattern. This model is presented here, though it was ultimately rendered meaningless for the sake of the LTC case study due to the limitation described above.
4.7.1 Stop Addition

An initial estimate of chance of dwell was calculated for existing stops based on the existing archived data for each stop along the trip pattern, as follows:

\[ C_j = \frac{s_j}{n_j} \quad [30] \]

For \( j = 1, 2, \ldots, k \)

Where:
\( j = \text{sequential stop number along a trip pattern (unitless)} \)
\( k = \text{total number of stops in a trip pattern (unitless)} \)
\( C_j = \text{chance of dwelling at Stop } j \text{ (unitless)} \)
\( s_j = \text{total number of dwells at Stop } j \text{ (unitless)} \)
\( n_j = \text{total number of trips for which Stop } j \text{ was in the pattern (unitless)} \)

The initial chance of dwell calculated for existing stops was assumed to be unaffected by the addition of new stops. While it is likely that the model below slightly overestimates a stop’s chance of dwell, it is an improvement on the assumption that the chance of dwell is 100 percent for all stops.

Chance of dwell at new (added) stops was calculated as a weighted average of the likelihood of dwelling at existing adjacent stops. Chance of dwell was calculated based on the total number of average boardings plus alightings estimated at the new stop. Refer to Figure 5 (Chapter 4.6.1) for an explanation of the stop layout of a hypothetical trip pattern undergoing stop addition.

\[ C_\Omega = \frac{C_b T_{\Omega b} + C_c T_{\Omega c}}{T_{\Omega b} + T_{\Omega c}} \quad [31] \]

Where:
\( C_\Omega = \text{chance of dwelling due at new Stop } \Omega \text{ (unitless)} \)
\( C_b, C_c = \text{chance of dwelling at Stops } b \text{ and } c \text{ (unitless)} \)
\( T_{\Omega n} = \text{passengers redistributed to new Stop } \Omega \text{ from any existing Stop } n \text{ (pass)} \)
4.7.2 Stop Removal & Limited-stop Operation

Initial chance of dwell was calculated for existing stops using the methodology described in Chapter 4.7.1. However, in the case of SR and limited-stop operation, the chance of dwell at existing stops was adjusted when adjacent stops were removed. Consider three consecutive existing Stops, a, b, and c, with existing (i.e., initial) chance of dwell $C_{a_i}$, $C_{b_i}$, $C_{c_i}$:

If no stops are deleted, let

$$C_{bf} = 1 - (1 - C_{bi})$$  \[32\]

If Stop a is deleted, and a nonzero number of passengers are redistributed to Stop b according to the model described in Chapter 4.6.2, let

$$C_{bf} = 1 - (1 - C_{bi})(1 - C_{ai})$$  \[33\]

If Stops a and c are deleted, and passengers are redistributed to Stop b from both a and c, let

$$C_{bf} = 1 - (1 - C_{bi})(1 - C_{ai})(1 - C_{ci})$$  \[34\]

Where:

- $C_{ai}$, $C_{bi}$, $C_{ci}$ = initial chance of dwell at existing Stops a, b, and c (unitless)
- $C_{bf}$ = final chance of dwell at remaining Stop b (unitless)

In a situation where Stop c is deleted, a bus dwelling at Stop b could choose to dwell for any one of the following reasons:

- To serve passengers initially accessing Stop b;
- To serve passengers initially accessing Stop c;
- To serve passengers initially accessing Stops b and c.

Similarly, if Stops a and c are deleted, a bus could dwell at Stop b to serve any one of seven different combinations of passengers (a; b; c; ab; bc; ac; abc).
Once the final chance of dwell has been calculated for remaining (i.e., non-deleted) stops, the number of passengers redistributed from deleted stops can be weighted according to initial chance of dwell at the deleted stops in question. For example, if examining three consecutive existing Stops, a, b, and c:

- Let \( C_{a_i} = 0.80 \),
- \( C_{b_i} = 0.70 \), and
- \( C_{c_i} = 0.60 \);
- Assume stop c is to be deleted, and Stops a and b are to remain;
- From [33],

\[
C_{bf} = 1 - (1 - C_{b_i})(1 - C_{c_i}) \\
= 1 - (1 - 0.70)(1 - 0.60) \\
= 0.88
\]

The calculated value of \( C_{bf} \), 0.88, represents the combined chance that the bus will dwell at Stop b for all possible combinations of passengers; however, this chance can be broken down into its probabilistic components, as follows:

- Chance of needing to serve only passengers initially accessing Stop b:

\[
C_{bb} = (C_{b_i})(1 - C_{c_i}) \\
= 0.28
\]

- Chance of needing to serve only passengers initially accessing Stop c:

\[
C_{bc} = (1 - C_{b_i})(C_{c_i}) \\
= 0.18
\]

- Chance of needing to serve passengers initially accessing Stops b and c:

\[
C_{b_{bc}} = (C_{b_i})(C_{c_i}) \\
= 0.42
\]
- Summing the three probabilities calculated above, the combined final chance of dwelling at Stop b is calculated:

\[ C_{bf} = C_{bb} + C_{bc} + C_{bbc} = 0.28 + 0.18 + 0.42 \]

\[ = 0.88 \]

*Where:*

- \( C_{bb} \) = chance of dwelling to serve only passengers initially accessing Stop b (unitless)
- \( C_{bc} \) = chance of dwelling to serve only passengers initially accessing Stop c (unitless)
- \( C_{bbc} \) = chance of dwelling to serve passengers initially accessing Stops b and c (utls)
- \( C_{bf} \) = final chance of dwell at remaining Stop b (unitless)
Building on [23], average passenger counts weighted by chance of dwell can be established:

\[
\bar{A}_{bf} = (C_{bi})a_{bi} + (C_{ci})a_{bc_i} \tag{35a}
\]

\[
\bar{B}_{bf} = (C_{bi})b_{bi} + (C_{ci})b_{bc_i} \tag{35b}
\]

Where:

\(\bar{A}_{bf}\) = final average alightings at Stop b, weighted by chance of dwell (pass.)

\(\bar{B}_{bf}\) = final average boardings at Stop b, weighted by chance of dwell (pass.)

\(C_{bi}, C_{ci}\) = initial chance of dwelling at Stops b and c (unitless)

\(a_{bi}, b_{bi}\) = initial alightings and boardings at existing Stop b (passengers)

\(a_{bc_i}, b_{bc_i}\) = alightings/boardings redistributed from c to b using [22] (pass.)

Next, using [33a] and [33b] to solve the example above,

- Assume two alightings and five boardings initially choose to access Stop b, and
- Assume two alightings and four boardings are redistributed from Stop c to Stop b using the model described in Chapter 4.6.2.

\[\bar{A}_{bf} = (0.70)(2) + (0.60)(2)\]

\[= 2.6 \rightarrow \text{round to 3.0}\]

\[\bar{B}_{bf} = (0.70)(5) + (0.60)(4)\]

\[= 5.9 \rightarrow \text{round to 6.0}\]

Thus, the final expected average boarding and alighting count at Stop b would be three alightings and six boardings, with a chance of dwell of 88 percent. It is worth noting that the same result can be found by applying [33] broken down into its components shown in the example above. That is, given that,

\[C_{bf} = C_{bb} + C_{bc} + C_{b\text{bc}} = 0.28 + 0.18 + 0.42 = 0.88\]
\[ \bar{A}_{bf} = C_{bb}(a_{bi}) + C_{bc}(a_{bci}) + C_{bbc}(a_{bi} + a_{bci}) \]

\[ = (0.28)(2) + (0.18)(2) + (0.42)(2 + 2) \]

\[ = 2.6 \rightarrow \text{round to 3.0} \]

\[ \bar{B}_{bf} = C_{bb}(b_{bi}) + C_{bc}(b_{bci}) + C_{bbc}(b_{bi} + b_{bci}) \]

\[ = (0.28)(5) + (0.18)(4) + (0.42)(5 + 4) \]

\[ = 5.9 \rightarrow \text{round to 6.0} \]

Where:

- \( C_{bb} \) = chance of dwelling to serve only passengers initially accessing Stop b (unitless)
- \( C_{bc} \) = chance of dwelling to serve only passengers initially accessing Stop c (unitless)
- \( C_{bbc} \) = chance of dwelling to serve passengers initially accessing Stops b and c (utls)
- \( C_{bf} \) = final chance of dwell at remaining Stop b (unitless)
- \( \bar{A}_{bf} \) = final average alightings at Stop b, weighted by chance of dwell (pass.)
- \( \bar{B}_{bf} \) = final average boardings at Stop b, weighted by chance of dwell (pass.)
- \( C_{bi}, C_{ci} \) = initial chance of dwelling at Stops b and c (unitless)
- \( a_{bi}, b_{bi} \) = initial alightings and boardings at existing Stop b (passengers)
- \( a_{bci}, b_{bci} \) = alightings/boardings redistributed from Stop c to b using [24] (pass)

Again, the final expected average boarding and alighting count at Stop b would be three alightings and six boardings, with a chance of dwell of 88 percent.

The above example shows the methodology used when only Stop c is deleted. However, the same logic is applied to situations in which Stops a and c are both deleted.

4.8 Acceleration and Deceleration Time Penalty Estimation
As introduced in Chapter 4.6, there are two fundamental factors affecting bus and passenger time savings/penalties as a result of SR and SA. The first factor, passenger redistribution, is described in Chapter 4.6, and to a lesser degree in Chapter 4.7.2. Acceleration and deceleration time savings/penalties represented the second fundamental change to anticipated run times as a result of SR and SA. For the purposes of this study, acceleration and deceleration time savings/penalty estimates were fixed regardless of location, and relied upon the following assumptions:

- Assume that buses are able to reach maximum operating speed before deceleration at the next stop is required;
- Assume buses accelerate from a full stop to a maximum operating speed of 60 km/hr (16.67 m/s) at a constant rate of 1.67m/s$^2$ (i.e., buses require 10 seconds to accelerate from a full stop to a maximum speed of 60km/hr);
- Assume buses decelerate to a full stop from maximum speed (16.67 m/s) at a constant rate of 2.08m/s$^2$ (i.e., buses require 8 seconds to decelerate from a maximum speed of 60km/hr to a full stop); and,
- Assume acceleration and deceleration is constant (linear).
These assumptions yielded the following time savings/penalty estimates:

\[
\nu_{avg} = \frac{v_{max} - v_0}{2}
\]

\[
= \frac{v_{max}}{2}
\]  \[36\]

\[
t_{lost} = (1 - \frac{\nu_{avg}}{v_{max}})(t_{accel} + t_{decel})
\]  \[37\]

Subbing [36] into [37],

\[
t_{lost} = (1 - \frac{v_{max}}{2 \, v_{max}})(t_{accel} + t_{decel})
\]

\[
= (1 - 0.5)(10 + 8)
\]

\[
= 9 \text{ sec/stop}
\]

Where:

\(\nu_{avg}\) = average bus speed during acceleration and deceleration (m/s)

\(v_{max}\) = maximum bus speed (m/s)

\(t_{accel}\) = time spent accelerating (s)

\(t_{decel}\) = time spent decelerating (s)

\(t_{lost}\) = time lost due to acceleration and deceleration (s)

Notably, Furth & Rahbee (2000) also made an assumption of nine seconds lost per stop due to acceleration and deceleration, though this was calculated using different assumptions of acceleration and deceleration values and durations.

Acceleration and deceleration time savings/penalties generally exceeded the amount of time saved/lost due to passenger loading and unloading at low-use stops (i.e., the stops most likely to be slated for SR). This result was anticipated.
4.9 Access and Egress Penalty Estimation

As mentioned briefly in Chapter 2.3, adding stops along a route reduces the average distance that users must travel, by other modes, in order to access the transit service, saving access time. Of course, the opposite is true when stops are deleted. This section outlines the process for determining the distance and time penalties associated with SR. For brevity, discussion of access/egress distance and time savings caused by SA will be omitted, as the same principles apply whether a stop is added or deleted. Note that in the context of this thesis, the terms “access/egress penalty”, “distance penalty”, “time penalty”, etc. all refer to the incremental penalty experienced by passengers when comparing existing stop layouts with adjusted stop layouts, unless described otherwise.

Given the assumption stated in Chapter 4.3 that London’s neighbourhoods are conducive to walking to transit, this analysis assumed that all passengers access transit by walking.

A weighted average of incremental walk distances to the closest remaining stops was calculated for each deleted stop. Combined with an assumed average walk speed, this incremental walk distance was used to determine the penalty incurred by passengers for deleting a given stop. As described in Chapter 4.3, it was assumed that all passengers walk to the nearest remaining stop, though in reality some passengers would choose to walk a longer distance in order to access a downstream stop. The road network was assumed to function as a perfect grid for the sake of calculating changes to access/egress distance to/from stops. As discussed in Chapter 4.3, uniform passenger demand distribution was assumed within each stop’s catchment area.

The methodology used in this study was similar to one described by Wagner & Bertini (2014), who used a weighted average “to determine the average net additional walking distance to the nearest remaining stops”. The method used by Wagner & Bertini is as follows:

\[
D_w = \frac{(D_n^2 + D_f^2 + 4D_nD_f - 2D_nD_t)}{2D_t}
\]  

[38]
Where:

- \( D_w = \) average additional walk distance to remaining stops
- \( D_n = \) distance to near stop
- \( D_f = \) distance to far stop
- \( D_t = D_n + D_f \)

Putting sample values to [38], the following incremental walk distance is calculated:

- Let \( D_n = 100 \text{ m} \)
- Let \( D_f = 200 \text{ m} \)
- \( D_t = 300 \text{ m} \)

\[
D_w = \frac{100^2 + 200^2 + 4(100)(200) - 2(100)(300)}{2(300)}
\]

\[
= 116.67 \text{ m}
\]

One assumes that \( D_w \) is in fact the proposed average final walk distance, rather than the additional walk distance as the paper states (116.67m is far too large to be the additional walk distance, as shown below). While this thesis used a similar approach, a different value for this weighted average incremental walk distance was determined, as follows, where existing Stop b is deleted, Stops a and c remain, and the midpoint between Stops a and c, \( \theta \), is shown:

![Diagram](image)

**Figure 7:** Deleted stop b catchment distance, and midpoint (\( \theta \)) from Stop a to Stop c.
First, Stop b’s catchment distance is divided into three sections, as shown in Figure 7. Using the same sample values discussed above,

\[ \text{Let } d_{ab} = 100 \text{ m}, \quad d_{bc} = 200 \text{ m} \]

Prior to deletion of Stop b, average access distance for passengers accessing Stop b can be found as follows:

\[ l_i = \frac{d_{ab}}{2} = 50 \text{ m} \]

\[ l_{ii} = d_{b\theta} = \frac{d_{ab} + d_{bc}}{2} - d_{ab} = \frac{100+200}{2} - 100 = 50 \text{ m} \]

\[ l_{iii} = \frac{d_{bc}}{2} - d_{b\theta} = \frac{200}{2} - 50 = 50 \text{ m} \]

\[ x_{0i} = \frac{l_i}{2} = 25 \text{ m} \]

\[ x_{0ii} = \frac{l_{ii}}{2} = 25 \text{ m} \]

\[ x_{0iii} = \frac{l_{iii}}{2} + l_{ii} = 75 \text{ m} \]

Weighting the average initial walk distances for each segment by the length of each segment,

\[ X_i = \frac{(l_i)(x_{0i}) + (l_{ii})(x_{0ii}) + (l_{iii})(x_{0iii})}{l_i + l_{ii} + l_{iii}} \quad [39] \]

**Where:**

- \( l_i, l_{ii}, l_{iii} \) = length of section i, ii, and iii, respectively (m)
- \( x_{0i}, x_{0ii}, x_{0iii} \) = average initial walk distance for passengers within each of sections i, ii, and iii (m)
- \( X_i \) = weighted average initial walk distance for all passengers initially accessing Stop b (m)
In this example, the weighted average initial walk distance for all passengers initially accessing stop b is found to be 41.67 m:

\[ X_i = \frac{(50)(25) + (50)(25) + (50)(75)}{50 + 50 + 50} = 41.67 \, m \]

After Stop b is deleted, average access distance to the next closest stop for passengers previously accessing Stop b can be found:

\[ x_{fi} = \frac{l_i}{2} + \frac{d_{ab}}{2} = 75 \, m \]
\[ x_{fii} = d_{ab} + \frac{d_{b\theta}}{2} = 125 \, m \]
\[ x_{fiii} = d_{\theta c} - \frac{l_{iii}}{2} = 125 \, m \]

Weighting these average final walk distances for each segment by the length of each segment,

\[ X_f = \frac{(l_i)(x_{fi}) + (l_{ii})(x_{fii}) + (l_{iii})(x_{fiii})}{l_i + l_{ii} + l_{iii}} \]

[40]

Where:
\[ x_{fi}, x_{fii}, x_{fiii} = \text{average final walk distance for passengers within each of sections } i, ii, \text{ and iii (m)} \]
\[ X_f = \text{weighted average final walk distance for all passengers initially accessing Stop b (m)} \]

In this example, the weighted average final walk distance for all passengers initially accessing Stop b is found to be 108.33 m:

\[ X_f = \frac{(50)(75) + (50)(125) + (50)(125)}{50 + 50 + 50} = 108.33 \, m \]

The calculated weighted average final walk distance in this example, 108.33 m, is close to the “incremental” value estimated by Wagner & Bertini’s model (116.67 m). It should be noted that
using this thesis’s methodology, the weighted average incremental walk distance in this example was found to be 66.67 m, a far cry from 116.67 m:

\[ X_{inc} = X_f - X_i = 108.33 - 41.67 = 66.67 \text{ m} \]

With the incremental average walk distance known, the incremental average walk time (i.e., the access and egress penalty incurred due to SR) is found using an assumed walk speed. Passenger access/egress walk speed was assumed to be 4.8km/hour, based on walk speeds described by the TCQSM, 3rd edition (2013) (refer to Assumption 11 in Chapter 4.3 for more information).

From [16]: \( t_{inc} = \frac{X_{inc}}{w} \)

Where:

\[ t_{inc} = \text{incremental average walk time (s)} \]
\[ w = \text{walk speed} = 4.8 \text{ km/hr} = 1.33 \text{ m/s} \]

Following the above example,

\[ t_{inc} = \frac{66.67 \text{ m}}{4.8\left(\frac{1000}{3600}\right) \text{ m/s}} = 50 \text{ s} \]

As previously mentioned, the same principles can be applied to calculate access/egress distance and time savings caused by SA.

4.10 Cycle Time, Headway, Average Speed, Productive Capacity, Maximum Load Section Load Factor, and Passenger Wait Time Estimation

As discussed in Chapter 2.3, adding or deleting stops to a trip pattern is expected to increase or decrease total bus run time, respectively. Knock-on effects of changing the bus run time include changes to the cycle time, vehicle headway, average speed, productive capacity, vehicle loading factor (specifically at the maximum load section), and passenger wait times.
4.10.1 Cycle Time Increase/Decrease Estimation

Total initial and final bus run times, and the difference between the two, were calculated as follows:

\[
\overline{T_{bus}} = \sum_{j}^{k} t_{Dj} + t_{op} \]  \hspace{1cm} \text{[41]}

From [1]: \[ \Delta \overline{T_{bus}} = \overline{T_{busf}} - \overline{T_{busi}} \]

Where:
- \( \overline{T_{bus}} \) = average bus run time (s)
- \( t_{Dj} \) = average bus dwell time at Stop \( j \) (s)
- \( t_{op} \) = average total operating time while a bus is in motion (s)
- \( k \) = total number of stops in the trip pattern (unitless)
- \( \overline{T_{busi}} \) = initial (existing) average bus run time (s)
- \( \overline{T_{busf}} \) = estimated final (adjusted) average bus run time (s)
- \( \Delta \overline{T_{bus}} \) = time savings or penalty resulting from changes to trip pattern (s)

Cycle time, defined as the sum of a bus’s run (active) time and its terminal time – time during which a vehicle is scheduled to be idle at its terminal(s), or terminal stop(s) – represents the time required for a bus to complete one full cycle of its trip pattern.

\[
T_c = \overline{T_{bus}} + t_t \]  \hspace{1cm} \text{[42]}

It was assumed that changes to terminal time would be proportionate to any change to run time caused by SR or SA. The adjusted terminal and cycle times were calculated as follows:

\[
t_{tf} = \left( \frac{\overline{T_{busf}}}{\overline{T_{busi}}} \right) t_{ti} \]

From [4]: \[ T_{cf} = T_{ci} - t_{ti} + t_{tf} - \Delta \overline{T_{bus}} \]
From [5]: \[ \Delta T_c = T_{c_i} - T_{c_f} \]

Where:
\( T_{c_i} = \) initial (existing) cycle time (s)
\( T_{c_f} = \) final (adjusted) cycle time (s)
\( t_{t_i} = \) initial total terminal time (s)
\( t_{t_f} = \) estimated final total terminal time (s)

4.10.2 Headway Increase/Decrease Estimation

By changing cycle time, assuming the number of vehicles assigned to the trip pattern remains constant, headway can be adjusted accordingly without the transit agency occurring additional costs.

From [2]: \[ \bar{H}_f = \frac{T_{c_i} - \Delta \hat{t}_{bus}}{N_{bus}} \]

From [3]: \[ \Delta \bar{H} = \bar{H}_f - H_i \]

Where:
\( T_{c_i} = \) initial (existing) scheduled cycle time (s)
\( \Delta \hat{t}_{bus} = \) time savings or penalty resulting from changes to trip pattern (s)
\( N_{bus} = \) fleet size operating on pattern of interest (vehicles)
\( H_i = \) initial (existing) scheduled headway (s)
\( H_f = \) final (existing) headway (s)
\( \Delta \bar{H} = \) change in headway resulting from changes to trip pattern (s)

4.10.3 Average Speed Increase/Decrease Estimation

Change in average run speed was calculated as a function of the estimated change in run time, given a constant trip pattern length. This value is symbolic in that vehicles never consistently maintain this speed while in operation; rather, it provides the transit agency with a means of comparing the performance of vehicle travel on distinct trip patterns.
From [6]: \[ \bar{v}_i = \frac{D_{a\Omega}}{T_{bus_i}} \]

From [7]: \[ \hat{v}_f = \frac{D_{a\Omega}}{T_{bus_f}} \]

From [8]: \[ \Delta \hat{v} = \hat{v}_f - \bar{v}_i \]

Where:
\( \bar{v}_i = \text{initial (existing) average run speed (km/h)} \)
\( \hat{v}_f = \text{final (adjusted) average run speed (km/h)} \)
\( \Delta \hat{v} = \text{speed increase/decrease due to stop layout changes (km/h)} \)
\( D_{a\Omega} = \text{total trip distance from first stop (} \alpha \text{) to last stop (} \Omega \text{) (km)} \)
\( T_{bus_i} = \text{initial (existing) bus run time (h)} \)
\( T_{bus_f} = \text{final (adjusted) bus run time (h)} \)

4.10.4 Productive Capacity Increase/Decrease Estimation

Change in productive capacity is proportional to change in average bus run speed, and like run speed, is a symbolic value providing an objective means of comparing vehicle (trip pattern) performance across the network.

From [9]: \[ \bar{C}_{pi} = \bar{v}_i(C_{bus}) \]

From [10]: \[ \bar{C}_{pf} = \hat{v}_f(C_{bus}) \]

From [11]: \[ \Delta \bar{C}_p = \bar{C}_{pf} - \bar{C}_{pi} \]

Where:
\( \bar{C}_{pi} = \text{initial (existing) productive capacity (psgr.km/h)} \)
\( \bar{C}_{pf} = \text{final (adjusted) productive capacity (psgr.km/h)} \)
\( \Delta \bar{C}_p = \text{productive capacity change due to stop layout changes (psgr.km/h)} \)
\( \bar{v}_i = \text{initial (existing) average run speed (km/h)} \)
The vehicle loading factor, or the degree to which a vehicle is “full”, is an important measure of service quality for transit agencies. Given constant before-and-after passenger demand, the loading factor (passenger loading) along a trip pattern can be assumed to vary linearly and proportionately to fleet headway. A trip pattern’s maximum load section (MLS) load factor – the limiting factor in terms of providing sufficient line capacity – was assumed to vary accordingly.

\[
\alpha = \frac{P}{C} = \frac{P}{(\text{freq})(C_b)} = \frac{HP}{60C_{bus}} \quad [43]
\]

\[
\frac{\alpha_f}{\alpha_i} = \frac{\frac{H_f P_f}{60C_b}}{\frac{H_i P_i}{60C_b}} = \frac{H_f}{H_i} \quad [44]
\]

From [12]: \( \hat{\alpha}_f = (\alpha_i)\left(\frac{H_f}{H_i}\right) \)

From [13]: \( \Delta \hat{\alpha} = \hat{\alpha}_f - \alpha_i \)

Where:

- \( \alpha = \) MLS load factor (unitless)
- \( P = \) passenger loading volume (pass.)
- \( C = \) trip pattern passenger capacity (pass./h)
- \( C_{bus} = \) passenger capacity on an individual bus (pass./h)
- \( \text{freq} = \) frequency of buses on a given trip pattern (bus/h)
- \( \alpha_i = \) initial MLS load factor (unitless)
- \( \hat{\alpha}_f = \) estimated final MLS load factor (unitless)
- \( H_i = \) initial scheduled headway (s)
- \( H_f = \) estimated final headway (s)
4.10.6 Wait Time Increase/Decrease Estimation

It is widely understood that, assuming no variability in headways, frequent service (roughly every 10 minutes or less), and random passenger arrivals (i.e., arrivals not timed to bus schedules), average expected passenger wait times are equal to half the duration of one headway.

\[ T_w^j = \left( \frac{H_f - H_i}{2} \right) B_j \]

*Where:*
- \( j = \) sequential stop number along a trip pattern (unitless)
- \( \hat{T}_w = \) incremental average passenger wait time (s)
- \( H_i = \) initial bus headway (s)
- \( \hat{H}_f = \) estimated final bus headway (s)
- \( B_j = \) average boarding count at stop \( j \) (passengers)

However, in the case of the LTC, changes to average passenger wait time were expected to be negligible. The vast majority of routes in the LTC’s network have long headways (15+ minutes), indicating that passengers are likely to time their bus stop arrival based on the route’s schedule, in which case headway increases (or reductions) are unlikely to have a significant impact on passenger wait times.

4.11 Limited-Stop Operation Preference Estimation

A model was constructed to estimate the percentage of passengers benefitting from limited-stop operation over existing local operation. The goal of this model was to provide the user with an estimate of the percentage of vehicles which should be converted from local operation to limited-stop operation, based on the assumption that the transit agency would be more likely to create a limited-stop trip pattern running in parallel to, rather than in place of, an existing local operation trip pattern.

As with the models described above, this model can be applied to any trip pattern of interest with known average ons and offs at each stop. This model was constructed to provide results as part
of a SR procedure; any SA procedure would need to be completed in a separate phase of analysis.

Chapters 4.5 through 4.9 describe the methodologies used to calculate dwell times, time savings and penalties, and passenger redistributions for a given number of ons and offs. This analysis is completed at the stop level. While this information can be aggregated to calculate bus run time savings (or penalties), for example, it is not granular enough to determine run time savings (or penalties) at the individual passenger level. As limited-stop preference estimation was measured at the individual passenger level, disaggregate measures of the above calculations were required in order to determine the benefits that individual passengers would likely experience from SR along a trip pattern. As described above, the obvious means of disaggregating calculations was to do so at the origin-destination level, with each element in the matrices representing a unique combination of origin stop and destination stop. This is true for each matrix described in this section.

A series of matrix calculations were required to produce the limited-stop operation preference model. These matrices will be described in order, and include O-D estimation; time-saved estimation (due to reduced bus running time); access/egress penalty estimation; passenger redistribution estimation; and, finally, limited-stop benefit estimation. Strictly speaking, each of the matrices described above was an O-D matrix; the first matrix simply calculated the information typically contained in an O-D matrix, i.e., boarding and alighting counts. Thus, in the passenger-count matrix as well as all following matrices, rows represented origin stops, and columns represented destination stops.

Though a limited-stop operation would be expected to have a shorter cycle time than its local operation counterpart, and therefore greater frequency (all other things being equal), this analysis assumed that any headway reductions realized as a result of implementing a limited-stop operation in parallel to local operation would be shared evenly between the local and limited-stop operations; that is, the number of vehicles serving each type of operation were assumed to remain balanced such that headways are equal for local and limited-stop patterns.
4.11.1 Passenger-Count Origin-Destination (O-D) Matrix Estimation

An accurate, up-to-date passenger-count origin-destination matrix is an essential tool for many transportation planning objectives, and a useful tool for many others. In the case of transit network planning, local and regional passenger-count matrices can be used to determine where to add routes, and which transit mode is appropriate for a given route, for example. At the transit operations level, decision-makers can use transit-specific passenger-count matrices to make informed decisions about headway and schedule adjustments, and tweaks to routes or trip patterns. Use of an Automatic Fare Card (AFC) system is the easiest way for a transit agency to collect reliable, disaggregate passenger-count O-D information at low marginal cost.

To estimate the percentage of passengers who would prefer limited-stop operation over existing local operation for any trip pattern of interest, a passenger-count matrix at the trip pattern level was required. As the LTC did not have fare cards at the time of writing, it was necessary to estimate a passenger-count matrix using data gathered by the LTC’s AVL and APC systems. This thesis made use of a method proposed by Navick and Furth (1994) and adapted by Schwarcz (2004) in her thesis work analyzing bus service for heavy-use corridors. Schwarcz employs the passenger-count matrix estimation described below as a starting point to determine what frequency share of a heavy-use corridor’s bus fleet should be converted to limited-stop operation. The same concept has been applied in this thesis, though the strategy to determine frequency share has been simplified to ignore headway and reliability considerations (reliability values are specific to each route and/or trip pattern, and this level of detail was not readily available).

In essence, Schwarcz’s adapted method for estimating a passenger-count matrix uses one-directional downstream passenger travel distance as the impedance function for a gravity-based model. First, a one-directional distance matrix was created. An example distance matrix is shown below in Table 6.
Table 6: Example One-Directional Distances Between Stops (metres)

<table>
<thead>
<tr>
<th>Destination Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin Stop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>150</td>
<td>350</td>
<td>650</td>
<td>750</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>200</td>
<td>500</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>300</td>
<td>400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

In order to employ a gravity model, an impedance function is required. The following impedance (propensity) function was provided by Navick & Furth:

\[ p(d_{ij}) = d_{ij}^\alpha e^{-d_{ij}\beta} \]

However, Schwarcz points out that the decay function cannot be identified in the case of one-directional travel. Moreover, Schwarcz found that the best fit for the power function parameter, \( \alpha \), was 1.0. As such, the impedance function simplified to become:

\[ p(d_{ij}) = d_{ij} \]

Using this impedance function, and average boarding and alighting counts gathered for each stop along a trip pattern, iterations of a doubly-constrained gravity model were completed until a minimum error threshold was reached. Example boarding and alighting counts, and an example estimated passenger-count matrix, are shown below in Table 7 and Table 8, respectively.

Table 7: Example Average Boarding and Alighting Counts (passengers)

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total Avg. Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Total Avg. Alightings</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>
Table 8: Example Estimated Passenger-Count O-D Matrix

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total Avg. Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.98</td>
<td>2.37</td>
<td>0.92</td>
<td>1.72</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1.58</td>
<td>0.82</td>
<td>1.60</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1.26</td>
<td>2.74</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2.00</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

| Total Avg. Alightings | 0 | 1 | 4 | 3 | 8 | 16 |

There are a number of other methods for estimating passenger-count matrices. One common technique, employed by transit agencies such as New York City Transit, involves inferring destinations where only origins are known from AFC system data (typical for transit agencies who require users to “tap on”, but have no requirement to “tap off” when exiting the transit system). This technique examines regular travel patterns, and assumes that if a user boards a transit vehicle at stop A every morning, and boards a vehicle at stop G every afternoon, then stop G is the morning’s destination, and stop A is the afternoon’s destination. Obviously, this technique was not employed for this study, due to the lack of available AFC data.

4.11.2 In-Vehicle Time Savings Matrix Estimation

Once an estimated passenger-count matrix was computed for the trip pattern of interest, and stops were selected for consolidation, a time-saved matrix was produced. This matrix determined the one-directional in-vehicle travel time (IVTT) savings for a passenger travelling from any Stop A to any downstream Stop B. IVTT was calculated as a function of dwell time savings and acceleration/deceleration time savings estimated using the models described in Chapters 4.5 and 4.8. An example time-saved matrix is shown below in Table 9. Deleted stops are highlighted in grey, and the amount of time saved by stop is listed in brackets beside the destination stop number. For passengers boarding or alighting at a deleted stop, no time savings (or penalties) are realized at that stop, as described in Assumption 4 (Chapter 4.3).
In reality, some remaining (i.e., non-consolidated) stops would impose a minor penalty to on-board passengers following a program of SR, as a result of increased dwell times caused by relocated passengers from consolidated stops (as described briefly in Chapter 4.5). These minor penalties were taken into account in this model, but were not illustrated in the example above (Table 9) for clarity.

### 4.11.3 Access/Egress Penalty Matrix Estimation

As in Chapter 4.9, the terms “access/egress penalty” and “time penalty” refer to the incremental penalty experienced by passengers when comparing existing stop layouts with modified stop layouts, unless described otherwise.

An access/egress penalty matrix (referred to herein as an access penalty matrix, for brevity) was produced to estimate the incremental impacts of SR on anticipated passenger walk times for each origin-destination pair. In principle, this process was simple: a matrix was created to sum the access penalties for each unique combination of origin and destination. However, even the O-D matrix level was considered insufficiently granular for the sake of calculating access penalties.

When considering a single O-D matrix, only the average access penalty associated with SR for a given O-D pair is captured. However, more accurate results can be had when the average value is split into its subcomponents. For example, say that passengers initially accessing Stop B would prefer to use a limited-stop service if, when Stop B is consolidated, their access penalty is no greater than 40 seconds. Assume that half of passengers access Stop B from upstream of the stop, and half from downstream. It could be that passengers accessing Stop B from downstream incur

---

**Table 9: Example Time-Saved Matrix**

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2 (19 s)</th>
<th>3</th>
<th>4</th>
<th>5 (23 s)</th>
<th>6 (17 s)</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>42</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
a 55-second access penalty, whereas passengers accessing Stop B from upstream incur a 45-second access penalty, for an *average* access penalty of 50 seconds. In this case, the average access penalty associated with consolidation of Stop B correctly predicts that none of Stop B’s passengers would choose to use the limited stop service.

On the other hand, it could be that passengers accessing Stop B from downstream incur a 95-second access penalty, while passengers accessing Stop B from upstream incur a 5-second access penalty. In this case, the *average* access penalty would still be 50 seconds, indicating that none of Stop B’s passengers wish to use the limited-stop service, but in reality, the passengers accessing Stop B from upstream would gladly use the service.

Using the access/egress penalty model described in Chapter 4.9, any significant asymmetry of upstream and downstream average access time occurs as a result of the potentially significant (and theoretically infinite) discrepancy between catchment distances on either side of a single stop. A stop could have a catchment distance of 200m upstream and 200m downstream, or a catchment distance of 500m upstream and 50m downstream. In short, changes to an individual’s average access distance could vary drastically depending on whether they intend to access a location downstream or upstream of their preferred stop. Consequently, O-D matrices were produced to represent access penalties for four different groups of passengers:

1. Passengers travelling from *downstream* of an origin stop to *downstream* of a destination stop;
2. Passengers travelling from *downstream* of an origin stop to *upstream* of a destination stop;
3. Passengers travelling from *upstream* of an origin stop to *downstream* of a destination stop; and
4. Passengers travelling from *upstream* of an origin stop to *upstream* of a destination stop.

An example access/egress penalty matrix is shown below ([Table 10](#)) for passengers travelling from downstream of an origin to downstream of a destination. Penalties shown are in seconds.
Consolidated stops are highlighted in grey. For a description of the passenger redistribution matrix methodology, refer to Chapter 4.11.4.

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>70</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>40</td>
<td>40</td>
<td>110</td>
<td>70</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>70</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>70</td>
<td>30</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0</td>
<td>100</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Table 10: Example Access/Egress Penalty Matrix for Downstream/Downstream Passengers (seconds)

Out-of-vehicle (i.e., access/egress) travel time is widely considered more valuable than in-vehicle time in the perceptions of transit passengers. Many studies have proposed a ratio for this perceived value of out-of-vehicle travel time to in-vehicle travel time, ranging as high as 3.0 (El-Geneidy, Strathman, Kimpel, & Crout, 2006). Given the wide range of estimates, and the degree to which one would expect this ratio to be influenced by local climate, built environment, ridership attitudes, and viability of alternative modes of transport, it was left to the user of this software to choose a value for the ratio. This ratio was used as a weight factor to scale up estimated access/egress penalties when calculating the number of passengers estimated to benefit from limited-stop operation.

4.11.4 Passenger Redistribution Matrix Estimation

The total number of passengers initially accessing each O-D pair was estimated using the method described in Chapter 4.11.1, and relocation of passengers due to SR was estimated using the methodology described in Chapter 4.6.2. However, this passenger relocation estimation provided only an aggregate estimate for the total number of passengers boarding and alighting at each O-D pair. To accommodate the breakdown of access penalties into subcategories (classified by upstream and downstream access for both origins and destinations), a similar breakdown was required for the number of passengers expected to access each O-D pair. This breakdown was estimated by assigning relative likelihoods of accessing each stop from the upstream or
downstream side based on stop catchment distances. In the equations below, the total distance between consecutive stops serves as a proxy for catchment distances (which are equal to half the total distance), as it is only the relative distance of upstream versus downstream catchment distance that is significant.

\[ T_{C_{up}} = \frac{D_{bc}}{D_{bc} + D_{cd}} T_C \]  \[ 45 \]

\[ T_{C_{down}} = \frac{D_{cd}}{D_{bc} + D_{cd}} T_C \]  \[ 46 \]

Where:

- \( T_C \) = total passenger count accessing Stop \( c \) (passengers)
- \( T_{C_{up}} \) = total passenger count accessing Stop \( c \) from upstream (passengers)
- \( T_{C_{down}} \) = total passenger count accessing Stop \( c \) from downstream (pass.)
- \( D_{bc} \) = distance from Stop \( b \) (upstream of \( c \)) to stop \( c \) (m)
- \( D_{cd} \) = distance from Stop \( c \) to Stop \( d \) (downstream of \( c \)) (m)

For example, if an average of six passengers were to access Stop C, and Stop C had a downstream catchment distance of 100m and an upstream catchment distance of 200m, then:

\[ T_C = 6 \text{ passengers}, D_{bc} = 200 \text{ m}, D_{cd} = 100 \text{ m} \]

\[ T_{C_{up}} = \frac{200}{200+100} \times (6) = 4 \]

\[ T_{C_{down}} = \frac{100}{200+100} \times (6) = 2 \]

Ultimately, four matrices were produced to break down the total O-D passenger count into subcategories for upstream and downstream access, for both origins and destinations (as described in greater detail in Chapter 4.11.3). An example breakdown is shown below (Table 11 - Table 14). Consolidated stops are highlighted in grey. Each value represents a percentage of
total passenger loading (boardings + alightings) for a given O-D pair. A value of zero indicates that no passengers are expected to fall into the category described by the matrix in question. Note that the matrices below show the breakdowns only for passengers travelling both to and from consolidated stops. Passengers travelling to and/or from a remaining (i.e., non-consolidated) stop are categorized using similar means, but these breakdowns are omitted below for brevity.

Table 11: Example Passenger Redistribution Matrix: Downstream/Downstream Passengers (% of total)

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12: Example Passenger Redistribution Matrix: Downstream/Upstream Passengers (% of total)

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 13: Example Passenger Redistribution Matrix: Upstream/Downstream Passengers (% of total)

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 14: Example Passenger Redistribution Matrix: Upstream/Upstream Passengers (% of total)

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

4.11.5 Limited-Stop Preference Matrix Estimation

Limited-stop preference was estimated as a function of the matrices described in Chapters 4.11.1 through 4.11.4. In short, after a passenger-count O-D matrix was estimated (Chapter 4.11.1), an in-vehicle travel time savings matrix was produced for the same O-D matrix (Chapter 4.11.2). Following this, a series of access/egress penalty matrices were estimated based on probabilistic redistribution of passengers upstream and downstream from consolidated stops (Chapter 4.11.3). Finally, passengers accessing each O-D pair were divided into subcategories based on probabilistic redistribution upstream and downstream from consolidated stops (Chapter 4.11.4).

By comparing passengers’ time savings to their time penalties, it was determined whether passengers in each O-D pair subcategory would be likely to save time or lose time by switching to the user-defined limited-stop service. Combining this with the passenger redistribution matrices described in Chapter 4.11.4, the final product of the limited-stop methodology was a single matrix estimating how many passengers at each O-D pair would benefit from the user-defined limited-stop operation over existing local service. Summing the passengers who benefit from limited-stop operation across the entire max – i.e., across an entire trip pattern – provided an aggregate estimate of the percentage of service capacity which should theoretically be dedicated to the limited-stop operation defined by the user of the software. An example result matrix is shown in Table 15 below.
Table 15: Example Limited-Stop Benefit Matrix

<table>
<thead>
<tr>
<th>Stop</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.90</td>
<td>0.77</td>
<td>0</td>
<td>0.63</td>
<td>0.59</td>
<td>2.88</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
<td>0</td>
<td>0</td>
<td>0.12</td>
<td>0.07</td>
<td>0.20</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>0</td>
<td>0</td>
<td>0.90</td>
<td>0.85</td>
<td>0</td>
<td>0.63</td>
<td>0.77</td>
<td>3.15</td>
</tr>
</tbody>
</table>

The summed value from Table 15 indicates that 3.15 passengers are expected to benefit from adopting the hypothetical limited-stop service over the existing local service. Though clearly passengers in the real world must be counted using only whole numbers, the only goal of this limited-stop analysis is to determine what percentage of buses serving a trip pattern should be converted to limited-stop operation. In this example, if the existing local service had an average loading of 12 passengers per trip, of which 3.15 were estimated to benefit from the proposed limited-stop service, it is proposed that roughly 26.25 percent (or 3.15/12) of buses should be converted to limited-stop operation.

In truth, the above first-order model calculates the number of passengers benefiting from limited-stop operation assuming that all vehicles serving a route are converted to limited-stop operation. In reality, however, not all vehicles would be converted, and not all passengers would choose to use limited-stop vehicles. One corollary is that limited-stop vehicles would be expected to spend less time dwelling than this methodology predicts, and passengers would therefore experience greater time savings than predicted. Refining the time savings estimate would result in a small increase to the number of passengers expected to benefit from limited-stop operation; this in turn would shift a yet smaller number of passengers back to local operation, and so on. Thus, to be more precise and less conservative, this model should theoretically be run iteratively until convergence is achieved. However, this thesis’s modelling effort is by nature high-level and simplistic, and as described elsewhere, intended to provide rough results. Extreme precision
achieved through lengthy iterations was deemed at best unnecessary, and at worst misleading. As such, the software was not designed to complete successive iterations.

4.12 Stop Removal Benefit-Cost Analysis

This thesis’s software was designed primarily to answer “what if” questions posed by the user, rather than to provide prescriptive methods for reducing vehicle and passenger travel times. However, using a method similar in principle to the one proposed by Wagner and Bertini (2014), a Benefit-Cost (B/C) ratio was produced for each stop along a given trip pattern. When the software presented a trip pattern’s list of stops, two lists were provided: one sorted in chronological order, and the other sorted by B/C ratio in descending order (refer to Chapter 5.2 for screenshots). In other words, a feature was added to give the user an indication of the low-hanging SR fruit along each trip pattern. This feature was only made available for SR analysis.

The B/C ratio was calculated as follows:

From [12]: \[ Benefit_j = \bar{T}_P = \bar{P}_j (\Delta T_{bus}) \]

From [14]: \[ Cost_j = V_a (\bar{A}_j + \bar{B}_j) t_{inc} \]

\[ Benefit \text{ to } Cost \text{ ratio at Stop } j = \frac{Benefit_j}{Cost_j} \] \hspace{1cm} [47]

For \( j = 1, 2, ..., k \)

Where:

\( j = \) sequential stop number along a trip pattern (unitless)
\( k = \) total number of stops in a trip pattern (unitless)

\( Benefit_j = \) total onboard passenger time saved due to Stop \( j \) deletion (s)
\( Cost_j = \) total add’l passenger time for access/egress due to Stop \( j \) deletion (s)
\( \bar{T}_P = \) average cumulative trip time savings or penalty for all onboard passengers at Stop \( j \) (s)
\[ \bar{P}_j = \text{average number of onboard passengers at Stop } j \text{ (pass.)} \]
\[ \Delta \bar{T}_{bus,j} = \text{bus time savings or penalty resulting from SC or SA of Stop } j \text{ (s)} \]
\[ V_a = \text{user – set value of access/egress time to in – vehicle time} \text{ (unitless)} \]
\[ \bar{t}_{inc} = \text{incremental average walk time per passenger} \text{ (s)} \]
\[ \bar{A}_j = \text{average number of passengers alighting at existing Stop } j \text{ (pass.)} \]
\[ \bar{B}_j = \text{average number of passengers boarding at existing Stop } j \text{ (pass.)} \]

Thus, the user was able to quickly discern the most impactful stop changes and, by extension, the most impactful “what if” scenarios.

4.13 Summary of Models

This chapter presented a number of Key Performance Indicators (KPIs) relevant to transit service planning and operations activities, as well as the models used to drive the KPIs. The KPIs include metrics relevant to transit agencies (bus run time change; headway change; cycle time change), passenger metrics (passenger in-vehicle time change; passenger access/egress time change; passenger wait time change; and number of passengers expected to prefer limited-stop over local operation), and system KPIs relevant to both passengers and transit agencies (average bus run speed change; productive capacity change; Maximum Load Section load factor change).

As the intent of this thesis is to provide a widely applicable, high-level sketch-planning tool for service planners, simple probabilistic models were employed. New models were created where necessary to calculate the desired KPIs while making best use of the available archived data. New models were thus created to estimate average boarding and alighting volumes; likelihood of vehicle dwell; and the number of passengers expected to prefer a hypothetical limited-stop service over an existing local service. Standard probabilistic models were used to estimate average dwell time; passenger redistribution; acceleration and deceleration time losses/savings; access and egress time losses/savings; and changes to vehicle cycle time, headway, average speed, productive capacity, load factor change, and passenger wait time. A benefit cost ratio was calculated by dividing total passenger in-vehicle time savings by total passenger access/egress time penalties, weighted by the user’s desired value of access time to in-vehicle time.
5 Software Platform

The LTC’s databases were hosted in Microsoft SQL Server 2012, and scripts were created to filter, aggregate, and extract data. The models described in Chapter 4 were initially constructed and tested using Microsoft Excel. Following this, the models were scripted using C#, and a graphical user interface (GUI) was created using HTML and CSS. Computationally, the software ran nearly instantaneously, with a lag time of no more than one or two seconds noticeable to the user; typical use of the software (as demonstrated in Chapter 6) might take the user two or three minutes to complete. To facilitate use of the GUI, the data extract, transform, and load (ETL) processes were automated. Data filters were incorporated into the GUI, and are described in Chapter 5.1. Sample screenshots of the software are shown in Chapter 5.2.

5.1 Data Filters and Parameters

This research aimed to create a software platform enabling conduct stop removal and stop addition (i.e., stop-level) analysis, customized to the user’s desired scope. Facilitating scope customization required that the software sift through the large quantities of AVL and APC data captured by a transit agency; filters provided the mechanism for the user to dictate how the data should be sifted. The filters were designed to interact with one another by limiting the data selected for analysis to the union of the domains of all filters, as depicted in Figure 8.

![Figure 8: The intersection of filter domains was used to determine the subset of data for any analysis.](image-url)
The filters built into the software were as follows:

- Date Range;
- Day(s) of the Week;
- Time of Day;
- Route;
- Trip Name; and,
- Trip Pattern.

The software also allowed the user to set parameters specific to limited-stop analysis:

- Maximum Stop Spacing and Minimum Stop Usage (i.e., Min. Average Passenger Count);
- Value of Access/Egress Time relative to the value of in-vehicle travel time.

These filters and parameters are described in greater detail below. Refer to Chapter 5.2 for screenshots.

5.1.1 Date Range, Day(s) of the Week, and Time of Day

The Date Range filter enabled the user to confine analysis of a trip pattern to any continuous or non-continuous time period for which data was available within the AVL and APC databases. This included selecting particular years, months, days of the month, and days of the week. For example, the user could select any of the following time periods:

- The time period spanning the entire data set;
- The month of December in 2011-2012;
- All Mondays in the first half of each month during May, June, and July, during all years;
- Etc.

In the case of the LTC’s data, trip pattern identification numbers are updated by the LTC roughly every two to six weeks, meaning that only trip patterns falling within the user’s specified time period were made visible to the user and available for analysis. Similarly, any long-range time period set by the user which extended beyond the range of a desired trip pattern would be
meaningless inasmuch as the union of the domain of the time period with that of the trip pattern would negate the entire time period outside of the two- to six-week shelf life of the pattern.

Time of Day was limited more strictly than the Date and Day of Week filters. Because the analysis relied on cycle time and headway to calculate some of the KPIs, a consistent headway and cycle time was required across all trips being aggregated into a single analysis. This necessitated limitation of analysis to a single Time of Day, as defined within the AVL and APC databases.

5.1.2 Route and Trip Pattern

Strictly speaking, since each trip pattern in the LTC’s dataset was unique to a specific route, a Route filter was not necessary. (This is true for any transit agency). However, a Route filter was provided to assist users in honing in on their trip pattern of interest.

After selecting a route, a Trip Pattern Name filter was provided to enable users to narrow their focus to a specific chronological subgroup of trip patterns (representing the evolution of a single trip pattern over time) belonging to the route of choice. Finally, the Trip Pattern filter allowed the user to choose the trip pattern of interest for a given analysis. As described in Chapter 4.1, the trip pattern was considered to be the fundamental level at which analysis should be completed.

5.1.3 Maximum Stop Spacing and Minimum Stop Usage (Limited-Stop)

To expedite the process of limited-stop analysis, the software was designed to automatically determine which stops should be consolidated on the basis of stop spacing and minimum ridership. (Of course, transit agencies must consider other factors when selecting stops for consolidation, such as proximity to important infrastructure – e.g., major roads or schools – and so the user was provided the option to manually toggle the stops slated for consolidation by the software). During limited-stop analysis, the user was required to input a maximum allowable spacing between consecutive stops, as well as a minimum threshold for keeping existing stops based on their existing passenger loading. In other words, the software would remove as many stops as possible without going over the maximum allowable spacing set by the user, and without deleting any stops with an average number of passengers (ons + offs) greater than the threshold
set by the user. Table 16 below shows an example of the automated SR procedure, where maximum spacing between remaining (i.e., limited-stop) stops is set at 800m, and the minimum threshold for keeping a stop based on its passenger loading is five. Stops slated for consolidation are shown in grey.

Table 16: Example of Automated Stop Removal Procedure for Limited-stop Operation

<table>
<thead>
<tr>
<th>Stop</th>
<th>Distance from prev. existing stop (m)</th>
<th>Distance from prev. remaining stop (m)</th>
<th>Avg. Total Passenger Loading (Ons + Offs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>175</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>225</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>325</td>
<td>725</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>250</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>400</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>275</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>225</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>175</td>
<td>800</td>
<td>7</td>
</tr>
</tbody>
</table>

5.1.4 Value of Access/Egress Time

As discussed in Chapter 4.11.3, given the wide range of perceptions of the value of access/egress time relative to in-vehicle travel time, it was left to the user to set this relative value. The outcome of the limited-stop analysis — i.e., the number of passengers estimated to benefit from limited-stop operation — was in part determined as a function of this ratio.

\[ r = \frac{V_{AE}T_{AE}}{T_V} \] \[ 48 \]

\[ L_{OD}(r) = A_{OD} + B_{OD} \quad \text{for } 0 \leq r \leq 1 \]

\[ L_{OD}(r) = 0 \quad \text{for } r > 1 \] \[ 49 \]

Where:

\( r = \text{weighted ratio of access/egress time to in vehicle travel time (unitless)} \)

\( V_{AE} = \text{value of access/egress time relative to in vehicle time (unitless)} \)
\( T_{AE} = \text{incremental access/egress time, or out of vehicle travel time (s)} \)

\( L_{OD}(r) = \text{number of passengers preferring limited stop op. at OD pair (pass.)} \)

\( T_V = \text{incremental in vehicle travel time (s)} \)

\( A_{OD} = \text{average alightings at OD pair, split by OD pair subcategory (pass.)} \)

\( B_{OD} = \text{average boardings at OD pair, split by OD pair subcategory (pass.)} \)

5.2 Desktop Software Platform

The software was constructed as a desktop web-based tool hosted on a remote server. The backend for the software was built using C#, and the GUI was displayed using HTML and CSS. The server hosted the software using Microsoft’s Internet Information Services (IIS) platform. Trip patterns were plotted using the Google Maps Application Programming Interface (API).

Screenshots of the tool are shown below (Figure 9 through Figure 14).
Figure 9: Sample screenshot of overall software graphical user interface
Figure 10: Sample screenshot of filters (with the date and time filters open)
Figure 11: Sample screenshot of Google Maps plot of stops along selected trip pattern (Google Maps, 2015)
Figure 12: Sample screenshot of stops ordered by removal benefit-cost (BC) ratio

Figure 13: Sample screenshot of list of stops on selected trip pattern
A demonstration of the software is provided in Chapter 6.
6 Demonstration and Validation of Sample Results

While many routes in the London Transit Commission’s (LTC’s) network were examined and tested during the creation of this software, Route 13 was selected for use in a detailed demonstration as well as for validation testing. Route 13 is a heavily used route connecting downtown London to the University of Western Ontario, the largest post-secondary institution in southwestern Ontario. The corridor served by Route 13 is also currently being considered by LTC for implementation of a Bus Rapid Transit (BRT) route. Figure 15 shows a map of Route 13 as of the time of writing.

![Figure 15: Route 13 in the LTC Network (London Transit Commission, 2015)](image)

Two demonstrations of the practical applications of the software are provided in Chapter Error! Reference source not found.. One demonstration displays a combined stop removal and addition (i.e., consolidation) assessment, while the second displays a limited-stop conversion assessment.
Validation of simplistic sample SR and SA results are provided in Chapter 6.1.2. Validation was attempted by comparing archived AVL and APC data from select trip patterns on Route 13, which provided the basis for “before-and-after” scenario comparison. Sample results were produced for the each of the following hypothetical trip pattern changes:

- Removal of one stop; and,
- Addition of one stop.

However, while sample results were generated for all KPIs described in Chapter 0, validation was attempted only for KPI (i) (total bus run time savings/penalty). This was a practical limitation imposed by the nature of the available archived data: obviously, no before-and-after data related to passenger wait times, access/egress times, etc. was available within the LTC’s AVL and APC databases, and headway and cycle times were not adjusted by the transit agency following the addition or deletion of a single stop along a trip pattern (all but one of the sample trip patterns contained 63 or 64 stops).

No validation was attempted for the results of the sample trip pattern conversion to limited-stop operation, as no such before-and-after data was available in the LTC’s archived databases.

6.1 Practical Application Demonstration

This thesis’s software is intended to provide transit service planners with estimates of the effects of stop layout changes to a given trip pattern. Neither transit agencies nor passengers will benefit substantially from proposed layout changes unless significant changes are made to the pattern of interest. The demonstrations that follow attempt to showcase the value and efficiency of the software, while addressing typical problems that a transit service planner might face. The demonstrations will focus on removal and addition of stops along Pattern 50649 (Route 13) from the LTC database. Figure 16 below shows a Google Maps plot of the pattern’s original stop layout.
Figure 16: Initial stop layout of Route 13, Pattern 50649
6.1.1 Stop Addition and Removal (Stop Consolidation)

As discussed in Chapters 1 and 2, stop consolidation (removal and addition of stops in tandem) is a common and effective technique for increasing average bus run speed. This demonstration assumes that a service planner has been tasked with finding a stop layout which enables buses to achieve an average run speed of 22.75 km/h during the weekday evening period.

Upon setting the appropriate filters, without selecting any stops for addition or removal, the service planner can run an initial analysis to determine the current average run speed (found to be 21.83 km/h) during the weekday evening period. As seen in Figure 17, the planner is presented with a list of BC ratios pertaining to stops along the pattern.

![Figure 17: BC Ratios for the Monday-Friday evening period along Pattern 50649](image)
At this point, the planner might remove, say, the nine stops with the highest BC ratios (Stops 10, 15-19, and 22-24, as seen in Figure 17) and then rerun the analysis. The stop with the ninth highest BC ratio, Stop 19, has a ratio of 2.69, or high enough to consider stop removal even if access time is considered up to 2.69 times as valuable as in-vehicle time. Removing the abovementioned stops yields an estimated final average run speed of 22.98 km/h (refer to Figure 18).

While the service planner has now met (exceeded) the target final average speed of 22.75 km/h, they might decide that removing a number of consecutive stops (15-19, and 22-24) would be politically unwise, or place undue hardship on the few passengers regularly accessing the removed stops. To counter this, the planner might choose to add one new stop along each of those stretches of the pattern.

The two stops being added by the planner are shown in Figure 18, and the pattern’s final stop layout and final results are shown in Figure 19. The results are summarized in Table 17.
Figure 18: Interim results of stop removal process, and addition of new stops between stops 14 and 20, and stops 21 and 25, along pattern 50649
Figure 19: Final results of Pattern 50649 hypothetical stop removal and addition
Table 17: Results of stop removal and addition (stop consolidation) along Pattern 50649

<table>
<thead>
<tr>
<th>KPI #</th>
<th>KPI Description</th>
<th>Estimated Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Total bus run time saved</td>
<td>107 s</td>
</tr>
<tr>
<td>ii</td>
<td>Bus headway decrease</td>
<td>139 s</td>
</tr>
<tr>
<td>iii</td>
<td>Bus cycle time decrease (run time + terminal time savings)</td>
<td>139 s</td>
</tr>
<tr>
<td>iv</td>
<td>Final average bus run speed</td>
<td>22.71 km/h</td>
</tr>
<tr>
<td>v</td>
<td>Productive capacity increase</td>
<td>62 psngr km/h</td>
</tr>
<tr>
<td>vi</td>
<td>Maximum Load Section load factor change</td>
<td>4 psngrs (4%)</td>
</tr>
<tr>
<td>vii</td>
<td>Total passenger in-vehicle time saved</td>
<td>3,162 s</td>
</tr>
<tr>
<td>viii</td>
<td>Total passenger access time added</td>
<td>213 s</td>
</tr>
<tr>
<td>ix</td>
<td>Total passenger wait time saved</td>
<td>N/A*</td>
</tr>
<tr>
<td>x</td>
<td>Number of passengers benefitting from limited-stop op.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* As described in Chapter 4.10, changes to the LTC’s network were considered unlikely to significantly impact passenger wait times.

As can be seen from the results, the estimated final average speed for the route is 22.71 km/h, which the planner might or might not decide is close enough to the target of 22.75 km/h. An estimated 3,162 passenger in-vehicle seconds will be saved as a result of the stop consolidation effort, while only an estimated 213 seconds will be added to total passenger walk (access/egress) time. The results described in Table 17 seem reasonable, and are well within the realm of plausibility. On first inspection, it may appear that removing stops has little effect on the average run speed of a bus. However, it is worth noting that average run speeds in excess of 20 km/h are quite fast for a transit agency, with 10-15 km/h being more common in large cities. Increasing average run speed by 1 km/h would therefore be equivalent to a roughly 7-10 percent increase for many transit agencies, which is quite substantial.

It is also worth noting the ratio of passenger (in-vehicle) time saved to passenger (access/egress) time lost due to the stop consolidation program demonstrated above (Table 17) predicts a ratio of 14.85 (3,162 / 213), indicating that passengers, taken as a single aggregate unit, would be likely to save a considerable amount of time as a result of implementing the demonstrated stop consolidation program. This highlights the significant benefit that can be delivered to passengers by targeting for removal stops with high BC ratios. Other benefits not considered in this thesis
such as potential increases to total ridership) may also be achievable in some cities if long-term reductions to total travel time are significant.

Refer to Chapter 6.2 for more detailed validation analysis.

6.1.2 Limited-stop Operation Preference Sample Results

As discussed in Chapters 1 and 2, limited-stop operation is introduced by transit agencies looking to provide faster service for passengers who travel long distances, the majority of whom access a number of key stops. As with the demonstration in Chapter 0, Pattern 50649’s weekday evening service will be analyzed. Pattern 50649 makes an appropriate candidate for this demonstration as it is a busy route with two major nodes of passenger activity: downtown London, and the University of Western Ontario. This demonstration assumes that a service planner has been tasked with finding a stop layout which facilitates conversion of 50% of a pattern’s fleet to limited-stop service.

Figure 16 in Chapter 6.1 shows Pattern 50649’s initial stop layout. Three properties allow the service planner to influence how many stops will be automatically recommended for removal by the software: the planner can set the maximum distance between consecutive stops, the minimum threshold for maintaining existing stops, and the value of access/egress time relative to in-vehicle time (these properties are described in greater detail in Chapter 5.1). Initially, the planner might decide to set the access time to in-vehicle time ratio to 1; the maximum distance between stops to 900m; and the minimum passenger threshold to four (i.e., for a stop to be removed, it must have a boarding + alighting count of four or fewer passengers, on average). Using these parameters, the stop layout and KPI results shown in Figure 20 are obtained. The results are summarized in Table 18.
Figure 20: Pattern 50649 limited-stop analysis, with maximum 900m spacing
Table 18: Pattern 50649 Limited-Stop Results, 900m spacing (Psngr. Threshold = 4, Walk Value Ratio = 1)

<table>
<thead>
<tr>
<th>KPI #</th>
<th>KPI Description</th>
<th>Estimated Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Total bus run time saved</td>
<td>578 s</td>
</tr>
<tr>
<td>ii</td>
<td>Bus headway decrease</td>
<td>754 s</td>
</tr>
<tr>
<td>iii</td>
<td>Bus cycle time decrease (run time + terminal time savings)</td>
<td>754 s</td>
</tr>
<tr>
<td>iv</td>
<td>Final average bus run speed</td>
<td>27.61 km/h</td>
</tr>
<tr>
<td>v</td>
<td>Final productive capacity</td>
<td>1,933 psngr km/h</td>
</tr>
<tr>
<td>vi</td>
<td>Maximum Load Section load factor change</td>
<td>-21%</td>
</tr>
<tr>
<td>vii</td>
<td>Total passenger in-vehicle time saved</td>
<td>12,451 s</td>
</tr>
<tr>
<td>viii</td>
<td>Total passenger access time added</td>
<td>7,181 s</td>
</tr>
<tr>
<td>ix</td>
<td>Total passenger wait time saved</td>
<td>N/A*</td>
</tr>
<tr>
<td>x</td>
<td>Number of passengers benefitting from limited-stop op.</td>
<td>46 (42.1%)</td>
</tr>
</tbody>
</table>

* As described in Chapter 4.10, changes to the LTC’s network were considered unlikely to significantly impact passenger wait times.

The results show a total of 38 stops would be removed, and 26 stops would remain in place. Given that only 42% of the total ridership is expected to benefit from the proposed limited-stop configuration, the planner might choose to reduce the maximum spacing between stops to 800m. Rerunning the analysis with 800m maximum spacing, the stop layout and KPI results shown in Figure 21 are obtained. The results are summarized in Table 19.
Figure 21: Pattern 50649 limited-stop analysis, with maximum 800m spacing
Table 19: Pattern 50649 Limited-Stop Results, 800m spacing (Psngr. Threshold = 4, Walk Value Ratio = 1)

<table>
<thead>
<tr>
<th>KPI #</th>
<th>KPI Description</th>
<th>Estimated Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Total bus run time saved</td>
<td>515 s</td>
</tr>
<tr>
<td>ii</td>
<td>Bus headway decrease</td>
<td>671 s</td>
</tr>
<tr>
<td>iii</td>
<td>Bus cycle time decrease (run time + terminal time savings)</td>
<td>671 s</td>
</tr>
<tr>
<td>iv</td>
<td>Final average bus run speed</td>
<td>26.83 km/h</td>
</tr>
<tr>
<td>v</td>
<td>Final productive capacity</td>
<td>1,878 psngr km/h</td>
</tr>
<tr>
<td>vi</td>
<td>Maximum Load Section load factor change</td>
<td>-19%</td>
</tr>
<tr>
<td>vii</td>
<td>Total passenger in-vehicle time saved</td>
<td>11,164 s</td>
</tr>
<tr>
<td>viii</td>
<td>Total passenger access time added</td>
<td>6,213 s</td>
</tr>
<tr>
<td>ix</td>
<td>Total passenger wait time saved</td>
<td>N/A*</td>
</tr>
<tr>
<td>x</td>
<td>Number of passengers benefitting from limited-stop op.</td>
<td>56 (50.8%)</td>
</tr>
</tbody>
</table>

The results show a total of 34 stops would be removed, and 30 stops would remain in place. Bus run time is expected to decrease by an average of 515 seconds (8.5 minutes), and average bus run speed is estimated to increase from 21.83 km/h to 26.83 km/h.

For the sake of comparison, the planner might recompute the results using a ratio of access time to in-vehicle time of 2. Notably, only KPI (x) would change as a result of changing this ratio:

Table 20: Pattern 50649 Limited-Stop Results, 800m spacing (Psngr. Threshold = 4, Walk Value Ratio = 2)

<table>
<thead>
<tr>
<th>KPI #</th>
<th>KPI Description</th>
<th>Estimated Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>Number of passengers benefitting from limited-stop op.</td>
<td>43 (38.8%) %</td>
</tr>
</tbody>
</table>

The result of KPI (x) from Table 19 and Table 20 suggests that about 51 percent of passengers would benefit from limited-stop operation assuming equal valuations of access and in-vehicle time, whereas only 39 percent of passengers would benefit from limited-stop operation if access time is valued at twice that of in-vehicle time. Understanding that the limited-stop frequency share of a pattern’s fleet be set at roughly the percentage determined for KPI (x), it is thus implied that either 51 percent or 39 percent of local operation buses should be converted to performing limited-stop operations, depending on which result the service planner selects. The discrepancy between Table 19 and Table 20 is particularly relevant to a service planner given that, according to Schwarcz (2004), limited-stop service is typically most effective when its frequency is no less than 50 percent of the total for the route.
Obviously, the type of limited-stop conversion program demonstrated above would require commitment and significantly more detailed planning from a transit agency in order to be successful. However, if implemented carefully and in appropriate locations, the magnitude of time savings described in Table 19 should be achievable.

The demonstration above suggests that buses converted to limited-stop operation would achieve an estimated average run speed of 26.83 km/h, 5 km/h faster than existing local operation speeds (described Chapter 0). This represents a 23 percent increase in average speed. For routes served by a large fleet, this magnitude of improvement would likely allow the agency to reduce its fleet size without lengthening existing headways, which would translate to a major cost savings.

Refer to Chapter 6.2 for more detailed validation analysis.

### 6.2 Validation

Validation analysis was undertaken for simplistic sample results for each of SA and SR. Two existing trip patterns were selected as base (control) cases to which changes were simulated using this thesis’s software (refer to Table 21). The patterns were selected on the basis of their mutual similarity, with the exception of the number of stops (the primary explanatory variable) and the average actual run time (the difference between the patterns’ respective run times was one of the results being estimated). The two patterns operated in the same direction along the same route, and the values shown were calculated using archived data gathered during non-holiday weekday evening trips (one of the LTC’s time-of-day categories as defined in the AVL and APC databases).

Simplistic sample results for SR and SA are shown in Chapters 0 and 6.2.2, respectively. Validation of the simplistic sample results is discussed in Chapter 6.2.3.

**Table 21: Test (Control) Trip Patterns Examined For Sample Results and Validation**

<table>
<thead>
<tr>
<th>Pattern ID</th>
<th>Dates of Operation</th>
<th>Number of Stops</th>
<th>Sched. Run Time / Headway (s)</th>
<th>Avg. Actual Run Time (s)</th>
<th>Avg. Psngr Load (Total Ons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50649</td>
<td>25 June – 1 Sept 2012</td>
<td>64</td>
<td>2,760 / 3,600</td>
<td>2,703</td>
<td>110</td>
</tr>
<tr>
<td>65414</td>
<td>24 July – 31 Aug 2013</td>
<td>63</td>
<td>2,760 / 3,600</td>
<td>2,694</td>
<td>111</td>
</tr>
</tbody>
</table>
6.2.1 Stop Removal (SR) Sample Results

Stop removal analysis was performed using Pattern 50649 as the base case. Stop 6 was selected for (simulated) removal as it was the only stop found in Pattern 50649 not present in Pattern 65414. Stop 6 had an average ridership of two boardings and one alighting during the evening period. The pattern’s fleet consisted of one vehicle. The following results were obtained from the simulation:

Table 22: Trip Pattern 50649 Stop Removal Sample Results

<table>
<thead>
<tr>
<th>KPI #</th>
<th>KPI Description</th>
<th>Estimated Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Total bus run time saved</td>
<td>15 s</td>
</tr>
<tr>
<td>ii</td>
<td>Bus headway decrease</td>
<td>20 s (0.56 %)</td>
</tr>
<tr>
<td>iii</td>
<td>Bus cycle time decrease (run time + terminal time savings)</td>
<td>20 s</td>
</tr>
<tr>
<td>iv</td>
<td>Average bus run speed increase</td>
<td>0.12 km/h</td>
</tr>
<tr>
<td>v</td>
<td>Productive capacity increase</td>
<td>8.2 psngr.km/hr</td>
</tr>
<tr>
<td>vi</td>
<td>Maximum Load Section load factor change</td>
<td>0.2 psngrs (0.56%)</td>
</tr>
<tr>
<td>vii</td>
<td>Total passenger in-vehicle time saved</td>
<td>405 s</td>
</tr>
<tr>
<td>viii</td>
<td>Total passenger access time added</td>
<td>223 s</td>
</tr>
<tr>
<td>ix</td>
<td>Total passenger wait time saved</td>
<td>N/A*</td>
</tr>
<tr>
<td>x</td>
<td>Number of passengers benefitting from limited-stop op.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* As described in Chapter 4.10, changes to the LTC’s network were considered unlikely to significantly impact passenger wait times.

Of course, the results indicate that the consolidation of Stop 1965 alone is inconsequential to route performance; total run time, headway, maximum load section load factor, and passenger travel times would all be expected to realize only minor improvements, most of which would be considered rounding errors in actual practice. However, this outcome is to be anticipated when deleting a single stop.

6.2.2 Stop Addition (SA) Sample Results

Stop Addition analysis was performed using Pattern 65414 as the base case. Stop 6 was selected for (simulated) addition as it was the only stop found in Pattern 50649 not present in Pattern
65414. The pattern’s fleet consisted of one vehicle. The following results were obtained from the simulation:

Table 23: Trip Pattern 65414 Stop Addition Sample Results

<table>
<thead>
<tr>
<th>KPI #</th>
<th>KPI Description</th>
<th>Estimated Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Total bus run time added</td>
<td>15 s</td>
</tr>
<tr>
<td>ii</td>
<td>Bus headway increase</td>
<td>20 s</td>
</tr>
<tr>
<td>iii</td>
<td>Bus cycle time increase (run time + terminal time savings)</td>
<td>20 s</td>
</tr>
<tr>
<td>iv</td>
<td>Average bus run speed change</td>
<td>0.11 km/h</td>
</tr>
<tr>
<td>v</td>
<td>Productive capacity change</td>
<td>8.4 psngr.km/hr</td>
</tr>
<tr>
<td>vi</td>
<td>Maximum Load Section load factor change</td>
<td>0.2 psngrs (0.56%)</td>
</tr>
<tr>
<td>vii</td>
<td>Total passenger in-vehicle time added</td>
<td>360 s</td>
</tr>
<tr>
<td>viii</td>
<td>Total passenger access time saved</td>
<td>87 s</td>
</tr>
<tr>
<td>ix</td>
<td>Total passenger wait time added</td>
<td>N/A*</td>
</tr>
<tr>
<td>x</td>
<td>Number of passengers benefitting from limited-stop op.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* As described in Chapter 4.10, changes to the LTC’s network were considered unlikely to significantly impact passenger wait times.

The results of the above SA procedure are similar in magnitude (though opposite in direction) to the results of the SR procedure shown in Chapter 0, as one would expect. KPIs (vii) and (viii), however, vary significantly between the models. On closer inspection, this result is to be expected given the relative change in passenger distribution along the route which developed during the intervening 12-month interval between life cycles of the two patterns:

Table 24: Passenger Distribution Along Patterns 50649 and 65414

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>20</td>
<td>0</td>
<td>(1)</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>…</td>
<td>…</td>
<td></td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>(5)</td>
<td>2</td>
<td>1</td>
<td>(5)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(6)</td>
<td>2</td>
<td>1</td>
<td></td>
<td>(stop removed by the LTC)</td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>2</td>
<td>1</td>
<td>(7)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>…</td>
<td>…</td>
<td></td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

As [Error! Reference source not found.](#) shows, while Pattern 65414 did not serve Stop 6, average passenger volumes at Stops 5 and 7 did not increase for Pattern 65414 relative to Pattern 50649, indicating that total average passenger volume across this stretch (Stops 5 through 7) was
actually lower for Pattern 65414. For KPI (viii), it is therefore reasonable to expect that, when holding ridership constant, passenger access time savings due to the addition of Stop 6 to Pattern 65414 would be smaller in magnitude than the penalty for removing Stop 6 from Pattern 50649. Similarly, for KPI (vii), it is logical that passenger in-vehicle time savings would be lower for Pattern 65414, which had fewer passengers boarding (and therefore lower departure loads) at stops leading up to the stop of interest when compared to Pattern 50649.

This ridership distribution discrepancy between the two patterns highlights the difficulty of selecting like patterns for before-and-after comparison; this is discussed in greater detail in Chapter 6.2.3.

6.2.3 Validation of Sample Results

To validate the results obtained in Chapter 0 and 6.2.2, the test trip patterns (50649 and 65414) were compared to a number of validation trip patterns which exhibited similar characteristics. It is worth stressing that validation was attempted using archived before-and-after data, not performed actively with the LTC’s cooperation. A plethora of characteristics varied from one trip pattern to the next, of which many could not be controlled for in their entirety; this variation is almost certainly true across trip patterns defined by any transit agency which has not intentionally made an attempt to the contrary. For example, consider the incomplete list below of characteristics which can vary between trip patterns, even within a single route:

- Route direction;
- Stop layout;
- Timepoint layout/designation;
- Cycle time and/or run time;
- Temporal changes (time of day, day(s) of the week, month, season, year);
- Construction or other temporary events impacting operational efficiency;
- Average number of passengers accessing the service;
- Passenger distribution across the service;
- Any systematic bias in the way data is recorded (e.g., perhaps GPS reception is poor in downtown areas, and therefore data is recorded more frequently and more accurately in low density areas).

- Etc.

In recognition of the difficulty of controlling such characteristics, five independent validation trip patterns were selected to provide an indication of the success of the simulation. The test and validation trip patterns are described as follows in Table 25: all data pertains only to “evening” trips in the same direction along Route 13.

### Table 25: Test and Validation Trip Pattern Characteristics

<table>
<thead>
<tr>
<th>Pattern ID</th>
<th>Dates of Operation</th>
<th>No. of Stops</th>
<th>Sched. Run Time (s)</th>
<th>Avg. Actual Run Time (s)</th>
<th>Avg. Psngr Load (Total Ons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50649 (Test)</td>
<td>25 June - 1 Sept 2012</td>
<td>64</td>
<td>2760</td>
<td>2703</td>
<td>110</td>
</tr>
<tr>
<td>65414 (Test)</td>
<td>24 July - 31 Aug 2013</td>
<td>63</td>
<td>2760</td>
<td>2694</td>
<td>111</td>
</tr>
<tr>
<td>60819 (Validation)</td>
<td>19 Feb - 27 Apr 2013</td>
<td>63</td>
<td>2760</td>
<td>2894</td>
<td>133</td>
</tr>
<tr>
<td>62790 (Validation)</td>
<td>29 Apr - 8 May 2013</td>
<td>63</td>
<td>2760</td>
<td>2762</td>
<td>111</td>
</tr>
<tr>
<td>63447 (Validation)</td>
<td>9 May - 22 June 2013</td>
<td>63</td>
<td>2760</td>
<td>2749</td>
<td>114</td>
</tr>
<tr>
<td>63460 (Validation)</td>
<td>10 May - 22 June 2013</td>
<td>85</td>
<td>3480</td>
<td>3342</td>
<td>108</td>
</tr>
<tr>
<td>64104 (Validation)</td>
<td>24 June - 23 July 2013</td>
<td>63</td>
<td>2760</td>
<td>2681</td>
<td>107</td>
</tr>
</tbody>
</table>

A quick visual comparison of the average actual run times in Table 25 indicates that both the number of stops and the average number of boardings during a trip contribute positively to a trip’s run time. Linear regression analysis (with a 95% level of confidence) of the data from Table 25 confirmed this finding, with each additional stop estimated to incur a penalty of roughly 29 seconds, as shown below in Table 26.

### Table 26: Linear regression of passenger load and number of stops vs. actual observed run time

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.95</td>
<td>213.40</td>
<td>0.04</td>
<td>0.97</td>
</tr>
<tr>
<td>Passenger Load</td>
<td>7.88</td>
<td>1.40</td>
<td>5.62</td>
<td>0.01</td>
</tr>
<tr>
<td>----------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Number of Stops</td>
<td>29.21</td>
<td>1.50</td>
<td>19.42</td>
<td>0.00</td>
</tr>
</tbody>
</table>

However, given Table 26’s minute sample size (n = 7), while the general trends highlighted by the linear regression might be considered indicative of the dataset, little stock should be placed in the magnitude of the coefficients. Moreover, Pattern 63460 contained a detour to access additional stops, and thus was responsible for travelling a longer distance than the other patterns in the validation dataset, contributing to longer run time. The regression model’s estimated 29-second penalty for additional stops thus internalized any run time penalties associated with additional travel distance.

A comparison of Patterns 50649, 65414, and 64104 appears superficially to provide excellent validation (refer to Table 25). Pattern 50649 had one extra stop and one fewer rider (on average) than Pattern 65414, and its average run time exceeded that of 65414 by nine seconds. While this value is less than the 15 seconds predicted by the simulation in Chapter 0, this discrepancy can be at least partially explained by the additional rider on Pattern 65414 (the simulation assumes constant ridership). Similar study reveals that Pattern 50649 had one extra stop and three extra riders in comparison to Pattern 64104, and that 50649’s run time was 22 seconds longer than that of Pattern 64104 on average. This value exceeds the 15 seconds predicted by the simulation, which again indicates that the change in ridership contributed to the observed difference in average run time. Moreover, the simulated change in run time – 15 seconds – falls nicely between the pair of actual results – 9 seconds and 22 seconds – and thus appears to roughly validate the prediction of the simulation.

In reality, the above validation narrative may be no more than a reflection of the adage, *if you look hard enough, you’ll find it*. Within the millions of possible comparisons between similar trip patterns which could be conducted using the LTC’s archived data, one is bound to find many examples where the simulation predicts the exact change in run time observed in reality. Similarly, one is bound to find many examples to the contrary, including, for instance, Patterns 62790 and 63447, which had only marginally higher average ridership than Pattern 50649, yet experienced distinctly longer average run times – despite having one fewer stop. Perhaps most
telling of all, though Pattern 63447 had higher ridership than Pattern 62790 and the same stop layout, 63447 actually experienced a shorter average run time.

Thus, while the sample results and validation above are generally consistent with expectations and true observed run time changes, and may provide an indication of the success of the models, this should not be taken as conclusive validation. Conclusive validation would require careful and intentional before-and-after changes to a pattern made by the LTC, with careful observation and analysis of the results, which was not feasible within the scope of this thesis.

In lieu of conclusive validation, the time savings/penalty estimates produced using this software were compared with observed run time changes noted in other studies testing the effects of SA and SR.

Virtually all net bus run time savings/penalty estimates produced using this thesis’s SR/SA methodology were in the range of 14-17 seconds, and the modal, mean, and median time savings estimate were each 15 seconds. Though determined anecdotally rather than through exhaustive systematic testing, this result was unsurprising: consolidation of any stop was estimated to save nine seconds due to reduced acceleration/deceleration time losses (as described in Chapter 4.8), and six seconds of door open/door close time (set in the dwell time estimator), for a total of 15 seconds. Given an assumption of constant ridership along a route, and this thesis’s reliance on the TCQSM’s dwell time estimator rather than observed average dwell times (as discussed in Chapter 3.3), it is predictable that consolidation or addition of a stop would generally redistribute, rather than change, the time required for passengers to board and alight; only minor differences to a trip’s average total loading time are predicted based on pattern-specific configurations of ons and offs at each stop, so a time savings/penalty range of 14-17 seconds per stop is fitting.

As discussed in Chapter 2.3, Furth & Rahbee (2000) estimated that bus run time savings following an optimized SR program would be in the order of 14-15 seconds per stop, of which nine seconds would be saved due to elimination of acceleration and deceleration time.
Guided by an initial “rule of thumb” of 20 seconds saved per deleted stop, de Vries Kehoe (2004) produced a regression model to estimate effects of SR observed during a study of Seattle’s transit system. The model assumed constant before-and-after ridership. Using de Vries Kehoe’s model, consolidation of a Stop C, located halfway between Stops A and B (where stops A and B are spaced 400m apart), would result in a time savings of 15 seconds (refer to Chapter 2.3 for definition of the model).

El-Geneidy, Strathman, Kimpel, & Crout (2006) found that bus running times along their test route segment in Portland were reduced by about 5-6 seconds per stop, in contrast to their regression model which estimated a run time reduction of 42.2 seconds per consolidated stop.

Tetreault & El-Geneidy (2010) determined that SR produced an average run time savings of 12.9 seconds per stop in their Montreal study, with the majority of time savings coming from elimination of acceleration and deceleration time. These findings were mirrored in a study by Stewart & El-Geneidy (forthcoming).

In their benefit-cost evaluation method for determining appropriate stops for consolidation, Wagner and Bertini (2014) set the time saved per consolidated stop at 15 seconds, selected as the midrange value within the 10-20 second range described by the Washington Metropolitan Area Transit Authority (2009).

Interestingly and perhaps most meaningfully, as of the time of writing, the STM in Montreal has set schedule times for its 67 (local) route and 467 (limited-stop) route, which travel the same corridor, such that the 467 is expected to save roughly 12-24 seconds per consolidated stop, depending on the time of day.

In short, there is a substantial body of literature which has attempted to estimate time savings per consolidated stop, and generally speaking, the findings of the literature closely match the findings of this paper.
7 Conclusions, Limitations, and Future Work

This section summarizes the intent, methodology, limitations, and findings of this thesis, and describes avenues of future research.

7.1 Summary

As a matter of course, transit agencies put substantial effort towards minimizing expenditures and vehicle and operator active time, as well as attracting new customers. Doing so without impacting existing service quality requires increasing system efficiency. One means of increasing system efficiency is to increase stop spacing through a stop removal (SR) process; this is especially relevant for surface transit routes in the United States, where stop spacing is typically in the order of 160–230m (Furth & Rahbee, 2000), and Canada, which has similar average stop spacing. Stop removal is a regular occurrence and a common technique for increasing average vehicle speed, reducing crowding and headways, saving passenger and operator time, and ultimately reducing operator costs. A means of assessing the effects of SR and the reverse process, stop addition (SA), is thus critical for any transit agency.

Any quantitative analysis requires data, and all other things being equal, the better the quality and quantity of relevant data, the better the resulting analysis. Data analytics, the process of analyzing raw data to answer specific business or scientific questions, is the engine that drives fact-based decision-making. Both terms fall under the umbrella of data science. The introduction of Automatic Data Collection (ADC) systems such as Automatic Vehicle Location (AVL) and Automatic Passenger Counters (APCs) on transit vehicles has created an opportunity for transit agencies to gather huge quantities of data at very low marginal cost. While much of this data is generated primarily for the sake of managing real-time operations such as dispatching, the data is also being recorded in long-term storage which facilitates use of data analytics to enable highly disaggregate fact-based decision-making. The value of disaggregate, situation-specific fact-based decision-making has become widely recognized in the transit industry in recent decades.

As in many industries, fact-based decision-making for transit performance management is rooted in before-and-after comparison of objective Key Performance Indicators (KPIs). In this vein, the
LTC engaged IBI Group and the University of Toronto in October 2013 to develop near-real-time Business Intelligence (BI) dashboards for descriptive reporting of KPIs based on their AVL and APC data. Like many transit agencies, the LTC had previously been conducting manual data analysis to estimate KPIs on a quarterly basis.

Following the completion of the above pilot project with the LTC, this thesis was conceived to expand the use of data analytics to include predictive reporting tools powered by LTC’s AVL and APC databases. Specifically, this thesis sought to create data analytics software to produce the following KPIs as a means of estimating the benefits of SR, SA, or introduction of limited-stop operation:

i. Total bus run time saved/added, broken down by stop;
ii. Bus headway increase/decrease;
iii. Bus cycle time increase/decrease;
iv. Average bus run speed increase/decrease;
v. Productive capacity increase/decrease;
vi. Maximum Load Section load factor demand increase/decrease;

vii. Total passenger in-vehicle time saved/added, broken down by stop;
viii. Total passenger access time saved/added, broken down by stop;
ix. Total passenger wait time saved/added based on changes to vehicle headways, broken down by stop;

x. Number of existing passengers expected to benefit from limited-stop over local operation, given the particular stop layout the user has selected.

While this thesis’s methodology is generally applicable to any transit agency, the research was tailored to accommodate the idiosyncrasies of LTC’s dataset, and may not be directly applicable or optimal for analysis of automatically generated data from other transit agencies.

While previous studies have focused SR or limited-stop assessment on a small number of important or representative routes, the intention of this study was to produce an automated, standardized approach that could be applied to any route, given the necessary raw data. This software was designed to accommodate modifications to existing bus trip patterns rather than
existing bus routes, as trip patterns are the basic building blocks for trip and run organization by a transit agency.

Algorithms were created to clean and aggregate the LTC’s archived data according to filters applied by the user. Filters included the operation year(s), month(s), day(s), day(s) of the week, and time of day. Given that the LTC’s AVL data was incomplete (with arrival and departure times only recorded at the timepoint level, rather than the stop level) and suffered from systemic and insurmountable inaccuracies at some mid-pattern stops, dwell times were estimated using an estimator provided in the Transit Capacity and Quality of Service Manual, 3rd edition, or TCQSM (Kittelson & Associates Inc., Parsons Brinckerhoff, KFH Group Inc., Texas A&M Transportation Institute, & Arup, 2013). The primary inputs dictating dwell time estimation using the TCQSM’s model are passenger boardings and alightings, and vehicle door open and close times. Thus, stop-specific passenger counts recorded in the LTC’s APC database were used to estimate vehicle dwell times in lieu of the unreliable AVL data.

The cleaned and aggregated data was used to power a number of simple probabilistic models estimating passenger relocation, dwell times savings/penalties, a vehicle’s likelihood of dwell, losses due to acceleration and deceleration, and passenger access and egress penalties, based on the user’s desired stop layout changes. The models were used in combination to compute the desired KPIs. For the limited-stop operation procedure, stops were selected for consolidation by the software based on user-set maximum spacing between consecutive stops and minimum passenger loading threshold for maintaining service to any given stop.

Assumptions behind the models are summarized in Chapter 7.3.

The models were coded using C#, and a graphical user interface was created to facilitate easy implementation for the end user.

The tool was demonstrated and sample results were produced for trip patterns belonging to Route 13 in the LTC’s network, and a validation procedure was conducted. Refer to Chapter 0 for further information.
7.2 Conclusions

As with any new technology, the introduction of ADC systems for collection, curation, and analysis of transit data has brought with it many new and surprising challenges in addition to the multitudes of benefits. Numerous reports over the last decade have described both the pitfalls and the best practices that agencies have adopted, with best practices including prioritization of time-at-location data over location-at-time data; storage of data at its most granular level; integration of various on-board sensors (such as APC and AFC systems); and careful database design to render post-operation analysis feasible (TCRP Web Document 23 (2003); Furth, Muller, Strathman, and Hemily (2004); Hammerle, Haynes, and McNeil (2005); TCRP Report 113 (2006)). Unfortunately, however, these recommendations remain topical as these best practices have by no means achieved universal adoption. ADC systems installed in numerous municipalities in southern Ontario (including the LTC’s system), though installed within the last 5-10 years, fail to follow all guidelines outlined in these reports. While the ability of transit agencies to store and process large quantities of data has improved vastly over the last decade in step with advances in computer processing power, data analysis can only ever be as good as its raw data. The limitations imposed by ADC system imperfections are tangible; many are described in Chapter 7.3.

There are a number of risks when using data analytics to assess archived data. While automating the analysis process tends to greatly reduce human error, machine error (such as sensor miscalibration) is introduced. Additionally, analysis of millions or billions of data points is impossible for the human eye to review exhaustively. Systemic mistakes tend to camouflage and can be difficult to root out, especially where they fail to cause remarkable results. One inevitable solution is to create algorithms to automate the quality assurance process, employing strategies and screening procedures and minimize the effects of error propagation. For example, TCRP 113 (2006) recommends a strategy for balancing passenger counts at points of known zero passenger loads on buses. Simple spot-checking to ensure accurate calculations and plausible results is also critical.
In line with findings of El-Geneidy, Horning, and Krizek (2011), this thesis confirmed the impracticality of assessing KPIs at the route level, and concludes that analysis must be conducted at the trip pattern or trip segment level, subcategorized by time of day, in order to avoid measurement confusion. This is especially important when unremarkable specious results render spot-checking ineffective.

Vehicle run time savings associated with SR (simulated using this thesis’s software) typically ranged from 14-17 seconds per consolidated stop, with 15 seconds being the mean, median, and modal value where spot-checked. Such values are in line with estimates in existing literature (de Vries Kehoe (2004); Tetreault & El-Geneidy (2010); Washington Metropolitan Area Transit Authority (2009)). Given the minor time savings associated with consolidation of any given stop, this thesis concludes that consolidation of a small number of stops along a trip pattern is unlikely to have a significant impact on a trip pattern’s average speed, schedule, on-time performance, headway, etc. Deletion of a substantial portion of a pattern’s stops is likely required to enable meaningful schedule or headway changes, especially if SR is intended to attract additional ridership. Vehicle run time penalties associated with SA were likewise roughly 15 seconds per added stop.

The (albeit anecdotally) observed low variation in time savings/penalty estimates was not anticipated, but in retrospect, this result is unsurprising given this paper’s universal estimate for acceleration and deceleration time losses, and estimated door open/close time. Notably, summing the estimated acceleration and deceleration time losses (nine seconds total per stop) plus the estimated door open/close time (six seconds total per stop) yielded a value of 15 seconds – equal to the typical predicted time change for adding or deleting a stop. This result is logical. Because time savings were computed using a static estimator tool, rather than true, dynamic, observed dwell times, site-specific passenger loading conditions were not taken into account. Combined with an assumption of constant before-and-after trip pattern ridership, it is reasonable that the model would generally redistribute, rather than change, the time required for passengers to board and alight. Only minor differences to a trip’s average total loading time are predicted based on pattern-specific configurations of ons and offs at each stop, so an average time savings of 15 seconds per stop is virtually inevitable given fixed costs of 15 seconds per stop.
The fact that run time savings were consistent regardless of which stop was deleted from a trip pattern is a shortcoming of relying upon archived APC data to estimate AVL (i.e., dwell time) data. This served as an indication that true vehicle dwell times would be needed in order to produce more nuanced results.

Sample results were produced for hypothetical SR and SA changes to trip patterns on Route 13 in the LTC’s network. Validation was attempted for KPI (i) (predicted vehicle run time savings/penalty) by comparing the sample results with true before-and-after data recorded in the LTC’s databases. The sample results were in line with changes observed in some before-and-after pattern comparisons, but not all. For example, some before-and-after comparisons showed that a trip pattern with a lesser number of stops had a longer average run time than a previous incarnation of the same pattern with a greater number of stops, even after controlling for a number of factors including total average ridership. Clearly, not all factors affecting total vehicle run time could be controlled for using only archived data for before-and-after comparisons. Conclusive validation would require a participating transit agency to make intentional before-and-after changes to a pattern, with careful observation and analysis of the results, which was not feasible within the scope of this thesis. In lieu of conclusive validation, the time savings/penalty estimates produced using this software were compared with observed run time changes noted in other studies, as discussed at the beginning of this section.

Sample results were also produced for a hypothetical introduction of limited-stop service, though no relevant data was available to enable validation. The results indicate that changing the assumed ratio of perceived value of access time to in-vehicle time has a significant effect on the number of passengers expected to choose the limited-stop service.

Initial testing of the software using sample data from the LTC dataset also indicated there is a need for more active and frequent management of trip pattern schedules. Many trip patterns in the dataset appear to have a standard scheduled run time which is consistently exceeded; other trip patterns have the opposite problem. It was hoped that by simplifying the process of conducting this sort of analysis, users of this software would be able to complete more regular and more effective schedule tweaks.
7.3 Limitations

Three types of limitations exist to this paper’s findings: *methodological* limitations, consisting primarily of simplifications such as the assumption of constant ridership regardless of stop layout changes to a trip pattern; *data-imposed* limitations, such as the lack of granularity of AVL data recorded in the LTC’s database; and *data-imposed methodological* limitations, referring to limitations which could be eliminated by improving this thesis’s models, given a dataset of sufficient quality and granularity. Obviously, the methodological limitations imposed by the models in this thesis would apply to any transit agency’s dataset, regardless of data quality, whereas the data-imposed limitations are specific to this thesis’s findings, and may be alleviated through use of a more complete, accurate, and granular dataset. The third category of limitation falls somewhere in between.

7.3.1 Methodological Limitations

Each assumption embedded into this thesis’s software is in truth a limitation to its use. Given the intent to create *high-level* models for estimating the effects of stop layout changes, the following simplifying assumptions were made:

- The existing first and last stop along each trip pattern were assumed to remain as the terminus stops, and it was assumed that no dwell time savings or penalties would be realized at these stops, even if an adjacent stop was added or consolidated;

- A maximum of one stop could be added between any existing pair of stops, though this process could be performed iteratively;

- It was assumed that passengers would walk to the nearest stop;

- Uniform passenger demand distribution was assumed *within* each stop’s catchment area;

- The catchment distance for a stop was assumed to extend halfway to immediately adjacent stops in each direction;
- It was assumed that no change in total pattern ridership would be triggered by changes to
  the stop layout, given the relatively minor changes being made;

- It was assumed that bus capacity to accommodate demand is unchanged from existing
  conditions as a result of changes to the trip pattern stop layout;

- It was assumed that the road network functioned as a perfect grid for the sake of
  calculating changes to access/egress distance to/from stops;

- No passenger wait time savings were expected as a result of any moderate headway
  changes caused by stop layout changes.

7.3.2 Data-Imposed Methodological Limitations

Because the LTC’s AVL system recorded departure and arrival time data only at timepoints,
rather than at all stops, and because many mid-route arrival and departure times were deemed
untrustworthy, dwell time savings and penalties were estimated using stop-level passenger count
data to inform the TCQSM’s dwell time estimator tool. This resulted in two data-imposed
methodological limitations. First, dwell time savings/penalties needed to be estimated using a
generic tool, rather than being estimated using true historical dwell time values observed at the
stops of interest. Second, because total trip pattern ridership was assumed to remain constant,
there was little variation in dwell time savings/penalty estimates; for example, deleting any
single stop along a trip pattern was predicted to save roughly 15 seconds of bus run time,
regardless of the stop’s ridership or local geography. Using true (observed) dwell times in place
of estimator values would by definition take into account all factors affecting the vehicle’s dwell
time at the stop in question.

Another noteworthy limitation to the dataset was the lack of Automatic Fare Card (AFC) data
with which origin-destination matrices could have been estimated. While the LTC’s AFC system
is expected to be up and running in the near future, no data was available at the time of this
research. A methodology to estimate rough origin-destination matrices, borrowed from Navick &
Furth (1994) and adapted by Schwarcz (2004), was used instead.
7.3.3 Data-Imposed Limitations

The untrustworthy nature of the AVL data also led to two data-imposed limitations. Recorded values of passenger ons and offs were considered unreliable in cases where the AVL system failed to identify its location using the GPS. As a result, the majority of the database was discarded from analysis, weakening the findings of this thesis and reducing the value of the software to the LTC. Additionally, as a result of discarding data, all vehicles were recorded as dwelling at every stop along their trip pattern, and so the likelihood of dwell at any given stop was calculated as 100%. While the various models described in Chapter 4 were designed to take likelihood of dwell into account, this was irrelevant for the sake of this thesis’s LTC case study.

Finally, the success and accuracy of this software is limited by the quality of whichever supporting dataset is used.

7.4 Future Work

There are a number of avenues for expanding this research. First and foremost, acquiring and analyzing a dataset with reliable dwell time data would enable more accurate estimates of vehicle (and by extension, passenger) time savings/penalties caused by SR, SA, and limited-stop operation. Of course, this would require production of a new dwell time savings/penalty model, but doing so would greatly assist transit agencies with choosing appropriate stops for consolidation and addition.

In conjunction with modelling true (observed) dwell times, replacing the assumption of constant ridership with a model for predicting ridership changes as a result of stop layout changes would create a more flexible and meaningful software package. Ultimately, the goal of changing stop layouts is to improve operational efficiency and service quality, and in many cases attract new riders, which the current software does not address.

To improve the accuracy of calculating incremental access/egress distance changes due to SR or SA, local stop geography could likely be taken into account by leveraging the Google Maps API. Use of the API would supplant the need to assume that the transit agency’s metropolitan area is designed as a perfect grid.
Of course, any number of additional KPIs could be produced to evaluate the effects of SR and SA from new angles. Like the recommendations above, this could be completed in the near term.

In the medium term, this sort of stop-level analytics tool could be expanded to include route- and network-level analysis. For example, the effects of adding a new route parallel to an existing service could be simulated. Changes to ridership in another part of the network (perhaps following construction of a new large-scale development) could be modeled to predict the effect on other routes or sub-networks in the system.

Currently, the software produced for this thesis exists only in desktop format. However, a mobile application could be built to allow for small-scale analysis and/or presentation of previous analyses while on site, in meetings, or otherwise away from the transit agency’s office.

Finally, but crucially, in the long term, an ontology for ADC sensor equipment should be created in pursuit of establishing an industry-wide standard akin to the GTFS standard for schedule data. Such an ontology would pave the way for machine learning, easing the application of data mining, and would also significantly reduce the marginal effort required to clean data for additional transit agencies or for new ADC systems as they are developed. Given that the process of cleaning data still accounts for roughly 60-90% of time spent developing and implementing analytics tools, the industry has much to gain and little to lose from standardizing its formats.
References


Google. (2015). Retrieved May 22, 2015, from Google Maps: https://www.google.ca/maps/place/London,+ON/@45.3885624,-79.2449885,6z/data=!4m2!3m1!1s0x882ef20ea88d9b0b:0x28c7d7699a056b95


