MODELLING TRANSIT OPERATING COSTS

Peter (Yun Hao) Lai
Eric J. Miller, Ph.D.

August, 2020

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to express my sincere gratitude and appreciation to my supervisor, Professor Eric Miller, for providing me with the opportunity to work on this project and for his continued guidance and support throughout, and likewise to my colleagues at TMG - Luna Xi and James Vaughan - for all of the immense help provided along the way. I am also grateful to have this project funded by the Undergraduate Student Research Awards (USRA) from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Travel Modelling Group.

I. Introduction

This report documents the development of a transit operating cost model for transit agencies in the Greater Toronto & Hamilton Area (GTHA). This model estimates daily route-by-route operating costs based on the network service configuration. The final product of the project is a software tool module to extend the capabilities of GTAModel, the custom GTHA travel demand modelling framework developed by TMG that is currently in use by many municipalities in the region to assist with their transportation planning activities.

Cost modelling is an essential component of public transit operations and any transit planning endeavour, as financial considerations define the practical boundaries of design possibilities, inform the decision-making process, and root the project within the real world. The purpose of this project is to assist with transit service and network planning by providing a means to predict the economics of service configuration alternatives, from different route alignments to service frequencies, in order to better evaluate the sensitivity of various route and service design decisions.

Section II of this report presents the theoretical specification of the transit operating cost model. Section III describes the data that could be obtained to calibrate this model. In general, the data available to this project were rather limited, resulting in corresponding limitations in the models that could be developed from these data. Section IV discusses the model development, while Section V presents the model validation results. Section VI describes the software developed to implement the model in GTAModel. Finally, Section VII discusses a very preliminary attempt to develop a transit revenue estimation capability within GTAModel to complement the operating cost model developed in this project.

II. Transit Operating Cost Modelling

2.1 - Scope & Literature Review

A literature review was conducted in order to define the scope of the transit operating costs that would be estimated, explore existing methods to model those costs, and inform the model design.

Operating costs are typically associated with (Meyer and Miller 2013):

1. The day-to-day operation of transit vehicles, including operators' wages, vehicle fuel costs, etc.
2. The maintenance work done on the vehicles, including mechanics' wages, costs of replacement parts, tires, lubricants, etc.
3. The cost of running the various transit facilities, including wages for administrators and supervisors, costs of utilities of stations and garages, office supplies, and so on.

Some of these costs are dependent on service variables, which are associated with the extent of revenue service provided, including revenue hours, revenue kilometers, number of vehicles in service, passenger boardings, among others. For example, increased revenue hours would mean increased costs for operator wages. Other operation costs are fixed, meaning they do not change based on service modifications. For example, re-allocating buses belonging to the same garage from one route to another would have minimal or no impact on the cost of running the garage.

There are several approaches to modelling operating costs. From the literature review, the most popular model in the industry is the "cost allocation model", which uses a unit-cost approach to sum the total cost given the values of different service variables. Most models used in the literature and industry projects (Durham-Orange 2012) are small variations of this model with the same underlying principle (Bruun 2005; Meyer and Miller 2013; Miller and Rea 1973; Stopher et. al 1987). Other models found include the "resource build-up model" (Meyer and Miller 2013), which is also similar to the cost allocation model, and one paper that discussed modelling the costs and looking at economies of scale in transit operation using translog functions (Harmatuck 2005). It was deemed that the cost allocation model would be the most practical and suitable to this project, given the context in which this model would be implemented, as the traffic and transit assignment software (Emme) used by GTAModel only keeps track of limited transit service data. The rest of this section describes the model in detail.

2.2 -The Cost Allocation Model

The main principle of the cost-allocation model is to determine a unit cost for each service variable, multiply the unit cost by the quantity of its respective service variable, and sum the results across all service variables to obtain the total operating cost. This can be done on any scale, from a network-wide basis to a route-by-route basis, with the latter being the goal of this project. The idea of the cost-allocation model is represented in the following equation, with the service variables in this case being revenue hours (H), revenue kilometers (K), and number of vehicles in service during peak hours (V). The c_X variables are the unit costs associated with each service variable X:

$$Total\ Operating\ Cost = c_H \times H + c_K \times K + c_V \times V \quad [1]$$

Once the unit costs are determined, different values of the service variables can be substituted into the equation depending on the service modification being simulated, and the new total operating cost after this particular service modification could be determined (Bruun 2005; Miller and Rea 1973; Stopher et al. 1987).

2.3 – Determining Unit Costs

The basic steps to determine the unit costs (c_X) for each service variable X are as follows:

1. Identify the service variables of interest. Typically, the three key service variables are vehicle revenue hours (H), revenue kilometers (K), and number of vehicles in service during peak hours (V). There could be more, but these three are the most significant variables in transit service statistics and the data for them are the most widely available. In this project, these are the service variables used in the model.
2. Find existing reference values of the service variables for the agency (and route, potentially) of interest and find data on the operating expenses for the same agency and route during the same time period (typically on an annual basis), in the form of a list of individual cost items.
3. Assign each expense item to the best-fitting service variable that would serve as its primary driving factor. Each cost item must only be associated with one service variable, but each service variable could have multiple cost items associated with it. For example, the cost item of operator wages is more closely associated with the service variable of vehicle revenue

hours, and the cost of fuel is more fitting to be associated with revenue kilometers. Other costs, such as general administration, are assigned to peak vehicle count (Meyer and Miller 2013).

4. For each service variable X, sum the values for all cost items associated with it to obtain S_X, and then divide the result (S_X) by the value of the service variable X to obtain the unit cost (c_X). For the rest of the report, these unit cost values will be referred to as "Unit Cost Parameters" (UCP).

$$c_X = \frac{S_X}{X} \qquad [2]$$

Alternatively, if the source data of the service variables and costs are available for multiple time periods (typically years), linear regression can be used to determine c_X. A scatter plot of the values of X on the x-axis plotted against the allocated cost sums S_X on the y-axis for all years can be constructed. Then, a linear regression running through the origin can be fitted to the data, with resulting slope of the regression line defining the unit cost c_X (Miller and Rea 1973).

There is another method to estimate the UCPs of each service variable, without performing a cost allocation. It uses the technique of a multivariate linear regression that estimates the coefficients (equal to the UCPs) for each independent variable (Miller and Rea 1973). For each data point, only the values of the service variables and the total operating cost are required, which is useful for when a detailed expenses breakdown is not available, meaning a cost allocation is not possible. It is also useful for processing route-by-route data, where each route's service stats and cost would be a data point and data from all routes of the agency can be pooled for the regression. The limitation is that many data points are needed for an accurate and reasonable estimate of the coefficients. To implement this method in this study, the Python library *sklearn* was used.

All three of these methods were used to various extents to determine the UCPs used in the model, depending on the nature of the data obtained for each of the GTHA transit agencies.

2.4 – Adjusting for Inflation

When working with financial data spanning multiple years, it is important to account and adjust for inflation by converting all cost values into dollars associated with a constant reference year. The year 2016 was chosen, as the latest GTAModel network was based on 2016 demand and transit service data.

The Consumer Price Index (CPI) was used to perform the conversion. Statistics Canada (2020) publishes these indices on a monthly basis to indicate the change in the relative price of goods against a reference year (2002, with base value of 100), with a specific index for each region and category of goods. The cost value is converted to the target year by simply dividing it by the CPI of its original year and multiplying it by the CPI of the target year. The CPI from the month of October of each year was chosen as the representative annual CPI, as GTAModel networks have been built primary based on Fall service data. Ontario was chosen as the geographic location for the CPIs as that was the lowest geographic level for which the CPIs were available, at least for some categories.

For this study, the categories considered included:

- Operation of passenger vehicles
- Gasoline
- Passenger vehicle parts / maintenance / repairs
- Other vehicle operation expenses
- City bus and subway transportation

After some testing with the first set of data received from MiWay (see Section 3.2), it was determined that the CPI of category "city bus and subway transportation" led to the most accurate conversions. The conclusion was reached after converting obtained UCPs to 2016 dollars, multiplying them to the 2016 service stats, and comparing the estimated cost to the actual 2016 cost. Thus, the CPIs from this category, presented in Table 1, were used to convert all cost data used in this study.

Table 1: CPI from the month of October for years 2008 to 2019, under category "city bus and subway transportation" for Ontario.

Year	**CPI**	**Year**	**CPI**	**Year**	**CPI**
2019	163.8	**2015**	150.4	**2011**	133.3
2018	159.3	**2014**	145.3	**2010**	132.8
2017	159.0	**2013**	140.9	**2009**	121.7
2016	154.6	**2012**	138.0	**2008**	120.8

III. Data

3.1 – GTHA Transit Agencies & Data Search Process

The list of transit agencies whose networks are modelled in GTAModel (and hence whose data would be required to obtain unique UCPs for) is as follows:

- Brampton Transit
- Durham Region Transit
- GO Transit / UP Express (Metrolinx)
- Burlington Transit
- Milton Transit
- Oakville Transit
- Mississauga Transit (MiWay)
- Toronto Transit Commission (TTC)
- Hamilton Street Railway (HSR)
- York Region Transit (YRT)[1]

An initial search for publicly available operating cost data was conducted, in which transit agency websites and their financial reports were examined with the hope of finding transit operating cost breakdowns. Unfortunately, this was rather unsuccessful. While most agencies did publish agency-wide financial reports on an annual basis, the publicly accessible documents only contained very aggregate figures in terms of the operating cost, or only had a vague and basic breakdown into just a few sub-figures. This lack of detail did not allow proper cost allocations to be performed. An exception to this was Brampton Transit, which did publish detailed operating expense breakdowns into many sub-items, though

[1] In this report, 'YRT' includes both local routes as well as Viva routes.

the values were not actual recorded expenses, but rather budgets and full-year forecasts based on the spending at the half-year point, which was still not ideal. For big multimodal agencies like TTC and GO, there was also no mode-specific data which would be needed for accurate calibration.

In terms of route-by-route cost data that could be potentially used for a multivariable linear regression, it was even more scarce, with only the TTC publishing this data just for the years of 2011 and 2012, and only for its surface routes (i.e. bus and streetcar).

Given the lack of suitable publicly accessible sources, a formal request for transit operating cost data was sent to all agencies. The Canadian Urban Transit Association (CUTA) also publishes an annual Fact Book with agency-specific service and financial statistics that are very useful, and a request was made to obtain access to those data.

Due to the COVID-19 pandemic, all agencies in the region had been overwhelmed by extra work around modifying service levels and developing emergency plans due to reduced ridership, which made obtaining cost data from them more difficult, as understandably our request would not be treated as a priority. Nonetheless, in addition to the aforementioned public Brampton data, useful data were received from three of the largest agencies in the region: MiWay, TTC and YRT, as well as from CUTA, which had aggregate-level data for all agencies. The data are described in the following sections, along with the development of agency-specific UCPs from those data. Results of the requests sent to other agencies are summarized as follows:

- **Burlington:** No route-by-route data.
- **Durham:** Not allowed to release the data.
- **HSR:** Unable to allocate time to provide data.
- **Metrolinx:** Unable to provide data.
- **Milton:** Only CUTA data available.
- **Oakville:** Only CUTA data available.

3.2 – MiWay

The dataset received from MiWay was an Excel file containing several spreadsheets, which are outlined below along with the key relevant information from each sheet.

1. **Route-by-route operating statistics**: included daily revenue kilometres and daily revenue service hours. There were separate statistics for weekday, Saturday, and Sunday service. This was assumed to be 2019 data as “Signup / October 28, 2019” was marked on each sheet.
2. **MiWay CUTA Summary**: contained data from 2014 to 2019 published in the annual CUTA Fact Book. Key data were: total active vehicles, revenue vehicle kilometres, revenue vehicle hours, and total direct operating expenses. These were all annual data. This was useful because it had data from 2016, which was used to evaluate the cost model using 2016 GTAModel Emme data.
3. **MiWay CUTA Values**: contained total operating expenses each year from 2014 to 2019 *without overhead*, whereas the previous sheet presented the total operating expenses each year *with overhead*.

4. **2018 & 2019 MiWay UTS Form**: this was the most important part, as these sheets contained the detailed system-wide operating expense data broken into specific items, which was used for the construction of the cost model with the cost allocation method. These data were organized in five categories:

    a. Transportation operating expenses.
    b. Fuel and energy expenses for vehicles.
    c. Vehicle maintenance expenses.
    d. Premises and plant maintenance expenses.
    e. General and administration expenses.

    The previously mentioned “overhead” refers to the cost from the latter two categories, and so cost “without overhead” refers to the first three categories.

Using the data from the 2018 and 2019 “MiWay UTS Form” sheets, the cost model was constructed using the cost allocation method. This was done for each of the two years of data available and thus resulted in two sets of unit cost parameters. Operator-related costs (wages, etc.) were assigned to revenue hours, vehicle travel related costs (fuel, maintenance parts, etc.) were assigned to revenue kilometres, and vehicle-specific costs (insurance, etc.) and general costs (administration, etc.) were assigned to peak vehicles.

Then, two versions of the 2016 set of UCPs were estimated by applying inflation adjustment to each of the 2018 and 2019 UCPs. Both 2016-dollar sets were applied to the actual service variable values through the cost model to obtain estimates for the 2016 total operating cost, which were validated against the actual recorded 2016 operating cost, with the differences recorded as percentage error. It was found that the 2016 set obtained from the 2019 UCPs resulted in an error of just 0.08%, whereas the set based on 2018 UCPs had an error of 1.47%.[2] Therefore, the set obtained from the 2019 data was chosen as the final set of 2016 MiWay UCPs.

Table 2: MiWay unit cost parameters in 2016 dollars adjusted from 2018 and 2019 dollars, with the 2019 version chosen as the final 2016 MiWay UCPs.

Original Year	$ Per Revenue Hour	$ Per Revenue KM	$ Per Vehicle*
2018	$ 72.47	$ 1.41	$ 63,528.66
2019	$ 72.73	$ 1.38	$ 70,797.08

*Note: the unit cost per vehicle here represents the cost of owning / operating one vehicle *for the year*.

3.3 – TTC

While the TTC was unable to provide any custom data due to the pandemic, they did publish select operating statistics on their website. Out of those data, 2011 and 2012 route-by-route daily weekday

[2] Note that these errors assumed the UCPs were applied to the exact actual values of the service variables. In reality, the estimated service variables from Emme were a bit off from the actual values, leading to slightly bigger errors at the end.

service statistics (including rev. hr, rev. km, peak vehicle, and operating cost) were of great use in this study.

These datasets were prime candidates for multivariate linear regressions, as each year had 137 to 138 routes each with their own individual costs and service statistics, all serving as data points. To prepare the data, the costs were converted to 2016 dollars using CPI, and then both years were combined into one big set of 275 data points. Plotting each service variable against the total cost, two distinct trendlines were immediately visible. The points that form the line with the steeper slope were all streetcar routes and the points around the less steep line were all bus routes. This was a very good sign as it was to be expected that streetcar operation would have different – and higher – unit costs compared to buses.

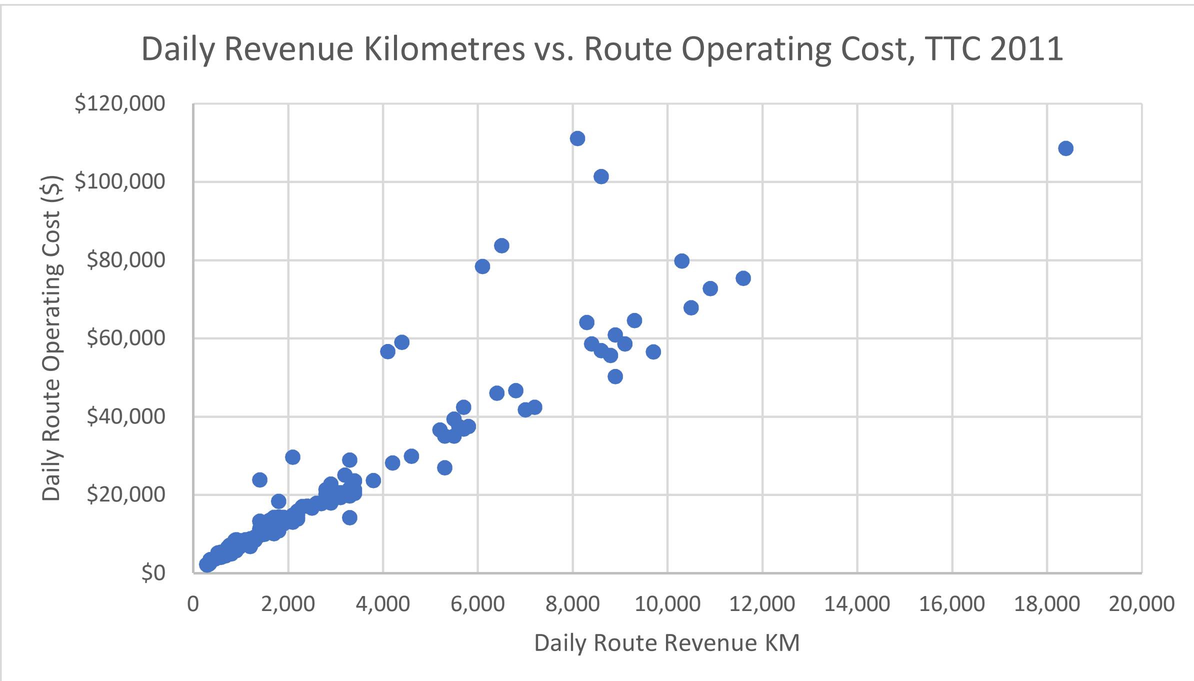

Figure 1: The scatter plot created by plotting the route-by-route daily revenue kilometres against the route operating costs, taken from the TTC 2011 dataset. Two distinct trendlines could be seen: the steeper one formed by the streetcar routes and the other by bus routes. The plots of the revenue hours and peak vehicles variables as well as the 2012 set exhibited similar behaviour.

Given this observation, the dataset was split into a streetcar set and a bus set, and multivariate linear regression was performed on each. Of course, the streetcar set was a lot smaller with only 16 data points (8 routes from each of the 2 years), but the regression still produced good results, as did the bus set. Below are the resulting UCPs for TTC bus and streetcar in 2016 dollars.

Table 3: Obtained 2016 sets of TTC bus and streetcar UCPs using multivariate linear regression.

Mode	$ Per Revenue Hour	$ Per Revenue KM	$ Per Vehicle Per Day*
Bus	$ 105.05	$ 1.90	$ 250.06
Streetcar	$ 84.07	$ 4.02	$ 1155.97

*: Since this set contained daily values to begin with, the resulting unit cost per vehicle was also on a daily basis instead of on an annual basis like with other agencies. The conversion between daily and annual is discussed in Section 4.7.

3.4 – YRT

YRT provided route-by-route cost data as well. The data were based on January 2020 service levels. Unfortunately, revenue hours information was missing, so a multivariate linear regression could not be performed. Nevertheless, it was still valuable for validating the route-by-route cost estimations generated based on CUTA data. To prepare the data, the cost figures were converted to 2016 dollars. It was also noticed that these costs did not include overhead costs, while the cost model being developed did include overhead. To mitigate this, a scaling factor between total cost and cost without overhead was calculated with a value of around 1.31 that allowed the overall operating costs on a route-by-route basis to be estimated.

3.5 – CUTA

CUTA *Fact Book* data for Ontario transit agencies had been provided by CUTA to MTO and were subsequently published online by the Ontario Legislature. The CUTA data contained a considerable amount of information from service statistics to ridership and operating expenses, for each of the ten agencies in the GTHA, for years 2011 to 2018, inclusive. This set of data became the primary source for the development of the final set of agency-specific UCPs. Key CUTA data fields that pertained to this project are:

1. Modal statistics: total annual revenue hours and revenue kilometres for each agency, broken down by mode.
2. Vehicles: the estimated peak vehicle count for each agency, again broken down by mode.
3. Operating expenses: the annual operating expenses, broken down into five categories:
    a. Transportation operations expenses.
    b. Fuel/energy expenses for vehicles.
    c. Vehicle maintenance expenses.
    d. Plant maintenance expenses.
    e. General/administration expenses
4. Operating revenues and ridership.

CUTA also directly provided the project with agency-specific historic data from 2006 to 2010. Unfortunately, they were unable to provide the peak vehicle counts for these years, as the vehicle information had been collected differently prior to 2011 and could not be converted to the current format. The revenue hours and kilometres information were also missing for all agencies for 2006 and 2007, making data from those two years effectively void for the project, since the independent variables (i.e. service variables) were missing, barring any sort of estimation of unit costs using regression.

Access to this full, multi-year set of service statistics for all GTHA agencies was crucial to developing a consistent operating cost model for the region. The aggregate operating expense categories, however, meant that only a very basic cost allocation to service variables could be done. This did negatively impact accuracy, compounded by the fact that these category sums varied wildly year to year for some agencies.

The first step in processing the CUTA data was to convert them to 2016 dollars. All three methods from Section 2.3 were attempted to obtain UCP sets for each agency. The first two – direct cost allocation and linear regression – required the assignment of cost items to one of the three service variables. As mentioned, there were only five cost items to assign, so unfortunately the processing could only be done in a coarse manner as opposed to the MiWay case, where almost 30 cost items were available for cost allocation. The cost allocations for the CUTA data were:

- Revenue hours: Transportation operations expenses.
- Revenue kilometres:
    - Fuel/energy expenses.
    - Vehicle maintenance expenses.
- Peak vehicle count:
    - Plant maintenance expenses.
    - General/administration expenses.

With the first method of direct cost allocation, 2016 costs allocated to each service variable were simply divided by their service variable values to obtain the 2016 set of UCPs. These are referred to as the "CUTA 2016" set.

For the linear regression method, each cost category was regressed against its allocated service variable for each year from 2008 to 2018 (with a zero intercept for the regression line). Note for that for 2008 to 2010, the peak vehicle information was not available, so for the regression of this variable, the data points were restricted to years 2011 to 2018. The final set of UCPs obtained from this method are referred to as the "CUTA LinReg" set.

The third method, multivariate linear regression, was also attempted. Unfortunately, because of how drastic some of the cost figures varied from year to year, and the fact that there were only eight data points (pre-2011 data could not be used due to lack of peak vehicle data), this was not a success. A lot of the UCPs produced (using the Python *sklearn* library) were even negative, which did not make sense given the context behind the cost model. Thus, nothing generated using this method was used.

Below is a table of the UCPs generated from the CUTA data using the first two methods for each agency. The accuracy of these sets is discussed in Section V.

Table 4: CUTA UCPs for each agency, generated using both direct cost allocation on 2016 data as well as linear regression on 2008-2018 data.

Agency	CUTA 2016			CUTA Lin Reg		
	UCP Rev Hr	UCP Rev KM	UCP Vehicle	UCP Rev Hr	UCP Rev KM	UCP Vehicle
Brampton	$72.82	$1.55	$37,555.16	$74.00	$1.61	$40,251.42
Durham	$73.51	$1.37	$79,242.71	$78.04	$1.53	$82,689.64
GO Transit	$136.97	$3.92	$1,053,386.08	$80.21	$3.15	$951,935.41
Burlington	$51.97	$1.16	$45,397.47	$53.69	$1.40	$45,897.65
Milton	$86.22	$0.42	$22,559.33	$86.01	$0.75	$25,570.67
Oakville	$67.27	$1.54	$41,800.55	$64.11	$1.58	$38,893.09
MiWay	$72.73	$1.38	$70,797.08	$75.43	$1.43	$81,347.43
TTC	$64.45	$2.13	$255,335.80	$66.83	$2.24	$252,416.67
Hamilton	$62.05	$1.83	$48,870.41	$65.47	$1.96	$46,114.60
YRT	$109.84	$0.46	$60,845.69	$114.24	$0.52	$65,201.45

Based on data from: Canadian Urban Transit Association, Ontario Urban Transit Fact Book, 2011 to 2018.

IV. Cost Computation Procedure

With the cost model defined and the UCPs obtained from data, this section presents the procedure to actually implement this model and produce the route-by-route cost estimations, given a service configuration (network and schedules). This procedure was implemented in the final software tool module that automates the entire process, as described in Section VI. The general idea is that the service variable values from a GTAModel network assignment need to be obtained, which could then be used in the cost model by multiplying the UCPs to them to determine the cost. The challenge was to develop a means to calculate these service variable values that could operate within the limits of Emme, the software used by GTAModel to process the network and demand information.

4.1 – Working with Emme

Ultimately, the cost model is applied each time the module is run, which will output operating costs for each individual route, for each of the five time periods of the day (AM, MD, PM, EV, ON) that are modelled by GTAModel. The issue was that Emme only tracked service statistical information on a time period basis, meaning it did not track individual transit vehicle trips.[3] This made calculating the three service variables within Emme difficult. However, the transit information in Emme came from the transit service table file, which was part of the existing input data within GTAModel. Thus, this service table was used to generate trip-specific information for each transit line that enabled the eventual calculation of the revenue hours, revenue kilometres, and peak vehicle count for each line.

Another quirk with Emme was that each branch and direction of an overall transit route was encoded as a separate line. For example, MiWay Route 7 ('M007') only had one branch (the main branch), but in

[3] In this report, a "trip" is the movement of a transit vehicle from one end of its route to the other, not the movement of a transit rider.

Emme the route was broken up into two lines called 'M007Aa' and 'M007Bb' to represent its two directions of service. For the purposes of this entire cost modelling project, all statistics and costs computed were in terms of each overall *route*, meaning that the individual *lines* within the route (directions and branches) as encoded in Emme were addressed in an aggregate fashion by *grouping them into routes*. The reason for this was that the finest level of detail in all the available cost data from transit agencies was at the route level and did not break down further into branch and direction level. Moreover, the estimation of the number of vehicles deployed on each route required the consideration of both directions at once as one cycle.

Therefore, the three service variable values of revenue hour, revenue kilometre, and vehicle count were all computed at the route level, for each of the five time periods. The calculation methods are summarized in the following sub-sections.

4.2 – Preliminary Calculations

To calculate the service variable values, the transit service table file was used, which contained every single scheduled transit trip along with each trip's start and end times. This file is produced by the TMG network modellers with each new network update. With this, the scheduled runtime, average headway, and the trip count for *each line* in Emme within each time period could be computed. The trip count is defined as the number of trips on the line that are scheduled to start within the time period. These calculations provided the necessary data for subsequent calculations.

4.3 - Revenue Hours

The revenue hours were computed for each line and then summed across all lines of the route to obtain a value for the overall route. For each line, revenue hours are the trip runtime multiplied by the trip count. The trip runtime value used was either the scheduled runtime obtained from the service table or the actual runtime obtained from the Emme transit assignment, whichever was larger. Sometimes, the actual runtime from the assignment was lower than the scheduled runtime, but this is not practical especially with suburban bus routes as it is not desirable for buses to be early. Thus, when this occurred, the scheduled runtime was used. On the other hand, if the actual runtime was higher than the scheduled runtime, it was reasonable to use the actual runtime as buses are often late due to traffic congestion and it is the actual runtime that gets counted in transit agencies' service statistics. Note that for the ON (overnight) time period, since there was no transit assignment done, scheduled runtimes were used for all lines in this time period.

$$Route\ revenue\ hours = \sum_{All\ lines\ of\ route} [\ \max(runtime_{scheduled}, runtime_{actual}) \times trip\ count\ ] \quad \text{.... [3]}$$

4.4 - Revenue Kilometres

The revenue kilometres were first computed for each line by multiplying the line length by the trip count, and then summed across all lines of the route.

$$Route\ revenue\ km = \sum_{All\ lines\ of\ route}[line\ length \times trip\ count] \quad [4]$$

4.5 – Peak Vehicle Count

The standard formula for computing the number of vehicles, N, required to deliver a given service headway, considering both directions of the route as one cycle is (Walker 2011):

$$N = ceiling\left(\frac{cycle\ runtime \times layover\ factor}{headway}\right), \text{ where } cycle\ runtime = runtime_{dir1} + runtime_{dir} \quad \text{…. [5]}$$

In this case, estimating the number of vehicles deployed for each route was a tricky process, as Emme did not store this information. Moreover, preliminary explorations found that simply applying equation [5] on an individual line/direction basis and summing the integer quotients to get the figure for the overall route tended to overestimate the vehicle count, especially for low frequency routes. This was due to the presence of the ceiling function,[4] and the fact that most routes operate on a cycle basis, meaning the vehicles usually turn around at the terminus after just a few minutes of layover time. The ceiling function rounds up the decimal quotient, but it does so twice for both directions of the route if the equation was applied to each direction separately, leading to the overestimation. Thus, both directions needed to be considered as one cycle in the calculation. However, Emme tracked directions as separate lines, which were scrambled with the other branches and directions of the same route in no apparent order. Multiple short-turn branches also existed, even though they used the same vehicles as the main branches but were still encoded as separate lines.

To address this, an algorithm was developed to filter out the short-turn lines, determine possible line pairings for each route, and identify the best pairing combination, i.e. it determines which two lines of the route are most likely opposing directions of the same branch, and does this for all branches of the route. This enabled the calculation of vehicle count for each pair of the route using the above formula by treating the pair as one cycle, and the overall vehicle count for the route is thus the sum of vehicle counts across all pairs of the same route.

Development, prototyping, testing, and optimization of the vehicle count estimation algorithm took considerable effort. TTC bus routes were initially tested and evaluated for accuracy against online databases and the original GTFS data (discussed further in Section 5.2), and then gradually the procedure was expanded to all other agencies in the region. Within the TTC bus routes alone, special cases, like the ‘F’ express branches of high-ridership routes that seemed like duplicates of their local branches, as well as the 141-145 downtown express routes, needed to be explicitly handled. In terms of the other agencies, special cases included GO Transit, where many of their routes run in one direction only during peak

[4] The ceiling function returns the value of the input rounded *up* to the nearest integer.

hours, eliminating the need for constructing two-directional pairs. The same applied to Oakville and Milton Transit, where there was no direction information encoded in the Emme network.[5]

Improvements and optimizations to the procedure itself were also gradually added, such as creating a recursive pairing function to generate possible direction pairing combinations, and making use of the starting and ending node IDs of each line to compare their geographical alignments during the evaluation of different alternative pairing combinations. Furthermore, it was discovered later that the 2011 network used a different direction encoding system where some lines were actually dual direction, which had to be explicitly addressed as well.

Over the course of repeated testing and iterating, an optimized set of constant values used in the algorithm was produced, containing the threshold values for deciding when a line is considered too infrequent to be included, and the layover factor optimized for each agency, etc. The values for the latter were also found through analyzing the original GTFS schedule data for each agency, where a script was coded to compute the typical layover times scheduled.

4.6 – Calculating Time-Period-Based & Daily Operating Costs

With the route-by-route service variable values determined for each time period, the operating cost for each route could be simply calculated by multiplying them with the unit cost parameters. Note that the unit costs per vehicle shown in Tables 2 and 4 represent the cost of owning and operating one vehicle *for the year*, and they assumed a constant peak vehicle count for the whole year. However, the service variables used were only for the duration of *one time period in one day*, which are only a few hours. Thus, it was necessary to first convert the unit cost per vehicle parameter from an annual basis to a daily time-period basis as follows:

$$UCP_{vehic\quad time\ period} = \frac{UCP_{vehicle_{annual}}}{weekday\ ratio \times 24} \times duration\ of\ time\ period \qquad [6]$$

Note that instead of dividing the annual figure by 365 days, the “weekday ratio” was used in the equation instead. This is discussed in the next section.

In addition, with the TTC UCPs obtained from the multivariate linear regression applied on the daily data, the UCP per vehicle for the time period was calculated as the daily UCP divided by 24 and multiplied by the duration of the time period.

With the operating cost per route determined for each time period, a summation across all routes gave the total operating cost for the time period. The same was done for revenue hours and revenue kilometres. For the *daily* revenue hours and revenue kilometres, they were obtained by performing a summation across all time periods. However, it was noted that to get the total daily operating cost, one *should not* simply add the costs from all time periods together. The reason for this was that the peak vehicles for the entire day is neither a sum of the vehicle count from all the time periods nor a weighted average, but rather it is simply the largest vehicle count from all time periods, and only one time period has that vehicle count. The costs from other time periods were calculated using lower vehicle counts, so the sum of costs from all time

[5] Same procedure as with GO: treated every line as single-direction where no pairing was needed, and simply aggregated data per line into the overall route.

periods would be lower than if one obtained the cost using values on an aggregated daily level. Therefore, the correct approach to getting the daily cost was to multiply the unit cost parameters by the daily service variable values directly, of which peak vehicles was taken from its largest value of all time periods. In addition, similar to before, a conversion for the peak vehicle UCP was performed, this time by dividing the annual version of the parameter by the "weekday ratio" to get the daily version.

4.7 – Estimating Annual Values: The Weekday-Annual Ratio

With the daily operating cost determined, the annual operating cost could be estimated, which was necessary for model validation. Like the daily cost, the annual service variable values were first computed, and then multiplied with the UCPs to obtain the annual cost. A complication was that the calculated daily service values applied to weekdays only. On the weekend, service levels are usually drastically reduced.

The crude method of estimating the annual values would be to simply multiply the daily values by 365, but that would lead to overestimations as it assumes every day of the year is a weekday, which is not true. The other method is to determine an agency-specific ratio of daily weekday service variable values to the total annual service variable values from existing schedule data. Since there are 251 weekdays in a year, this ratio is expected to be between 251 and 365 due to the lower (but not non-existent) weekend service levels.

To determine this weekday-annual ratio (abbreviated as weekday ratio) for each agency, both weekday and weekend revenue hours and revenue kilometres were computed from GTFS data. Weekly totals and thus annual totals were estimated, which enabled the calculation of the weekday to annual ratios. Overall, this ratio averaged to be around 312 across all the agencies, with TTC's being a bit higher at 330 and Milton's being the lowest at 289.

Table 5: Estimated weekday-annual ratios by agency.

Agency	Weekday Ratio	Agency	Weekday Ratio
Brampton	305.83	**Oakville**	312.48
Durham	301.96	**MiWay**	297.22
GO Transit	318.56	**TTC**	330.25
Burlington	301.02	**Hamilton**	320.80
Milton	289.08	**YRT**	302.77

With this ratio, annual revenue hours and revenue kilometres were then estimated, by multiplying the daily values with their respective weekday ratios. The annual peak vehicle count was the same as the daily value, as it simply represented the highest number of vehicles in service at any one point. Finally, the annual cost estimates were obtained by applying the UCPs to the newly estimated annual service variable values.

In [6], the weekday ratio was applied to the conversion of the peak vehicle UCPs between annual and daily bases as this was more accurate. In other words, instead of multiplying or dividing by 365 to convert between annual and daily UCPs, the weekday ratio was used instead.

V. Model & Data Validation

With the annual costs estimated using the cost model developed in the previous sections, validation was performed by comparing the estimated annual operating costs per agency with the actual recorded 2016 costs obtained through the data received. The same method was used for validation of the service variable values. Through repeated testing and refinement, an optimized set of agency-specific UCPs were produced that yielded the most accurate results.

5.1 – Validating Revenue Hours and Revenue Kilometres Estimations

Upon comparing the estimated revenue hours and revenue kilometres from Emme data against the actual 2016 numbers found in the CUTA data, it was found that Emme tended to underestimate the revenue hours and revenue kilometres, though not by a significant amount, within about 5% off for most agencies. As these values were dependent on the results of the static transit assignment in Emme, which couldn't be modified (or at least would involve sizeable work outside the scope of this project), it was easier to simply correct the differences by applying a scaling (adjustment) factor unique to each agency in order to bring the estimated values more in line with the actual values. Doing so has ultimately benefited the accuracy of the cost estimation. Below is a table of the adjustment factors for each agency for 2016.

Table 6: List of 2016 adjustment factors by agency and mode after validation of revenue hours and revenue kilometres.

Agency	Mode	Adj. Rev Hr	Adj. Rev KM
Brampton	Bus	1.0313	0.9951
Durham	Bus	0.9855	1.0157
GO Transit	Coach Bus	1.2863	0.9943
GO Transit	Commuter Rail	3.7580	1.0590
Burlington	Bus	0.9222	1.0449
Milton	Bus	1.1271	1.1174
Oakville	Bus	0.9276	0.9607
MiWay	Bus	1.0071	1.0036
TTC	Bus	1.0348	1.0498
TTC	Streetcar	0.8691	0.9286
TTC	Subway	1.0540	0.9785
Hamilton	Bus	1.1503	1.0173
YRT	Bus	1.0503	1.1324

Additional to these adjustment factors, an extra step had to be performed for TTC subway to account for the number of cars per train. The reported TTC data was on a per train-car basis, whereas the Emme calculations were on a per train-set basis. Thus, Emme revenue hours and kilometres for Line 1 and Line 2 needed to be multiplied by 6 (because they use 6-car trains), and Lines 3 and 4 values by 4 (because they use 4 car trains).

5.2 – Validating Peak Vehicle Count Estimations

To validate the estimated number of vehicles deployed per route during each time period, a set of route-specific calibration data were needed for the particular year of 2016, but this information proved to be very difficult to find. Out of all agencies, only the TTC officially published this vehicle information and only for a few years that did not include 2016. Thus, other sources were sought out.

The first source investigated was the wiki site of the CPTDB (Canadian Public Transit Discussion Board). It is an unofficial source maintained by transit enthusiasts that contains a very detailed per-agency database with regards to fleet and route information, including the number of vehicles used on each route for bigger agencies like the TTC. Since it was a wiki, all historic versions of the pages were stored and openly accessible, enabling Fall 2016 route vehicle information to be obtained. Using these data as a calibration set, validation, iteration, and optimization of the direction pairing and vehicle count estimation algorithm discussed in Section 4.5 were performed for all TTC routes. Unfortunately, CPTDB did not have similarly detailed vehicle information for other agencies, and it was not always accurate for the TTC either, which led to a search for a more reliable and robust data source.

It was discovered that GTFS data included a variable called the "block_id", which is a unique identifier that agencies assign to each of their vehicles to identify which trips are operated by the same vehicle in their schedule data. By examining the number of unique block_ids assigned to trips on a particular route in a time period, a good estimate of the number of vehicles required to be deployed on the route at the same time could be obtained. Of course, for interlined routes, this would not be accurate. When the TTC route vehicle data pulled from GTFS were compared to the CPTDB data, the results were largely consistent, except for a few interlined routes. The advantage of using this "block_id" method was that all agencies publish GTFS data, so calibration sets for route vehicle deployment for other agencies beyond the TTC could be constructed, which the CPTDB database lacked.

Furthermore, in addition to the data from GTFS, the total peak vehicle counts per agency from the CUTA data were also helpful as another point of reference when calibrating the vehicle estimation procedure.

5.3 – Validating Cost Estimations

The majority of the cost data available, and, hence, modelling undertaken in this study, were only available at the annual, system-wide level, and so most of the validation of the model necessarily occurred at this level. The one route-by-route validation that was possible was for YRT, for which the required observed route-level data were available. The results were satisfactory after adjusting for the overhead costs and inflation (as discussed in Section 3.4), with each route's estimated cost and the corresponding YRT-given cost being fairly similar with a median error of 12%. A scatter plot of the observed vs. estimated costs for each route is shown in Figure 2. The differences were acceptable as the observed data were based on January 2020 service levels and the data used to generate the UCPs were based on Fall 2016. Several routes had also undergone significant alignment changes during this period, leading to large error magnitudes, with the more extreme ones removed from the plot as outliers. Overall, given the < 1 slope of the trendline, it seems that the cost model tended to overestimate the cost, though the fit is decent as evident from the R^2 value.

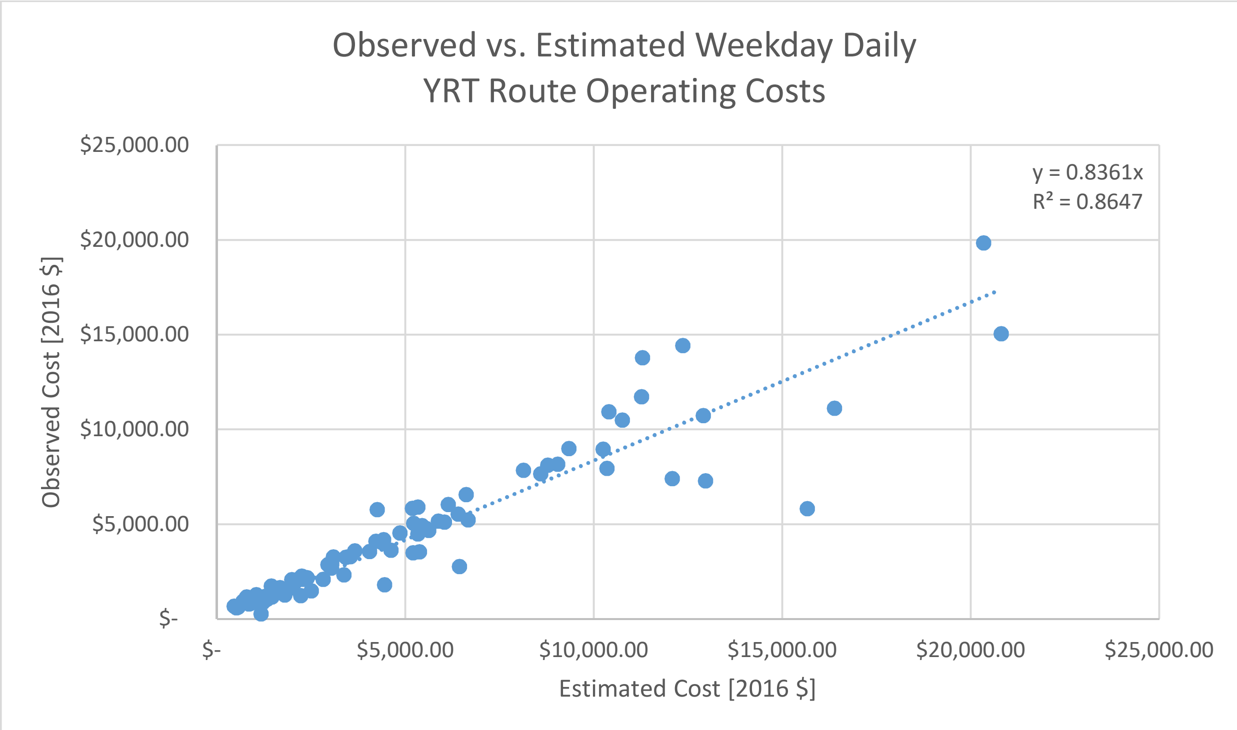

Figure 2: Observed vs. estimated weekday daily YRT route operating costs. The estimated costs were based on Fall 2016 service levels whereas the observed costs were based on January 2020 service levels. Both are shown in 2016 dollars.

For the annual cost validation, four different estimates were generated for each agency, mainly using four sets of UCPs:

- CUTA 2016 UCPs, with no adjustment factor on service variable estimations
- CUTA LinReg UCPs, with no adjustment factor
- CUTA 2016 UCPs, with adjustment factors applied
- CUTA LinReg UCPs, with adjustment factors applied

In addition, in the case of MiWay and the TTC bus and streetcar routes, the UCPs generated from the direct agency data were also considered in the validation in addition to the above CUTA sets.

After rounds of testing and refinement, a final set of optimized UCPs for each agency was produced, by taking the set that yielded the most accurate result when compared to actual 2016 numbers. Below is a summary of the results for 2016:

Table 7: Final optimized 2016 set of UCPs by agency and mode as well as comparison between estimated and actual annual costs with % error, all in 2016 dollars.

Agency / Mode	$ / Rev Hr	$ / Rev KM	$ / Veh*	Estimated Annual Cost	Actual Annual Cost	% Error
Brampton	$72.82	$1.55	$37,555.16	$127,709,146.19	$128,385,139.00	**-0.53%**
Durham	$73.51	$1.37	$79,242.71	$66,366,641.30	$66,445,884.00	**-0.12%**
GO	$136.97	$3.92	$1,053,386.08	$842,668,428.55	$823,990,986.00	**2.27%**
Burlington	$51.97	$1.16	$45,397.47	$15,547,099.33	$14,970,409.00	**3.85%**
Milton	$86.22	$0.42	$22,559.33	$3,867,018.00	$3,867,018.00	**0.00%**
Oakville	$64.11	$1.58	$38,893.09	$24,765,713.59	$24,616,745.00	**0.61%**
MiWay	$72.73	$1.38	$70,797.08	$174,542,382.00	$174,676,376.00	**-0.08%**
TTC Bus	$105.05	$1.90	$250.06 *			
TTC Stc	$84.07	$4.02	$1,155.97 *	$1,713,930,172.22	$1,659,757,328.00	**3.26%**
TTC Sbwy	$64.45	$2.13	$255,335.80			
HSR	$62.05	$1.83	$48,870.41	$86,793,167.45	$87,613,749.00	**-0.94%**
YRT	$109.84	$0.46	$60,845.69	$172,375,637.78	$175,014,016.00	**-1.51%**

*: For the TTC bus and streetcar UCPs only, the UCP per vehicle is on a daily basis while it is annual for the rest of the agencies.

Actual cost data source: Canadian Urban Transit Association, Ontario Urban Transit Fact Book – 2016 Operating Data, RTS-17-20, October 2017.

Additionally, the same validation process was completed for the 2011 network as well, using the 2011 service table and the corresponding 2011 CUTA cost allocation UCPs (plus MiWay's and TTC's own UCPs), all in 2011 dollars.

Table 8: Final optimized 2011 set of UCPs by agency and mode as well as comparison between estimated and actual annual costs with % error, all in 2011 dollars.

Agency / Mode	$ / Rev Hr	$ / Rev KM	$ / Veh*	Estimated Annual Cost	Actual Annual Cost	% Error
Brampton	$63.80	$1.39	$34,705.78	$83,293,164.21	$82,334,298.00	**1.16%**
Durham	$67.29	$1.32	$71,297.08	$51,440,939.25	$51,900,348.00	**-0.89%**
GO	$50.47	$2.25	$554,699.53	$488,393,781.59	$466,726,810.00	**4.64%**
Burlington	$46.30	$1.21	$39,574.11	$13,166,002.74	$13,025,455.00	**1.08%**
Milton	$93.82	$0.05	$33,636.46	$2,317,878.33	$2,301,678.00	**0.70%**
Oakville	$50.98	$1.43	$31,956.54	$19,216,663.43	$19,538,199.00	**-1.65%**
MiWay	$65.04	$1.23	$70,139.79	$145,641,898.95	$148,712,919.00	**-2.07%**
TTC Bus	$90.58	$1.64	$215.61 *			
TTC Stc	$72.49	$3.47	$996.71 *	$1,380,157,306.35	$1,419,892,261.00	**-2.80%**
TTC Sbwy	$59.92	$2.09	$231,794.71			
HSR	$56.45	$1.69	$39,761.17	$68,904,680.11	$69,739,170.00	**-1.20%**
YRT	$98.50	$0.44	$56,218.33	$135,248,202.78	$134,923,249.00	**0.24%**

*: For the TTC bus and streetcar UCPs only, the UCP per vehicle is on a daily basis while it is annual for the rest of the agencies.

Actual cost data source: Canadian Urban Transit Association, Ontario Urban Transit Fact Book – 2011 Operating Data, RTS-12-20, November 2012.

These two optimized sets for 2016 and 2011 would be used as the cost parameters input each time the final software tool module is run, depending on which of the two networks the tool is working with.

5.4 – Model & Data Limitations

While the final errors of the optimized cost estimations were small, the UCPs are difficult to extend to other years beyond 2016 and 2011 and for future predictions. This is mainly due to the lack of detailed cost data for calibration. The majority of the UCPs came from the aggregate CUTA data that only had five cost items available for allocation, which also varied unpredictably between years even after inflation adjustment, caused by its aggregate nature. The UCPs were either obtained through direct cost allocation done on the target year only, or through a linear regression of multiple years after allocation was done for each. The former is only representative of the target year and would be largely inaccurate if applied to other years due to the inconsistency of costs between years. While the latter considers the variations between years through a linear regression, the fit has been mediocre and even worse for some agencies.

With regards to other data received, while the TTC route-by-route data was very effective with the multivariate linear regression, it was from 2011 and 2012, almost a decade ago. For future predictions, this would become more inaccurate as the unit costs have likely changed by now, as the TTC have since expanded and modified their network and upgraded their fleet. Perhaps the most robust dataset received was the MiWay data, which was based on 2018 and 2019 data, and had a detailed cost breakdown for performing a proper cost allocation. This sets a solid foundation for MiWay UCP predictions with future years.

Another limitation is the insufficient modal-specific data received. For the TTC, there were at least data for buses and streetcars in 2011 and 2012. For GO Transit, however, attempts at obtaining data from them were unsuccessful, meaning the UCPs are identical for both the bus and commuter rail modes. This is unrealistic as the operation of the two modes should involve drastically different unit costs, so the current UCPs are overestimating the costs for bus routes and underestimating costs for commuter rail lines. Hopefully at some point in the future, the GO Transit modal-specific data could be obtained, ideally with a breakdown of annual costs by mode as well as route-by-route costs. This would produce much more accurate estimations for GO.

Furthermore, the cost allocation model itself is still a fairly simplistic model, being purely linear. In actuality, as the service levels change, the operating cost would likely change in a non-linear fashion in order to properly account for fixed costs. Thus, this current model would be the most suitable to be applied to relatively small service configuration changes where the linear nature would still provide a good approximation. This should cover the majority of the use cases with this model in the context of TMG.

VI. Emme Tool Creation

6.1 – TMG Toolbox / Python Development

The entire cost estimation procedure was originally prototyped and iteratively developed as a standalone Python script that took in the Emme transit line data from a manually prepared CSV file and ran for one agency and one time period at a time for easier modifications. Eventually, this script was cleaned and optimized into a more polished version so that it was able to process all agencies at once for each time period, requiring just one input containing transit line information for the entire network, in addition to the transit service table CSV. A "cost parameters" input CSV file was also created that contains all UCPs and adjustment factors for all agencies, which the script would parse and apply the right UCPs to the corresponding routes automatically. This allowed for rapid testing of different sets of UCPs, as all the cost information was contained in one file. The script would then output a CSV containing route-by-route revenue hours, kilometres, vehicles, and operating cost estimates for the specified time period.

As a final step, the polished standalone script was adapted to be fully compliant as a proper Emme Modeller tool according to the Emme API specifications. Logging and error handling procedures that are compliant with the Emme API were added as well. The finished form of the transit operating cost estimation tool now runs within Emme Modeller and interacts directly with Emme by pulling the line information from the specified scenario, which the user can freely select as long as they are in the same project. It is a fully tested and approved addition to the TMG Emme Toolbox. There are three other required inputs:

- Transit service table file: the same file found in the GTAModel input folder in CSV format.
- Cost parameters file: the CSV file according to specifications that contains all the UCPs and adjustment factors for all agencies.
- Report file path: the path to which the output file is to be saved.

For a more detailed description, please see the software documentation for the tool. The source code is available as part of the TMG Toolbox package on GitHub.

6.2 – XTMF Module / C# Development

An XTMF module was also developed that takes the inputs during a GTAModel run and automatically invokes the Python TMG Toolbox tool to run the cost estimation. This module, like the rest of XTMF, is written in C#. The code is on GitHub.

VII. Revenue Modelling

In addition to modelling operating costs, operating revenues should also be estimated, so that subsidies can be computed after subtracting the revenues from the total operating cost. To build a proper revenue model, there needs to be accurate information regarding the transit ridership of different demographics and age groups, as they each pay a different fare. While this was present in the CUTA data, the issue lied with the fact that the current transit assignment in Emme was not capable of estimating these numbers. Thus, instead of accounting for the different fares for different demographics, a simple "average fare per passenger" value was used. This was obtained by dividing the observed total revenue by the total ridership for each agency. The total revenue could then be simply estimated by multiplying the estimated ridership by the average fare.

It is possible to extract daily ridership estimates for each transit agency from GTAModel. Using the weekday to annual ratios developed for the cost model, GTAModel estimates of annual 2016 ridership by agency was compared with CUTA reported data, as shown in Table 9. As seen from this table, the outcome was mixed: some agencies had quite accurate estimates, notably MiWay, Hamilton, YRT, and GO, but others did not, especially the three Halton agencies.

Table 9: Estimated vs. actual annual ridership per agency in 2016.

Agency	Estimated Annual	Actual Annual	% Error
Brampton	20,387,047	23,129,596	-11.86%
Durham	12,100,601	10,192,324	18.72%
GO Transit	70,905,158	69,193,574	2.47%
Burlington	2,551,501	1,900,094	34.28%
Milton	639,058	455,246	40.38%
Oakville	3,899,681	2,851,368	36.77%
MiWay	39,056,737	38,597,356	1.19%
TTC	498,741,494	538,078,996	-7.31%
Hamilton	22,516,335	21,495,758	4.75%
YRT	23,078,716	22,414,974	2.96%

Actual ridership data source: Canadian Urban Transit Association, Ontario Urban Transit Fact Book – 2016 Operating Data, RTS-17-20, October 2017.

One way to mitigate this result was to apply an adjustment factor to bring the values closer to the actual value. Applying these adjustment factors to 2011 ridership estimates and comparing them to the actual 2011 figures yields Table 10. Use of the adjustment factors generally reduced the modelling error, except in the case of Brampton and Durham. More work is clearly required to develop a more robust revenue estimation model for the GTHA.

Table 10: Estimated vs. actual annual ridership % errors per agency in 2011, before and after applying the 2016 adjustment factors.

Agency	% Error Before Adj.	% Error After Adj.
Brampton	7.74%	22.24%
Durham	-3.21%	-18.47%
GO Transit	26.17%	23.13%
Burlington	50.64%	12.18%
Milton	83.76%	30.90%
Oakville	40.41%	2.66%
MiWay	7.78%	6.51%
TTC	-6.79%	0.57%
Hamilton	5.79%	1.00%
YRT	12.87%	9.62%

Actual ridership data source: Canadian Urban Transit Association, Ontario Urban Transit Fact Book – 2011 Operating Data, RTS-12-20, November 2012.

References

Bruun, Eric. 2005. "Bus Rapid Transit and Light Rail." *Transportation Research Record: Journal of the Transportation Research Board* 1927 (1): 11–21. doi:10.1177/0361198105192700102.

Canadian Public Transit Discussion Board Wiki. 2020. "Toronto Transit Commission." *Toronto Transit Commission - CPTDB Wiki*. https://cptdb.ca/wiki/index.php/Toronto_Transit_Commission.

Canadian Urban Transit Association. 2012. "Ontario Urban Transit Fact Book – 2011 Operating Data. RTS-12-20. November 2012."

Canadian Urban Transit Association. 2017. "Ontario Urban Transit Fact Book – 2016 Operating Data. RTS-17-20. October 2017."

Ceder, Avishai. 1991. "Transit Scheduling." *Journal of Advanced Transportation* 25 (2): 137–60. doi:10.1002/atr.5670250203.

Ceder, Avishai (Avi). 2005. "Estimation of Fleet Size for Variable Bus Schedules." *Transportation Research Record: Journal of the Transportation Research Board* 1903 (1): 2–10. doi:10.1177/0361198105190300101.

City of Brampton. 2017. "2018 Operating Budget." Accessed May 11, 2020. https://www.brampton.ca/EN/City-Hall/budget/2018%20Budget/Operating%20Budget/TRN_TRN_CC_2018.pdf.

Durham-Orange Light Rail Transit Project. 2012. "Operating and Maintenance Cost Methodology Report." Accessed May 11, 2020. https://gotriangle.org/sites/default/files/2.3-00659-om-cost-methodology-report-091012.pdf.

Government of Canada, Statistics Canada. 2020. *Consumer Price Index, Monthly, Percentage Change, Not Seasonally Adjusted, Canada, Provinces, Whitehorse and Yellowknife - Transportation*. Government of Canada, Statistics Canada. July. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000407.

Harmatuck, Donald J. 2005. "Cost Functions and Efficiency Estimates of Midwest Bus Transit Systems." *Transportation Research Record: Journal of the Transportation Research Board* 1932 (1): 43–53. doi:10.1177/0361198105193200106.

Meyer, Michael D., and Eric J. Miller. 2013. *Transportation Planning: A Decision-Oriented Approach*.

Miller, James H., and John C. Rea. 1973. "Comparison of Cost Models for Urban Transit." *HRB, Highway Research Record* 435: 11-19.

Stopher, Peter R., Len Brandrup, Byron Lee, and Stephen T. Parry. 1987. “Development of a Bus Operating Cost Allocation Model Compatible with UMTA Urban Transportation Planning System Models.” *Transportation Research Record* 1108: 31–41.

Toronto Transit Commission. 2011. “Ridership and Cost Statistics for Bus and Streetcar Routes, 2011.” *Surface Ridership 2011*. http://ttc.ca/About_the_TTC/Transit_Planning/Surface_Ridership_2011.jsp.

Toronto Transit Commission. 2012. “Ridership and Cost Statistics for Bus and Streetcar Routes, 2012.” *Surface Ridership 2012*. http://ttc.ca/About_the_TTC/Transit_Planning/Surface_Ridership_2012.jsp.

University of Wisconsin-Milwaukee Center for Urban Transportation Studies. n.d. “Transit Cost Analysis.” Accessed May 11, 2020. https://www4.uwm.edu/cuts/utp/cost.pdf.

Walker, Jarrett. 2011. “Basics: Operating Cost (02box).” *Human Transit*. July 5. https://humantransit.org/2011/07/02box.html.