

Examining the effects of autonomous vehicle ridesharing services on fixed- route public transit

by

Trevor Luke Vanderwoerd

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Civil Engineering

Waterloo, Ontario, Canada, 2020

© Trevor Luke Vanderwoerd 2020

Author's Declaration

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

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

Abstract

With the increased development of new transportation technologies such as autonomous vehicles and ridesharing fleet services comes the possibility of a new mode of transportation: autonomous vehicle ridesharing services (AVRS). The speed, convenience, accessibility, and low cost that AVRS is likely to offer will put it in competition with traditional fixed-route transit services. The effect of this competition is studied through the use of a four-step transportation demand model applied to hypothetical idealized urban networks. Under the assumptions that AVRS will cost 9% less per kilometer than owning and operating a personal automobile, that the AVRS service will have average wait times of 7 minutes, and that transit systems remain as they currently exist, the presence of AVRS in the network leads to an average loss per transit route of 49% of passenger-kilometers. Important transit route properties that correlate with decrease in passenger-kilometers include the passenger-kilometers before the introduction of AVRS and the headway. Additional effects of the introduction of AVRS could include an increase in delay due to congestion, an increase in travel times, and an increase in vehicle-kilometers travelled.

Acknowledgements

Without the help of those around me I would never have been able to complete this thesis. The aid they gave to me in the form of criticism, discussion of ideas, and emotional support was greatly appreciated. Among others, I would like to thank

my supervisor, Dr. Bruce Hellinga,

my parents, Jim and Valerie Vanderwoerd,

my siblings, and

the other graduate students that worked with me: Amir Zarinbal Masouleh, Wenfu

Wang, Darren Deng, Alan Xaykongsa, Andrew Orr, Ben Allen, Mohammad

Zarei.

Table of Contents

| | |
|---|------|
| List of Figures..... | vi |
| List of Tables..... | viii |
| 1 Introduction and Context..... | 1 |
| 2 Model Structure | 5 |
| 2.1 Route Table Generator | 10 |
| 2.1.1 Characteristics of Public Transit Routes in Canada..... | 12 |
| 2.1.2 Synthesizing Hypothetical Transit Routes | 26 |
| 2.2 Network Generator | 26 |
| 2.2.1 Road Grid..... | 28 |
| 2.2.2 Transit Route Alignments..... | 31 |
| 2.2.3 Traffic Analysis Zone Properties..... | 33 |
| 2.2.4 Road Network Properties | 39 |
| 2.2.5 Origin-Destination Matrix Calculation..... | 46 |
| 2.3 Four-Step Model | 48 |
| 2.3.1 Trip Generation..... | 50 |
| 2.3.2 Trip Distribution | 51 |
| 2.3.3 Mode Split | 53 |
| 2.3.4 Traffic Assignment..... | 62 |
| 2.3.5 Iteration | 63 |
| 3 Results | 66 |
| 3.1 Experimental Set-Up | 66 |
| 3.2 Impact of Autonomous Vehicle Ridesharing..... | 68 |
| 4 Limitations and Recommendations | 74 |
| References | 76 |
| Appendix | 80 |

List of Figures

| | |
|---|----|
| Figure 2.1: Overview of model structure | 7 |
| Figure 2.2: Distribution of coefficients of variation for subroute travel times..... | 15 |
| Figure 2.3: Distribution of coefficients of variation for subroute headways..... | 16 |
| Figure 2.4: Distribution of the number of trips in each subroute. | 18 |
| Figure 2.5: Distribution of Transit Route Length (km). | 21 |
| Figure 2.6: Distribution of Transit Route Travel Times (minutes)..... | 22 |
| Figure 2.7: Distribution of Transit Route Headways (minutes). | 23 |
| Figure 2.8: Distribution of Transit Route Operating Speeds (km/h)..... | 24 |
| Figure 2.9: Distribution of Transit Route Indirectness. | 25 |
| Figure 2.10: Network generator components | 28 |
| Figure 2.11: Visual definition of the road grid. Only one TAZ is shown for visual clarity. Similarly, in following figures the simplification in the top right corner is used. | 29 |
| Figure 2.12: Road grid with segments removed..... | 30 |
| Figure 2.13: Example transit route alignment, with diversion. | 33 |
| Figure 2.14: Distribution of dissemination area densities..... | 35 |
| Figure 2.15: Distribution of population densities | 36 |
| Figure 2.16: Final population density distribution..... | 37 |
| Figure 2.17: Final population density distribution, detail | 38 |
| Figure 2.18: BPR functions for automobile and bus transit. The solid line is automobile, the dashed line is bus..... | 44 |
| Figure 2.19: Vector field used to create a circular path. | 45 |
| Figure 2.20: City map showing an example configuration of road types. The thick yellow roads are highways, the red is arterial1, the orange arterial2, and the grey urban local. | 46 |
| Figure 2.21: Structure of the graph used to calculate transit origin-destination matrices.. | 48 |
| Figure 2.22: Four Step Model components with iteration. | 64 |
| Figure 3.1: Relationship between initial passenger kilometers and change in passenger kilometers. Note that the axes limits have been chosen to show the detail of the | |

| | |
|---|----|
| majority of routes at the cost of excluding outliers. Full figure can be found in the appendix, Figure 4.2. | 70 |
| Figure 3.2: Relationship between route headway and the relative change in passenger-kilometers. | 71 |
| Figure 4.1: Venn diagrams showing filtering logic. Numbers in diagram represent percent of total subroutes for that data set. The accepted subroutes are shown in black-outlined shape in the center. See section 2.1.1.2. | 80 |
| Figure 4.2: Extended view of Figure 3.1 including all data points. | 81 |

List of Tables

| | |
|--|----|
| Table 2.1: Model Inputs | 8 |
| Table 2.2: Model Parameters..... | 9 |
| Table 2.3: Transit Route Characteristics..... | 11 |
| Table 2.4: Raw GTFS Data Characteristics | 13 |
| Table 2.5: Shape File Excerpt | 13 |
| Table 2.6: Example of a Schedule Table | 14 |
| Table 2.7: Aggregated Data Counts | 17 |
| Table 2.8: Example Length and Trip Percentage | 19 |
| Table 2.9: Number of subroutes in each dataset | 19 |
| Table 2.10: Sample census data (Population and Dwelling Count Highlight Table) | 35 |
| Table 2.11: Amount of land by population density (example)..... | 36 |
| Table 2.12: Automobile-related parameters of link profiles | 40 |
| Table 2.13: Parameters for CCG signalized intersection approach delay model | 42 |
| Table 2.14: Bus transit-related parameters of link profiles | 43 |
| Table 2.15: Summary of TTS Trip Data | 59 |
| Table 3.1: Parameters used to create hypothetical networks | 67 |
| Table 3.2: Modelled City Characteristics..... | 68 |
| Table 3.3: Correlation coefficients for quantitative route characteristics..... | 70 |
| Table 4.1: Transit route dataset summary for bus routes in small cities..... | 82 |
| Table 4.2: Transit route dataset summary for bus routes in large cities | 82 |
| Table 4.3: Transit route dataset summary for rail routes..... | 83 |
| Table 4.4: Mode usage for modelled cities | 84 |
| Table 4.5: Trip Characteristics for modelled cities | 85 |

1 Introduction and Context

New technologies are often the largest driver in changes to transportation, ranging from more efficient and safe variations of existing technology to entirely new modes. More dramatic changes such as the development of a new transportation mode tend to lead to changes in the built environment, new infrastructure requirements, shifts in societal behaviour, and increased economic development. During these larger developments in transportation technology, it is important to correctly anticipate their impacts to aid in administrative decisions and prevent undesired side-effects.

In recent years, the most dramatic change in transportation has been the development and proliferation of ridesharing services provided by fleet operators such as Uber and Lyft. While this is not entirely a new form of transportation service, as taxis have been in existence for over 100 years, by taking advantage of the access to communication provided by smartphones and by restructuring wages, these ridesharing services have become more convenient and at lower cost for the customer.

Another recent transportation technology development is the autonomous vehicle. Both the manufacturers and researchers (Ohnemus et al., 2016) have projected that autonomous vehicles will be safer, more energy efficient, increase traffic flow, and remove barriers to transportation. These two technological developments are expected to be combined in the form of ridesharing services using autonomous vehicles to provide transportation to users in response to demand. This type of transportation can be thought of as a new mode of transportation for the purposes of modelling travel decisions, in particular mode choice. This mode goes by a number of names in the literature, including connected autonomous vehicles (CAV), autonomous taxis (aTaxis), and autonomous vehicle ridesharing services (AVRS). For this thesis, the term AVRS will be used.

In the context of this thesis, AVRS is assumed to operate similar to existing taxi services (i.e. a demand responsive mode providing transportation from origin to destination) but at a lower cost (due to the fact that there is no need to pay a driver). The term “ridesharing” indicates that the vehicle is not owned by the trip maker (as is the case with personally owned automobiles). It does not indicate that the service provides

simultaneous same-vehicle rides for multiple trip makers with different origins and destinations.

It is anticipated that the widespread adoption of AVRS will have an impact on traditional fixed-route public transit services such as government-operated bus and rail systems, and the impact will likely be a decrease in demand for these services. Traditional transit in Canada plays a secondary role to the main transportation mode (personally owned automobiles) and is commonly seen as a fallback that is used when automobiles cannot be used. This includes captive transit riders: groups of people who do not have reliable access to automobiles. Only in rare cases is public transportation a first option, typically where congestion is high enough to cause significant reductions in travel time and convenience. Due to the role transit plays in ensuring transportation for everyone, it is typically government managed and funded. AVRS will likely compete favourably with transit in terms of travel time and convenience, and potentially cost. As a result, the future role that transit will occupy is unknown, but it may involve shifting to incorporate AVRS, increasing the number of separate right-of-way routes, or withdrawing to only the busiest corridors.

The objective of this study was to gain an understanding of how the demand for traditional fixed-route public transit could change in the case where AVRS becomes a common form of transportation. More specifically, the objective is to quantify the changes in demand and to determine if changes in demand are correlated with the transit route characteristics. This includes how changes in transit demand will have an effect on the city as a whole, which characteristics of transit have relationships with changes in transit demand, and the nature of those relationships.

Research on AVRS or relevant to AVRS can be broadly divided into three groups: research on the autonomous vehicles themselves, research on the fleet operations and systems needed to run an AVRS company, and research on the potential impacts AVRS systems could have on the existing transportation landscape. The autonomous vehicles themselves are not particularly relevant for this work. Most research on AVRS itself tends to be research on fleet operations, including work to examine the costs (Bösch et al., 2018), wait times, algorithms (Ma et al., 2017, Hanna et al., 2016, Winter et al., 2016), and fleet sizes (Fagnant et al., 2014, Fagnant et al., 2016, Winter et al., 2017) of potential AVRS

systems, and how components of AVRS can be integrated into transit systems often in a last-mile context (Liang et al., 2016). Research into the effects of AVRS is most often on the topic of city-wide impacts, benefits with respect to changes in congestion levels, and changes in the traffic flows of various infrastructures (Ghiasi et al., 2017). Some research has been done on the competition between AVRS and transit and how the demand for transit will evolve, such as Levin and Boyle (2015), who integrate autonomous vehicles into a Four-Step model but the autonomous vehicles are privately owned (there is no ride- or carsharing) and the effects on specific transit routes are not examined.

In this thesis, impacts of AVRS on demand for public transit were investigated by constructing a four-step transportation demand model and applying this model to a set of urban networks of various sizes with and without the AVRS mode. A set of models were developed to create the networks, allowing for increased generalization and reducing data collection needs. Transit routes were measured and categorized by transit type, headway, length, and speed (among other measurements). These transit characteristics from the simulations with and without AVRS were compared to find relationships to the changes in demand, measured by passenger-kilometers per hour. Findings suggest that with the availability of AVRS, demand and usage of traditional fixed-route transit will decline dramatically (about 40%), and that this decrease is relatively consistent across the variations of transit routes. The characteristics that make transit routes more viable are short headways and separate rights-of-way. Additionally, AVRS could cause an increase in the vehicle kilometers travelled city-wide, increasing congestion and decreasing travel times for all travellers.

The remainder of this thesis is structured as follows:

Chapter 2 provides an overview of the methodological approach taken in the research, the structure of the model that was developed, and the details pertaining to each component of the model. Where appropriate, references are made to the literature with respect to elements of the model components, model parameter values, and/or data sources.

Chapter 3 describes how the developed model was used to simulate a set of different hypothetical urban networks and presents the ensuing model results as well as an interpretation of these results.

Chapter 4 identifies limitations associated with this study and presents a set of recommendations for future research to extend this work.

2 Model Structure

The overall goal of this research was to examine the impact that AVRS is likely to have on public transit ridership and thereby provide public transit agencies with some insight into how to plan for the introduction of AVRS.

At the outset of this work, it was necessary to determine the most appropriate structure for the modelling. It was determined that two main approaches could be considered.

The first would be to use conventional urban transportation models (such as the four-step model) to examine the impact of introducing AVRS into a specific community, such as Waterloo Region. The advantage to this approach is that the existing transportation network (e.g. roads, public transit routes, etc) and the land use data (e.g. population and employment distributions and density) would already be known and would be used as input to the four-step transportation model. The disadvantage is that the results (i.e. impact of AVRS) would be restricted to the specific characteristics of the urban area that had been modelled.

A second approach was to develop a modelling framework that creates hypothetical urban areas (i.e. cities) with user controlled attributes, and for these hypothetical cities, apply the four-step model to determine mode splits. The advantage to this approach is that the impact of introducing AVRS could be evaluated over a range of conditions (e.g. size of urban area, characteristics of road and transit network, etc.). The disadvantage is that this approach would require a model to generate the transportation network (roads and transit routes) and the population and employment densities in each zone and concerns about the validation of the modelling results.

Factors such as time and complexity, accuracy, scaling, and control were taken into account when deliberating between these two modelling approaches. Concerns associated with using a model to create the network are alleviated by recognizing that by nature this project has some limits to accuracy. The AVRS mode is an extremely new technology and mode, so its characteristics are not fully known. In light of having to make broad assumptions about AVRS costs, travel times, and more, it is unlikely that the increases in accuracy of a real-world network like the precise distribution of population and travel

destinations or the specific alignments of transit routes would have a substantial positive effect. This conclusion is strengthened by noting that the amount of variation found in real-world cities makes it unlikely that studying any particular city is any better than studying a variety of generalized cities created by the model. Using a self-made model to create networks also gives more control over experimentation.

Given these characteristics of the two different modelling approaches, it was decided to adopt the second approach, and develop a model to generate hypothetical urban areas and model each with the four-step transportation model.

The complete model has a three part structure which together create transit routes, create a city with roads and population/employment, and simulate the travel demand and transportation infrastructure use. The model was created for this study and for the purposes of studying how transit ridership could be affected by the emergence of a new mode, specifically autonomous vehicle ride-sharing services (AVRS). In order to accomplish that goal, there are a number of requirements. The model needs to be able to simulate transportation demand, both the more mechanical aspects such as congestion, route planning, and road use, and the decision-based aspects like mode choice and destination choice. Such transportation demand modelling is traditionally done with the Four-Step Model, but has also been done (sometimes more accurately) by using activity or tour-based models. The model especially needs to be able to handle decisions regarding mode choice across a variety of situations: different demands, trip lengths, congestion levels, and types of public transit. Four-step models handle these mode choice decisions explicitly and in isolation, whereas activity-based models handle them in conjunction with other decisions, but however they are handled they are a crucial aspect of the model for this project. The model must also take into account effects such as the costs of using the modes and the effects of congestion (and especially the different ways in which congestion affects the different modes).

The three part structure is made up of the Route Generator, the Network Generator, and a Four-Step model. The route generator was tasked with creating transit routes by creating sets of properties that were judged to be crucial to the nature of a transit route and to a travel (mode choice) decision. The network generator handled the creation of the city properties necessary for transportation modelling and for the modes being considered.

This structure is outlined in Figure 2.1. Each part of the overall model has inputs, many of which become outputs for another stage. There is a small number of inputs to the entire model which are used to control what type of city and transit scenario is modelled. The other inputs are calibrated parameters. These inputs and parameters are shown in Table 2.1 and Table 2.2.

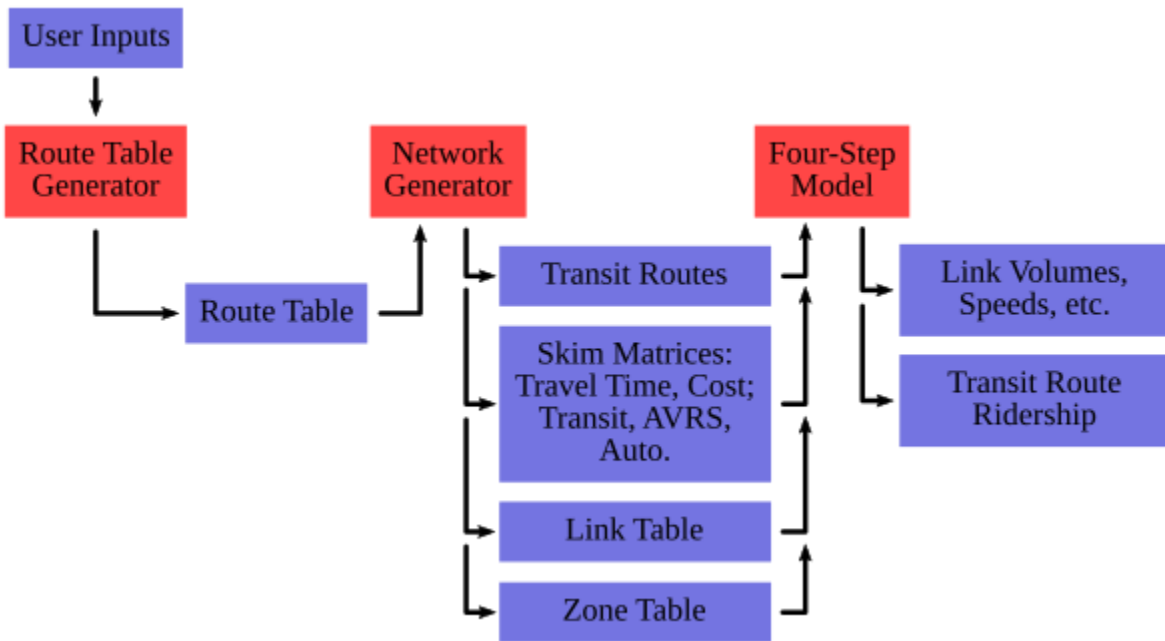


Figure 2.1: Overview of model structure

In these tables, inputs are considered to be the values that do not have a value with clear theoretical foundation, and therefore are values that need to be given by the user. They are also the values that are important in terms of controlling the scenario being modelled. Parameters are the values that do have a theoretical foundation and/or can be calculated/calibrated from applicable data, and therefore can be left to the default values. They can, however, still be changed to create a certain scenario.

Table 2.1: Model Inputs

| Input | Sub-model | Description |
|-----------------------|-------------------|--|
| Transit Dataset | Route Table | Controls whether bus routes are created to match a small city profile or a large city profile. |
| Number of Bus Routes | Route Table | |
| Number of Rail Routes | Route Table | |
| Population | Network Generator | Goal population for the city, the actual population will be slightly above this. |
| Minimum City Size | Network Generator | Minimum side length of the city square, in blocks (see block size). |
| AVRS Cost | Four Step Model | Cost per kilometer, in 2019 CAD, of using an AVRS service |
| AVRS Wait Time | Four Step Model | Average time in minutes that is spent waiting for AVRS services |

Table 2.2: Model Parameters

| Parameter | Sub-model | Description |
|-------------------------|-------------------|---|
| Block Size (b) | Network Generator | The distance in meters between parallel roads in the road grid, and therefore also the distance between zone centres. |
| Fill Density | Network Generator | Population density (in people/km ²) of rural areas surrounding the city. |
| Bridge Gap | Network Generator | Mean distance in blocks (see block size) between bridges |
| Bridge Deviation | Network Generator | Standard deviation, in blocks, of the distance between bridges |
| Remove Percent | Network Generator | Percent of roads to remove from the basic grid. |
| VDF parameters | Network Generator | α_a and β_a in the volume delay function (equation (2.20)) for each road type |
| Number of Lanes | Network Generator | The number of lanes for each road type |
| Road Capacity | Network Generator | Capacity, in vehicles per hour, for each road type. |
| Free Speed | Network Generator | Free speed, in kilometers per hour, for each road type. |
| Automobile Cost | Four Step Model | Cost per kilometer, in 2019 CAD, of owning and operating a personal automobile. |
| Transit Fare | Four Step Model | Cost per trip, in 2019 CAD, of public transit. |
| Walking Speed | Four Step Model | Walking speed, in meters per second |
| Trips per Person | Four Step Model | Trips produced per person per day. |
| Trips per Employment | Four Step Model | Trips attracted by each worker per day. |
| Peak Hour Trips per Day | Four Step Model | The proportion of daily trips that take place during the peak hour. |
| Impedance a | Four Step Model | a in the impedance function of the gravity model (equation (2.28)). |
| Impedance b | Four Step Model | b in the impedance function of the gravity model (equation (2.28)). |
| Mode Split Functions | Four Step Model | Includes the constants and coefficients for the utility function for each mode. |

When working on a project that seeks to integrate many different facets, it is important to be clear about the scope. Autonomous vehicles are projected to be part of technologies such as connected vehicles, automated intersections, platooning, and others that could change the nature of traffic flow compared to human-driven vehicles. These potential features are not included in this study because of the extra complexity, lack of consensus of how they operate or how they are modelled, and including them would mean

manually implementing the models. Another limitation to the scope is that the project assumes a Canadian context - the cities that the network generator creates are meant to capture the properties and distributions found in Canada. Where Canadian information is not present or available, information from the USA is used instead.

2.1 Route Table Generator

The job of the route table generator is to produce transit routes that are similar to those found in Canadian cities. There are many configurations of transit routes, and many ways to measure and characterize them. The route table generation should produce a variety of routes, but not a completely random variety as some of the possibilities may not be an accurate reflection of real transit routes. To address these needs, the route table generator samples from a data set of real-world transit routes, after that data set has been smoothed by kernel density estimation. This model has three areas that need configuration:

1. Which characteristics of transit routes to include in the model,
2. What data source to use as the basis for the model, and
3. What type of sampling and smoothing to use.

Transit routes have been characterized and described, qualitatively and quantitatively, for operational and research purposes. Vuchic (2007) describes some of the most common: the order of transit (i.e. higher order such as subway, and lower order such as fixed route bus), the headway, capacity, ridership, revenue/cost ratio, operating cost, passenger cost, and load factor. These characteristics have varying values to the route table generator, according to their ease of measurement, their significance in passenger mode-choice decisions, the ease of modelling their impact, and the consistency with which they apply to all trips made using the route. Using these considerations, 9 characteristics were chosen to be studied in depth. These are shown in Table 2.3.

Table 2.3: Transit Route Characteristics

| Characteristics | Symbol | Unit | Meaning |
|----------------------|--------|---------------|---|
| Headway | h | min | The time between successive visits by transit vehicles |
| Route Length | ℓ | km | The total distance covered by the transit vehicle from the beginning to the end of the route. |
| Geodesic Distance | g | km | The shortest distance between the starting and ending location of the route. |
| Route Indirectness | i_r | n/a | The route length divided by the geodesic distance; a higher ratio means the route takes a less direct path. |
| Total Travel Time | t | min | The total time elapsed during a single one-way trip. |
| Operating Speed | s | km/h | The route length divided by the travel time; the average speed of the vehicle during a single one-way trip. |
| Wait time percentage | w | n/a | The average wait time [†] divided by total travel time; the typical proportion of a user's trip that they spend waiting. [†] Average wait time is half of the headway. |
| Cost | c | \$ (2019 CAD) | The average amount paid by a single passenger to use the route. |
| Peak Ridership | d | passengers/hr | The number of people using the route during the busiest hour of the day. |

Of these route characteristics, some will be a direct result from the route table generator: headway, route length, geodesic distance, and total travel time. The operating speed, route indirectness, and wait time percentage can be calculated from the headway, length, geodesic, and travel time:

$$s = \frac{\ell}{t} \quad (2.1)$$

$$i_r = \frac{\ell}{g} \quad (2.2)$$

$$w = \frac{h}{2t} \cdot 50\% \quad (2.3)$$

Peak ridership is dependent on many factors and is not an inherent characteristic of the route. Instead, it is estimated during the simulation process and is an output of the four-step model. All of the characteristics produced by the route table generator are used as inputs to the network generator and indirectly to the four-step travel demand model.

For most Canadian transit systems, fare cost is the same across a given transit service (i.e. fixed fare regardless of the route(s) used), and therefore this is also assumed in this work.

With Table 2.3 and the exceptions above, for the purposes of the route table generator a transit route is a collection of four characteristics: headway, length, geodesic distance, and travel time. The next sections describe the creation of a data set that includes those four characteristics, and the model used to create hypothetical routes with that data set.

2.1.1 Characteristics of Public Transit Routes in Canada

Many transit agencies provide data to the public formatted according to the General Transit Feed Specification (GTFS) (Google 2019), which includes schedule and route alignment information. GTFS data was obtained for 7 cities in Canada, chosen to cover a range of city sizes, transit services, and geographies: Toronto (City of Toronto 2018), Montréal (Société de transport de Montréal 2018), Vancouver (Translink 2018), Calgary (City of Calgary 2018), Hamilton (City of Hamilton 2018), Waterloo Region (Grand River Transit 2018), and London (London Transit Commission 2018). The routes and trips in the raw data sets for these 7 cities are summarized in Table 2.4. This data was then aggregated to create datasets that have a large number of routes, each with values for the four characteristics.

GTFS data provides data as a set of tables. For this research, the most relevant tables are the route, trip, stop time, and shape tables. The route table lists all the routes managed by a transit agency, including their name and the type of transit (i.e. bus and rail). The trip table lists every trip made by a transit vehicle as well as information about what route the trip belongs to, the direction, and the shape of the trip. The stop times table lists the time that each trip is scheduled to stop at each stop along its trip, which can be used to know precisely when each vehicle is in which location. Finally, the shape table lists every physical path that the trips can take, by giving a sequence of latitude-longitude pairs for each shape.

Table 2.4: Raw GTFS Data Characteristics

| City | Routes | Bus Routes | Rail Routes | Trips | Bus Trips | Rail Trips |
|------------|--------|---------------|-------------|---------|-----------------|---------------|
| Calgary | 295 | 293 (99.3%) | 2 (0.7%) | 11,196 | 10,515 (93.9%) | 681 (6.1%) |
| Hamilton | 44 | 44 (100%) | 0 (0%) | 3,830 | 3,830 (100%) | 0 (0%) |
| London | 42 | 42 (100%) | 0 (0%) | 3,726 | 3,726 (100%) | 0 (0%) |
| Montreal | 228 | 224 (98.2%) | 4 (1.8%) | 20,553 | 18,803 (91.5%) | 1,750 (8.5%) |
| Toronto | 205 | 190 (92.7%) | 15 (7.3%) | 43,705 | 37,751 (86.4%) | 5,954 (13.6%) |
| Vancouver | 242 | 239 (98.8%) | 3 (1.2%) | 24,303 | 22,304 (91.8%) | 1,999 (8.2%) |
| Waterloo | 79 | 79 (100%) | 0 (0%) | 3,303 | 3,303 (100%) | 0 (0%) |
| All Cities | 1,135 | 1,111 (97.9%) | 24 (2.1%) | 110,616 | 100,232 (90.6%) | 10,384 (9.4%) |

2.1.1.1 Aggregation

To obtain the desired data set, the GTFS data for each city had to be reformatted and aggregated, using a two-stage aggregation. First, the 4 characteristics are found for each trip. Then, the trips are aggregated into subroutes according to their physical alignment.

The route length and geodesic distance were calculated using the shape files. An excerpt of a shape file is shown in Table 2.5. The shape file includes the sequence of locations the route visits as latitudes and longitudes, which can be used to calculate the route length, and the geodesic distance can be calculated using the first and last location.

Table 2.5: Shape File Excerpt

| Shape ID | Latitude | Longitude | Sequence |
|----------|-----------|------------|----------|
| 10011 | 43.435221 | -80.560327 | 10001 |
| 10011 | 43.434987 | -80.560079 | 10002 |
| 10011 | 43.434688 | -80.559763 | 10003 |
| 10011 | 43.434681 | -80.559753 | 10004 |
| 10011 | 43.434663 | -80.55978 | 10005 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 10011 | 43.422757 | -80.440732 | 660011 |
| 10011 | 43.422906 | -80.440808 | 660012 |
| 10011 | 43.423253 | -80.441242 | 660013 |

The travel time and the headway of each trip is calculated using a schedule table, such as the one shown in Table 2.6. The travel time and headways are calculated for each trip, then aggregated into a single value for the route. Let $s_{i,k}$ be the scheduled stop time of trip i at stop k . The travel time of trip i is

$$t_i = s_{i,kmax} - s_{i,kmin} \tag{2.4}$$

The trip headway is calculated from the average of the headway at each stop:

$$h_i = \frac{1}{n} \sum_k (s_{i_{next},k} - s_{i,k}) \quad (2.5)$$

where i_{next} is the index of the next trip after i that visits stop k and n is the number of stops. For example, trip 1 in Table 2.6 has a headway of

$$\begin{aligned} h_1 &= \frac{1}{45} \left((s_{2,100} - s_{1,100}) + \dots + (s_{2,142} - s_{1,142}) + (s_{3,143} - s_{1,143}) \right. \\ &\quad \left. + (s_{3,144} - s_{1,144}) \right) \\ &= \frac{1}{45} (5 + \dots + 5 + 10 + 10) \\ &= 5.22 \text{ minutes} \end{aligned}$$

If the headway at any stop would be greater than 90 minutes, then that stop is omitted from the calculation. This ensures that breaks in service won't result in headways that are multiple hours long.

Table 2.6: Example of a Schedule Table

| Trip | Time at Stop Number | | | | | |
|------|---------------------|-------|-------|-------|-------|-------|
| | 100 | 101 | 102 | 142 | 143 | 144 |
| 1 | 07:00 | 07:04 | 07:06 | 08:13 | 08:15 | 08:19 |
| 2 | 07:05 | 07:09 | 07:11 | 08:18 | n/a | n/a |
| 3 | 07:10 | 07:14 | 07:16 | 08:23 | 08:25 | 08:29 |
| 4 | 07:15 | 07:19 | 07:21 | 08:28 | n/a | n/a |
| 5 | 07:20 | 07:24 | 07:26 | 08:33 | 08:35 | 08:39 |

Each trip within a route can have different values for the 4 characteristics calculated above, meaning that aggregating all these trips together would lead to problems. In fact, it is possible for two trips from the same route to have different physical paths, ranging from short-turns that only cover half the route, to trunk-and-branch routes with many paths at the ends, to breaking up a route into multiple back to back trips. To account for all this variation, the trips are aggregated by *subroute*: two trips belong to the same subroute if they have the exact same route length and geodesic distance. Naturally, this removes any variation in the length and geodesic. The variation in travel time over the course of a day is mostly due to changes in congestion, while the variation in headway is due to changes in service level which is driven by demand. For travel times, the variation is typically a difference of about 5 minutes, which is not significant relative to the travel times themselves (typically at least 20 minutes). Therefore, the subroute travel time is found by averaging the travel time for each trip. Figure 2.2 shows the distribution of the coefficients of variation (standard deviation divided by the mean) for the aggregated subroute travel

times. This figure demonstrates that taking the average travel time is not introducing a large amount of variation - the coefficients of variation are nearly always below 0.16.

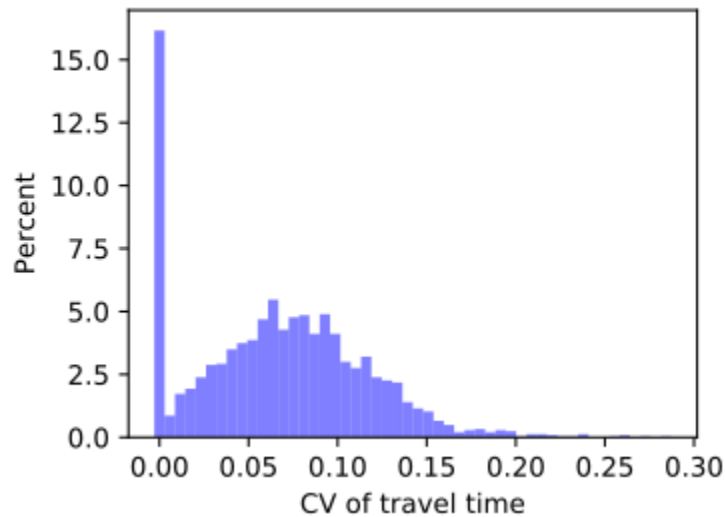


Figure 2.2: Distribution of coefficients of variation for subroute travel times.

For the headways, a common pattern is that they are significantly longer in the early mornings and late nights than during the remainder of the day. It is most accurate in these cases to say that the headway of the subroute is the headway that occurred most frequently. Therefore, subroute headways are found using the mean of trips which begin their journey between 6 AM and 7 PM. Figure 2.3 compares the distributions of the coefficients of variation for the headway calculated by including all trips, and by including only trips from 6 AM to 7 PM. Excluding early morning and late night trips decreases the variation, meaning that the trips that have been excluded are atypical.

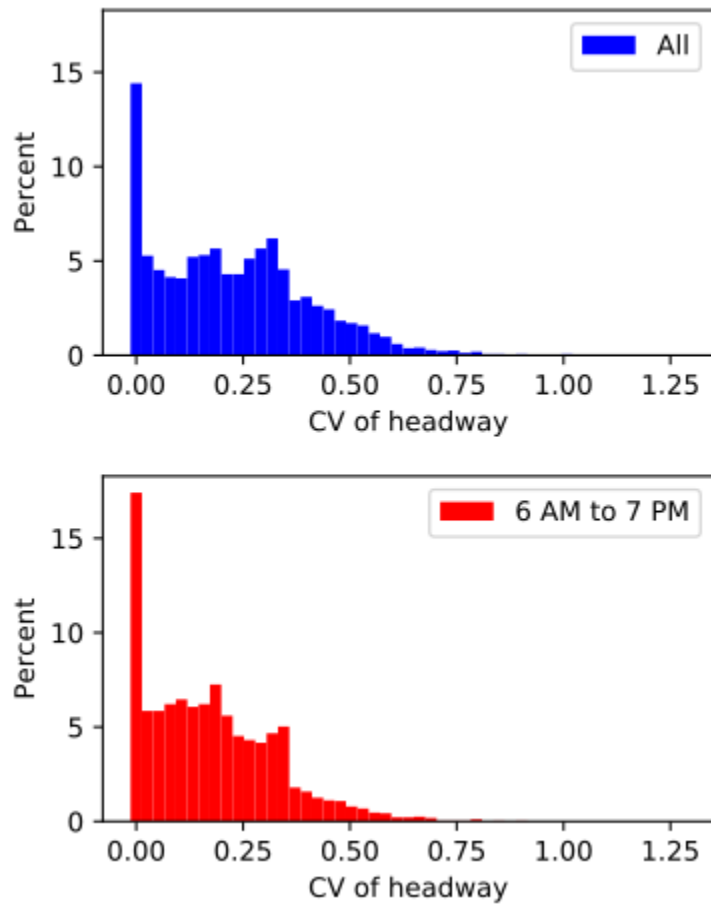


Figure 2.3: Distribution of coefficients of variation for subroute headways.

After aggregation, there are 3,164 subroutes, 3,019 bus subroutes and 145 rail subroutes. It should be noted that 380 subroutes (12.0%) do not have any trips that start between 6 AM and 7 PM, which means that these subroutes are effectively filtered from the dataset. Table 2.7 shows the subroute counts by city and mode, and includes the 380 subroutes without a daytime headway.

Table 2.7: Aggregated Data Counts

| City | Subroutes | Bus Subroutes | Rail Subroutes |
|------------|-----------|---------------|----------------|
| Calgary | 453 | 443 (97.8%) | 10 (2.2%) |
| Hamilton | 116 | 116 (100%) | 0 (0%) |
| London | 181 | 181 (100%) | 0 (0%) |
| Montreal | 560 | 551 (98.4%) | 9 (1.6%) |
| Toronto | 935 | 835 (89.3%) | 100 (10.7%) |
| Vancouver | 778 | 752 (96.7%) | 26 (3.3%) |
| Waterloo | 141 | 141 (100%) | 0 (0%) |
| All Cities | 3,164 | 3,019 (95.4%) | 145 (4.6%) |

2.1.1.2 Filtering

After the aggregation work, the subroutes are filtered to ensure that the final data set doesn't include subroutes that are unrealistic. Two filters are applied: one to eliminate routes that start and end in the same location, and the other to ensure that the subroute has enough trips to be considered a proper route.

The presence of routes that start and end in the same location (loop routes) creates a problem because they have a geodesic of zero, and therefore the route indirectness cannot be calculated. However, it is relatively easy for a route which, for all practical purposes, starts and ends at the same location to have a non-zero geodesic. For example, the starting and ending points may be on opposite sides of a street or building. To ensure that these routes are also eliminated, trips with a route indirectness greater than 6 and a geodesic less than 1 km are also considered loop routes. In short, after applying the loop route filter, all routes have a non-zero geodesic and must have an indirectness of 6 or less or a geodesic of 1 km or greater. The threshold of route indirectness was chosen after looking at the maps of various routes with high indirectness and observing that loop routes (as classified subjectively) tend to have route indirectness greater than 6.

Another issue with the data is that there are a fairly large number of subroutes that have either a much shorter route length as compared to other subroutes of the same route, or have many fewer trips than similar length subroutes of the same route. Figure 2.4 shows the large number of subroutes that have very few trips - in fact, around 20% of all subroutes have 4 or fewer trips. These subroutes are found in the data as short turns, or because the transit agency represented a route as a set of smaller, interconnected route fragments.

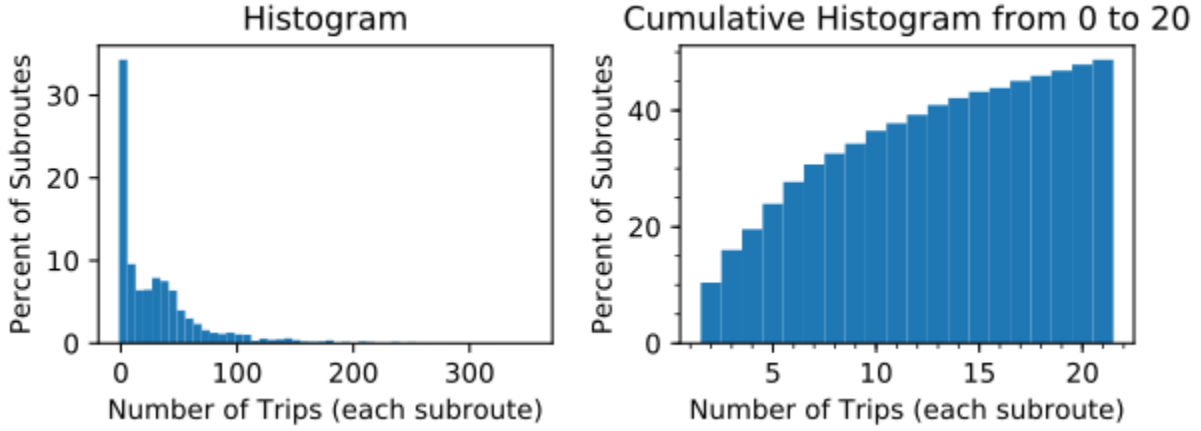


Figure 2.4: Distribution of the number of trips in each subroute.

Two additional measurements are introduced to filter out such routes: length percentage, and trip percentage. Length percentage is calculated for each subroute using Equation (2.6).

$$q_s = \frac{\ell_s}{\max_{r \in R} \ell_r} \cdot 100\% \quad (2.6)$$

where q_s is the length percentage of subroute s , ℓ_r and ℓ_s are the lengths of subroutes r and s , and R is the set of subroutes that make up the same route as s . Trip percentage is calculated as in equation (2.7).

$$z_s = \frac{\sum_{x \in G} n_x}{\sum_{y \in R} n_y} \cdot 100\% \quad (2.7)$$

where z_s is the trip percentage of subroute s , n_i is the number of trips in subroute i , R is the set of subroutes that make up the same route as s , and G is the set of subroutes with a length percentage equal to ℓ_s to the nearest ten. To clarify, if $round()$ is the rounding function to the nearest ten, then

$$G \equiv \{x \in R \mid round(\ell_s) = round(\ell_x)\}. \quad (2.8)$$

Subroutes are kept if they have a length percentage $\geq 25\%$ and a trip percentage $\geq 10\%$.

Note that the calculation for trip percentage is not simply the percentage of trips that a given subroute contributes to the route as a whole. It was found that using a naive approach eliminated too many subroutes and upon further examination, it eliminated subroutes that subjectively were not atypical. These situations occurred when a route consisted of many similar subroutes, which individually did not represent a significant

amount of the route's trips, but together did. In other words, some subroutes display small variations that did not fundamentally change the subroute, and so are grouped together when calculating the trip percentage. Examples of these calculations are shown in Table 2.8.

Table 2.8: Example Length and Trip Percentage

| Route | Subroute | Length (km) | Trips | Length % | Trip % |
|-------|----------|-------------|-------|-------------|--------|
| A | 1 | 8.6 | 6 | 100.0 (100) | 12.2 |
| A | 2 | 4.7 | 30 | 54.7 (50) | 79.6 |
| A | 3 | 1.2 | 4 | 14.0 (10) | 8.2 |
| A | 4 | 4.5 | 9 | 52.3 (50) | 79.6 |
| B | 1 | 10.0 | 107 | 100 (100) | 52.5 |
| B | 2 | 4.2 | 18 | 42 (40) | 45.6 |
| B | 3 | 3.9 | 19 | 39 (40) | 45.6 |
| B | 4 | 3.8 | 17 | 38 (40) | 45.6 |
| B | 5 | 4.1 | 19 | 41 (40) | 45.6 |
| B | 6 | 4.0 | 20 | 40 (40) | 45.6 |
| B | 7 | 2.0 | 4 | 20 (20) | 2.0 |

In combination, the filters described above as well as only using headways during daytime results in 5 separate filters. Figure 4.1 in the appendix details in the interactions of the logic more clearly, as well as containing information about how many subroutes are eliminated as a result of the filters.

2.1.1.3 Resulting Dataset

The resulting dataset consists of 2,264 subroutes, or 71.6% of the 3,164 initial subroutes. Since the transit routes come from a range of city sizes and transit modes, the data set is split into three: rail routes, bus routes from small cities (Calgary, Hamilton, Waterloo, and London), and bus routes from large cities (Toronto, Montréal, and Vancouver). The number of subroutes in each dataset is shown in Table 2.9. The appendix contains data summaries (Table 4.1-Table 4.3) and Figure 2.5 through Figure 2.9 show the characteristics of the datasets.

Table 2.9: Number of subroutes in each dataset

| Dataset | Number of Subroutes |
|-------------------|---------------------|
| Rail | 54 |
| Bus, small cities | 614 |
| Bus, large cities | 1596 |

The histograms of length and travel time (Figure 2.5 and Figure 2.6) display expected characteristics: the routes are the appropriate length to allow for mobility around a city, and the distributions are visually Normal with a slight right skew. Rail routes cover a larger range of lengths.

Figure 2.7 shows the distribution of headway, and it is clear that the three datasets have different characteristics: rail routes have short headways, as they typically serve higher demand. Large cities also tend to have higher demand bus routes which requires shorter headways, whereas small cities tend to have more low-demand routes which are serviced with longer headways. In both of the bus histograms, the effect of clock-facing headway policies (headways which are factors of 60 minutes) can be seen in the spikes at headways such as 5, 20, and 30 minutes.

Figure 2.8 shows the speed distribution. Buses typically travel at around 15-30 km/h, with buses in larger cities traveling slower than those in smaller cities. This is likely due to higher congestion in larger cities. For rail routes, there is a bimodal structure. The rail dataset includes Toronto and Montréal's metros, Vancouver's SkyTrains, Calgary's LRT, and Toronto's streetcars. The bimodal speed distribution is caused by the large differences between streetcars and higher right-of-way trains: the slower group of subroutes (slower than most buses) are the streetcars, whereas the dedicated right-of-way rail routes typically operate at 30-40 km/h, faster than most buses.

Figure 2.9 shows the distribution of route indirectness, and that rail routes are noticeably more direct than bus routes. This is likely because the higher infrastructure costs of rail, the increased permanence, and the higher capacity all mean that there is more focus on quick, direct, and efficient movement. It should be noted that there are 23 subroutes with route indirectness greater than 6 that are not shown in Figure 2.9, this is because if they are shown, the area near 1, which is where most routes are found, becomes a single spike.

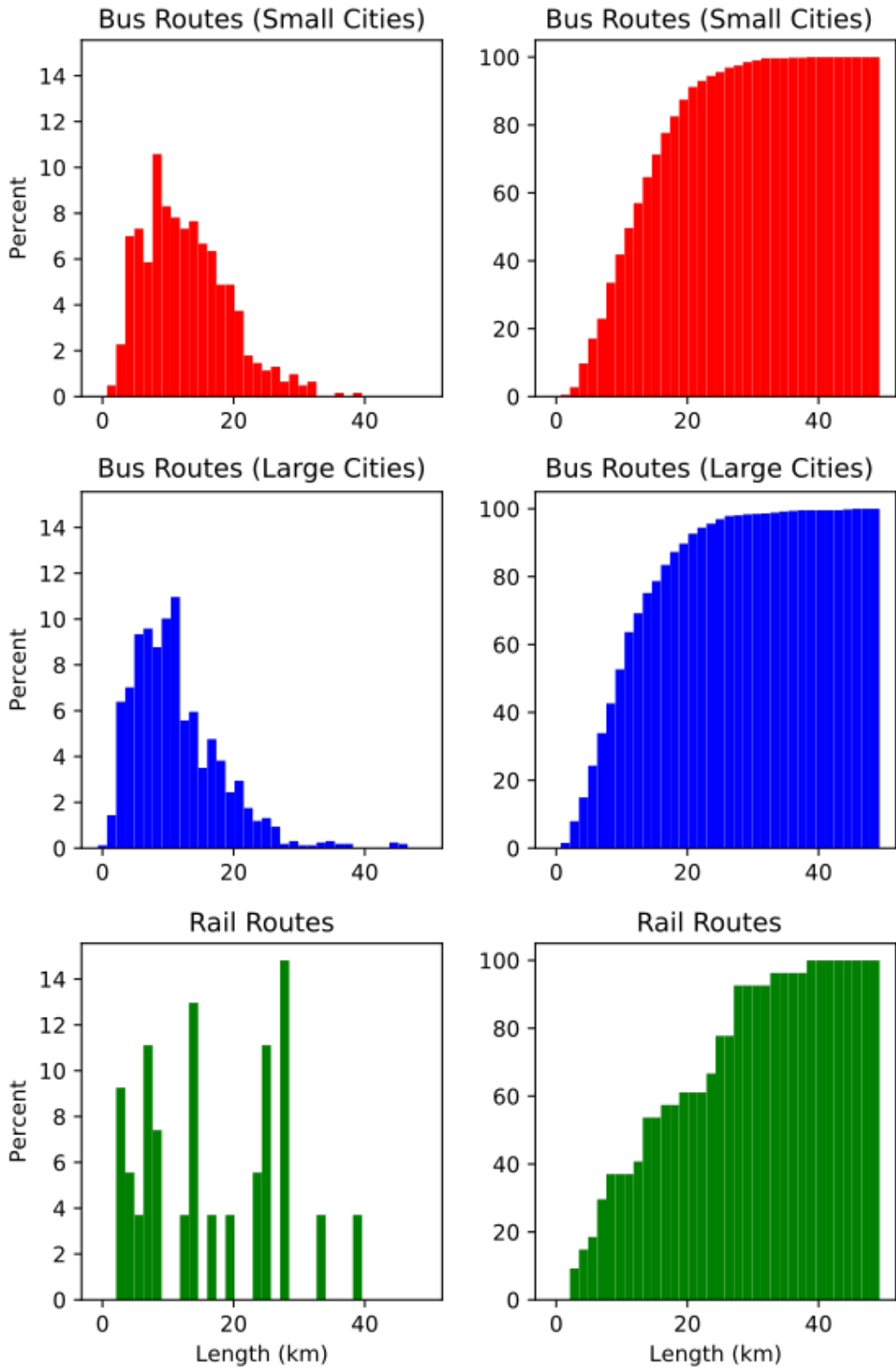


Figure 2.5: Distribution of Transit Route Length (km).

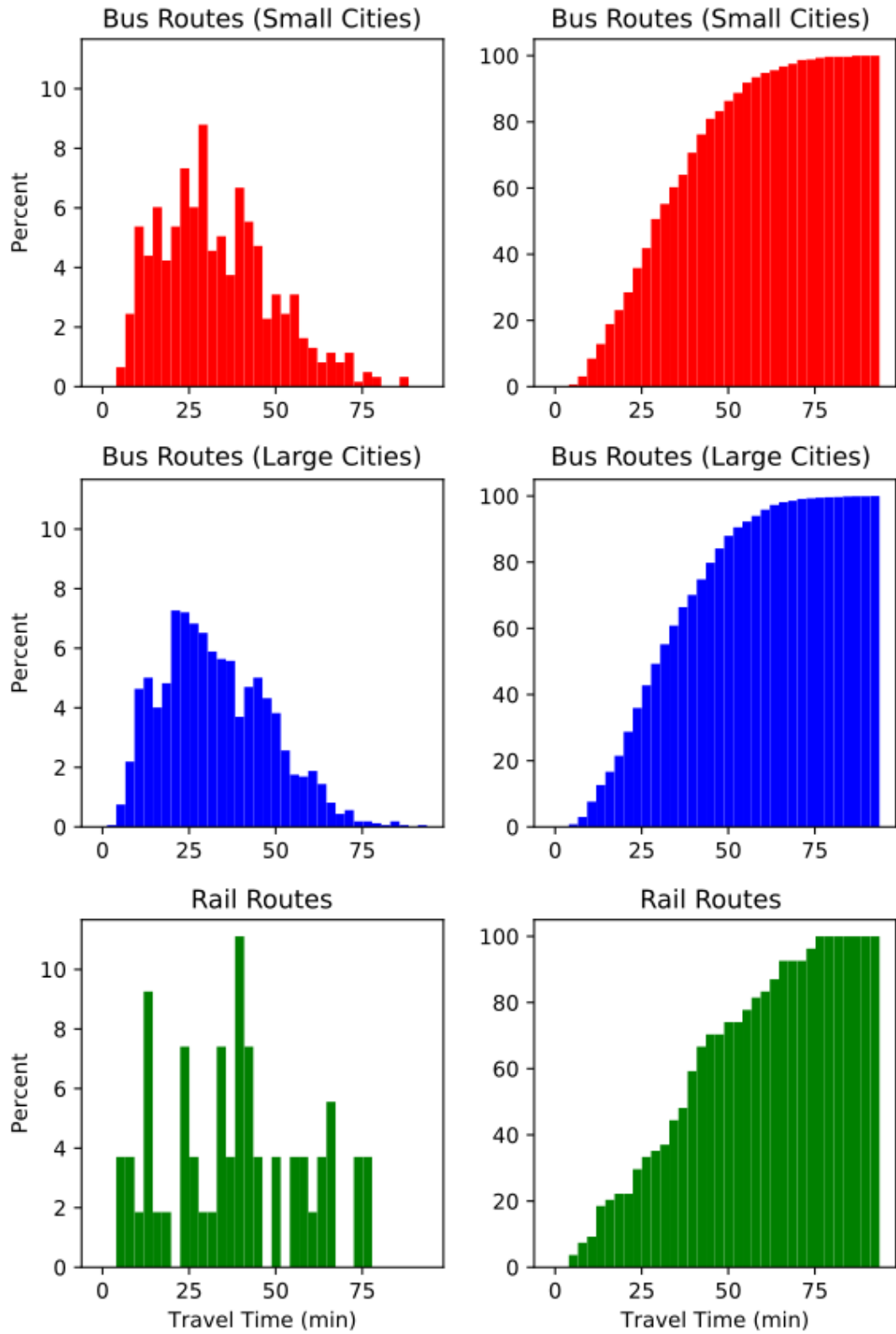


Figure 2.6: Distribution of Transit Route Travel Times (minutes).

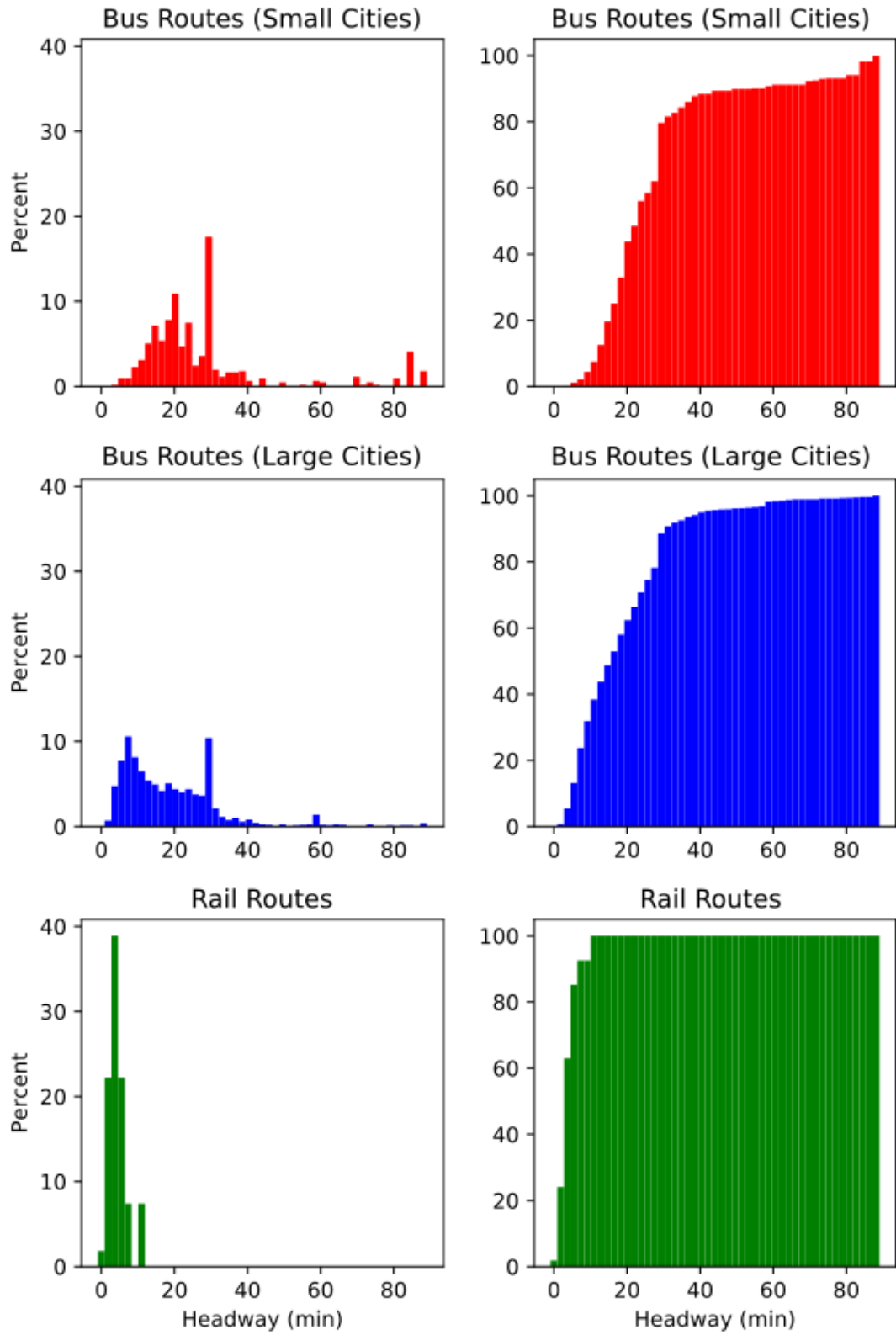


Figure 2.7: Distribution of Transit Route Headways (minutes).

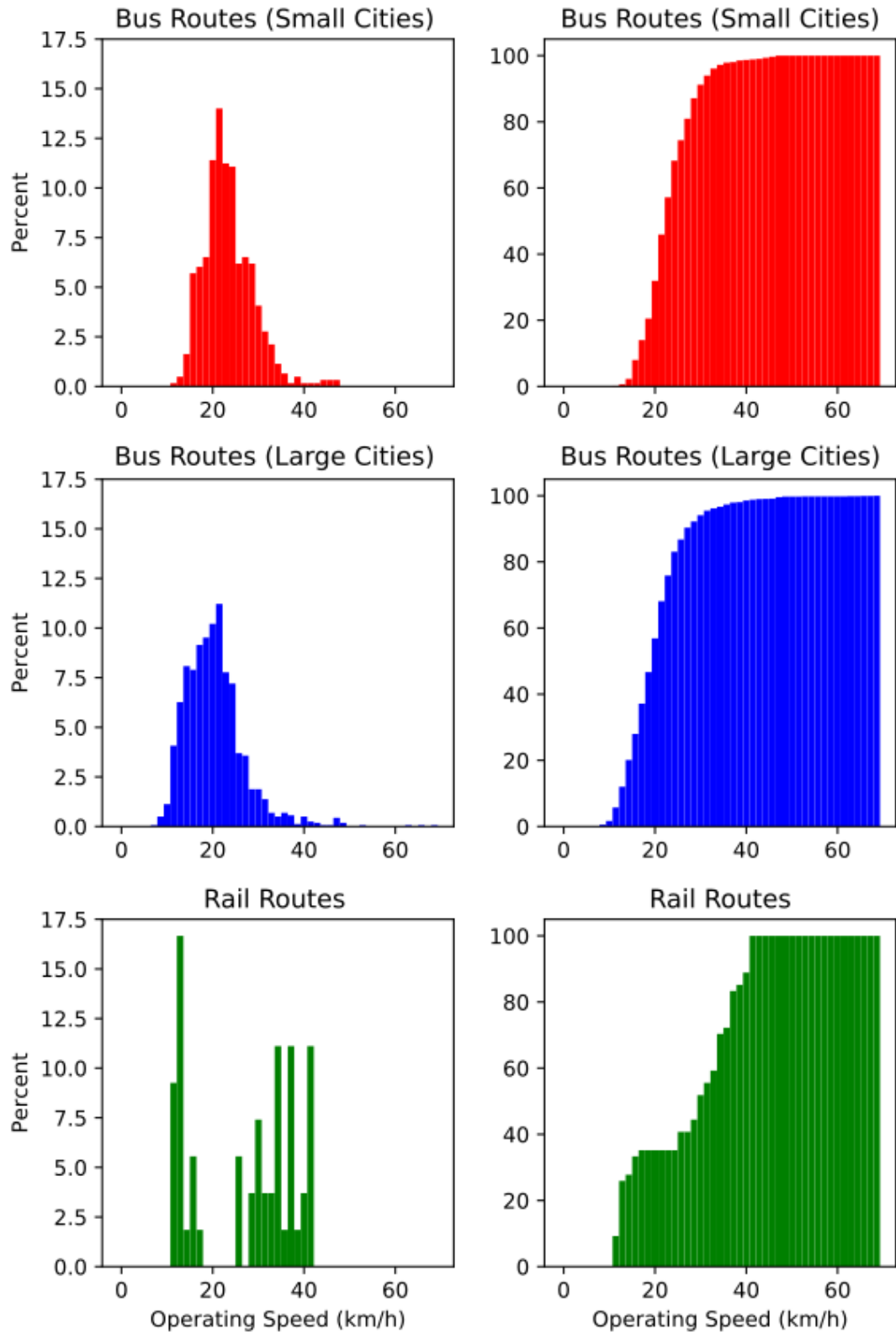


Figure 2.8: Distribution of Transit Route Operating Speeds (km/h).

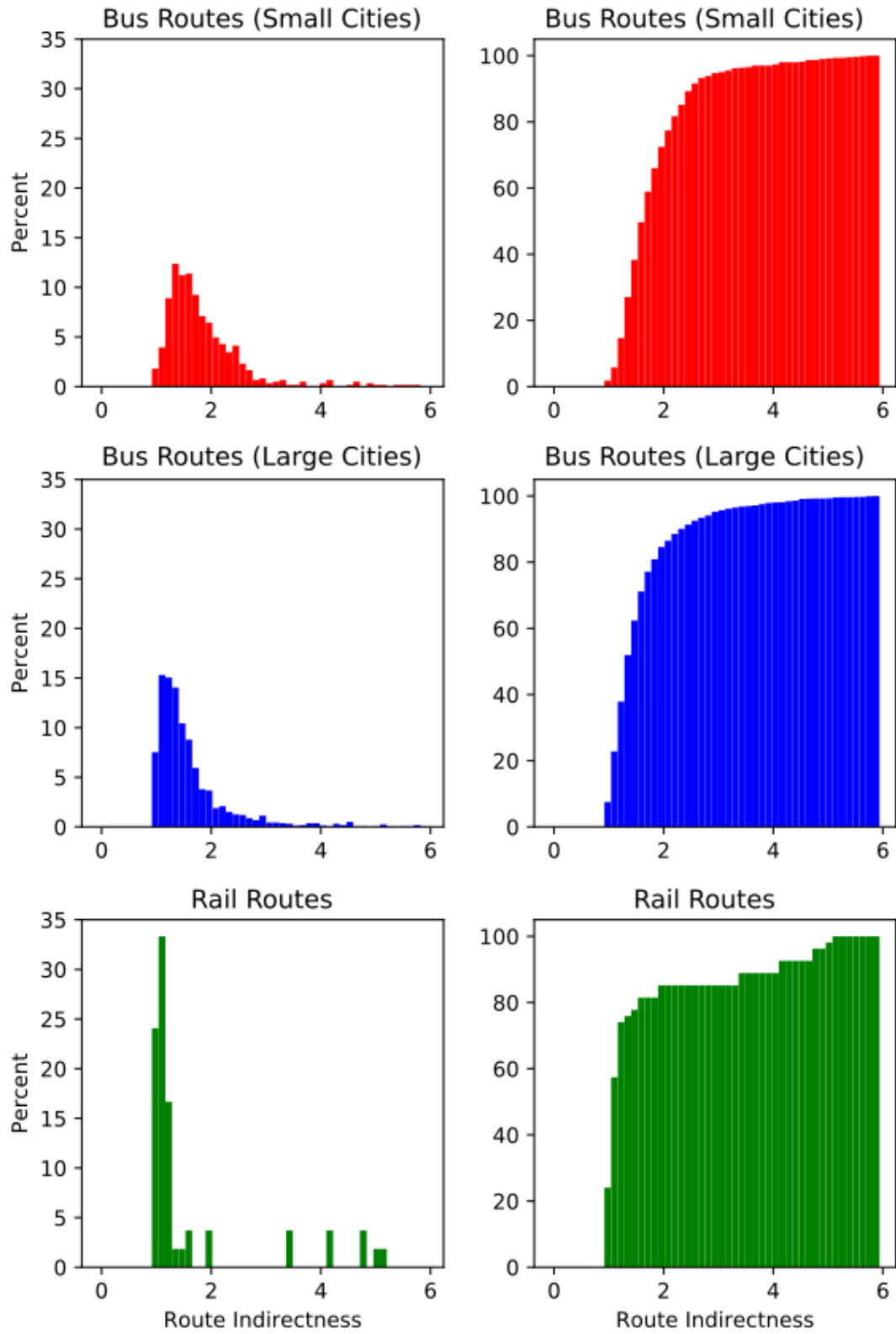


Figure 2.9: Distribution of Transit Route Indirectness.

2.1.2 Synthesizing Hypothetical Transit Routes

As described in the previous section, routes are a vector of four characteristics: headway, length, geodesic distance, and travel time. By combining data (vectors) from many different transit routes from urban centres across Canada, it is possible to create 4-dimensional empirical frequency distribution. The route table generator creates routes by drawing samples from the empirical data set. Kernel density estimation (KDE) was used to smooth the 4-dimensional distribution before sampling. Smoothing was used to eliminate gaps in the distribution that do not appear to have any theoretical justification, and to include vectors on the edges of the distribution that do not appear in the data set but could exist. Kernel density estimation was chosen as a simple way to create a smooth distribution, and because it can easily handle multiple dimensions.

Kernel density estimation creates a continuous distribution from a data set by placing a Gaussian (normal) distribution at each vector, with the Gaussian distributions being the same dimension as the vectors. Each Gaussian must have a standard deviation in each dimension, and by increasing the standard deviation, the distribution can be made smoother.

For the implementation of the KDE, the *scikit-learn* software package was used. This package includes the creation of the smoothed distribution, the selection of the standard deviations (more commonly called bandwidths in the context of KDE), and sampling from the created distribution. For more on kernel density estimation and the *scikit-learn* package, see VanderPlas (2016) and Pedregosa (2011). After samples are drawn, a basic filter is applied to ensure that the smoothing hasn't violated the following constraints:

$$l > 0 \tag{2.9}$$

$$g \geq b \tag{2.10}$$

$$t > 0 \tag{2.11}$$

$$h > 2.5 \tag{2.12}$$

$$l \geq g \tag{2.13}$$

2.2 Network Generator

Previously, a comparison was made between carrying out the demand modelling on city data collected from the real world or on data created through a model, with the conclusion being that model-created cities allowed more control, reduced the complexity of

fitting existing data to the exact requirements of this project, and did not sacrifice accuracy due to the goals and nature of the demand modelling. See the introduction to Chapter 3 for this discussion. The model that creates city data is called the network generator.

The job of the network generator is to create the city data that is required by the four-step travel demand model. In that model, cities are represented by:

1. Traffic analysis zone tables covering land use,
2. Link tables describing the road network and transportation options, and
3. Origin-destination matrices giving information such as travel times and costs for trips between those origins and destinations.

To create this data, the network generator has five components:

1. creating a road grid,
2. creating alignments of the transit routes,
3. giving properties to the traffic analysis zones,
4. assigning volume-delay functions for the road network, and
5. calculating origin-destination matrices.

In some cases, these components are directly creating the data required for the four-step model, and in other cases they are creating underlying structures that ensure an adherence to reality and can be converted into the four-step model inputs.

Before further describing each of these components, it is worth considering what properties of a city are important and the required level of detail for the output. For roads and the road network, it is not important to have very high-resolution data that includes the paths the roads take, and includes all the possible types of intersections and road patterns. Instead, what is important is that for any origin and destination within the network, there are a number of route options, of which some use higher-order infrastructure, some use transit, some are direct, etc. As well, some origin-destination pairs may not have many route options, due to being constricted by some sort of linear barrier such as bodies of water, changes in elevation, or transportation rights-of-way. This diversity of options creates a space for the travel demand model to assess the factors influencing mode choice decisions. Traffic analysis zones are a way to abstract the land use properties of the city. When creating TAZs for an existing city they are made descriptively, seeking to combine areas of similar land use and follow natural and human boundaries. In

this project, they are used prescriptively to create a spatial variation of land uses, and the actual physical shapes or sizes of the TAZs are not important. Since the physical layout of neither the roads nor the TAZs is important, the city is represented by a grid. This creates TAZs of equal area at each point on the grid, with roads between each. As real cities are not perfect grids, some roads are removed to introduce irregularities, and to account for linear barriers.

A visual summary of the components of the network generator is shown in Figure 2.10. These components are further described in the following sections.

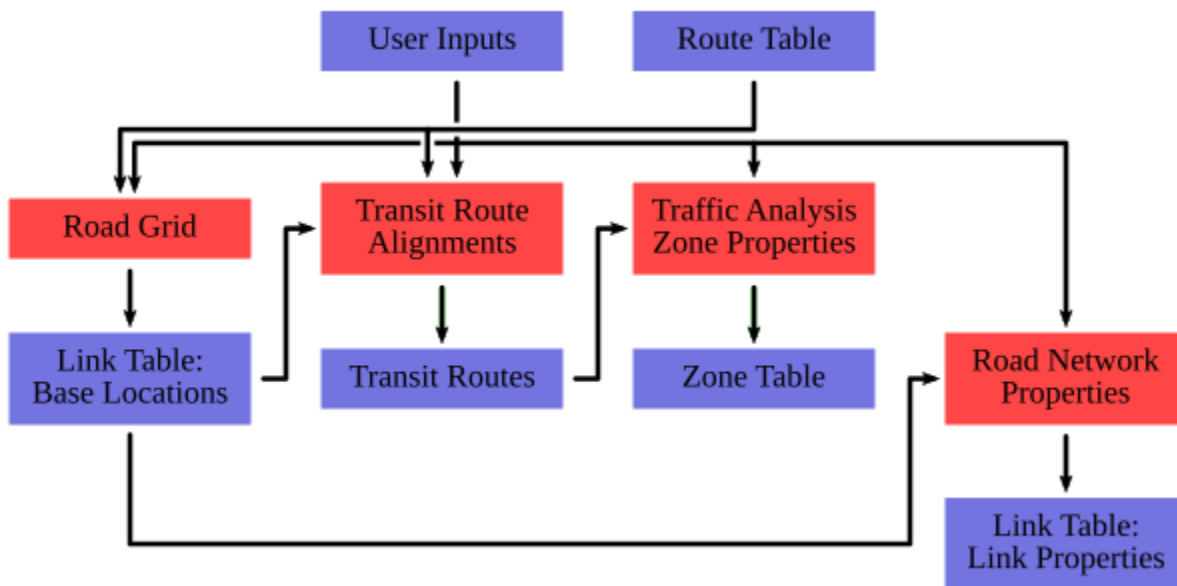


Figure 2.10: Network generator components

2.2.1 Road Grid

The road grid component produces a rudimentary link table containing information about how the links connect to make up the city. As discussed previously, the city is made up of a grid, shown in Figure 2.11. Each point on the grid, also referred to as the grid points, serve as the connection points for the links. Links here refers to the one-way connection from one grid point to the next, and a pair of links connecting the same grid points in opposite directions make up a (road) segment. This component uses the route table as an input, as well the following user inputs and parameters: minimum city size, block size, bridge gap mean, and bridge gap standard deviation. See Table 2.1 for more information on these parameters. Block size has a default value of 1000 m, chosen to reduce computation

time but maintain a resolution of TAZ similar to those used in the Transportation Tomorrow Survey (for more on the TTS, see Section 3.3.3). The bridge parameters' default values are 1.75 and 1.25 respectively, chosen through subjective analysis of bridge spacing carried out on maps of Canadian cities.

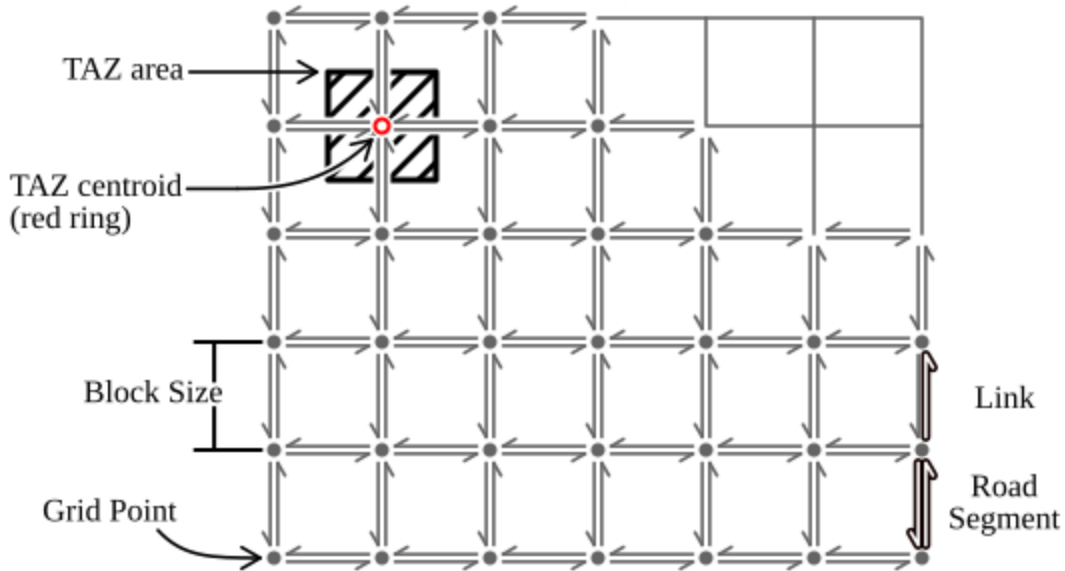


Figure 2.11: Visual definition of the road grid. Only one TAZ is shown for visual clarity. Similarly, in following figures the simplification in the top right corner is used.

The grid points also serve as the locations for the traffic analysis zones (TAZ, or zones). Defining the TAZ centroids in the same locations as the grid points reduces complexity and computation time, as it reduces the number of roads that have to be modelled. The size of the grid is $\max(n_{min}, n_g)$, where n_g is the number of blocks such that the width of the city is equal to the maximum geodesic distance of the transit routes. Dummy segments are used to connect the TAZ centroids to the nearest grid point, to model all the roads within zones. All road segments are made up of two links: there are no one way streets in the network.

Two adjustments are made to the base grid to introduce variation and route choice complexity in the road network: random removals and linear barriers. A small percentage of road segments are removed to simulate the effect of parks, large properties, and other irregularities. By default, 7% of segments are removed. Segments are eligible for removal if removing them does not create a dead-end segment, which are segments that serve as the

only connection between a TAZ and the rest of the city. Eligible segments are chosen to be removed randomly, each with an equal probability.

A linear barrier is also created to simulate the varieties of geographic barriers such as bodies of water, changes in elevation, and infrastructure rights-of-way. These types of barriers result in some segments being key routes as they magnify the differences between alternative routes and cause many origin-destination pairs to use the same key segment. Most often, these linear barriers are rivers, and the bridges are the key segments. To create a linear barrier, the network generator removes most links along a line passing through the city. To decide which links are kept and which are removed, the length of the gap, or the number of links removed between bridges, is defined by a normal distribution, using the bridge gap mean and standard deviation.

The output from this sub-model of the network generator is a road network topology. An example of a road grid with a linear barrier and randomly removed links is shown in Figure 2.12. Note that the classification of roadway types (e.g. arterial versus freeway) and associated capacity is determined in the fourth component as described in Section 2.2.4.

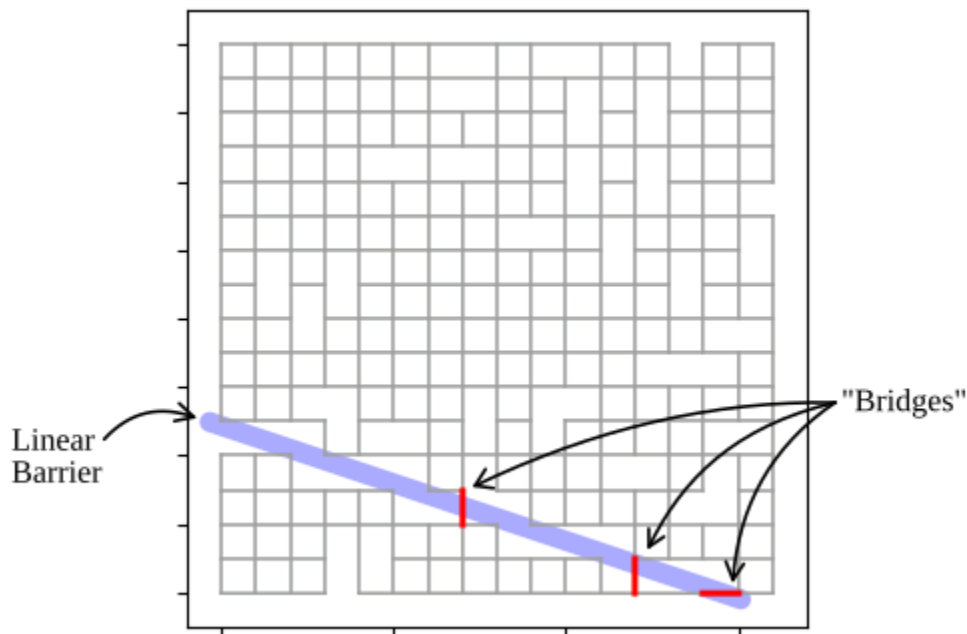


Figure 2.12: Road grid with segments removed.

2.2.2 Transit Route Alignments

The next step in the process is to determine transit route alignments. The user specifies the number of public transit (bus and rail) routes and the characteristics of each route is obtained from the route table (i.e. the output from the Route Table Generator). The Transit Route Alignments component generates alignments for the transit routes such that the route characteristics match those in the Route Table as closely as possible. In particular, care is taken to ensure that the geodesic distance and the route length match, since these are the measures that are most affected by the alignment of the route. The transit routes produced are represented by a list of the links that make up the route in the order they are traversed. It should be noted that due to the discrete nature of the road grid, the actual route characteristics may not exactly match the target characteristics as defined in the Route Table.

The route alignment generation is stochastic subject to the need to adhere to the target characteristics as obtained from the Route Table. It may seem unintuitive for the transit route alignment to be generated absent of an underlying structure considering that in reality transit routes are carefully placed to be useful to travellers. However, in this model, we wish to simulate a set of transit routes with prescribed characteristics (i.e. to represent those existing in Canadian cities) and therefore, we define the routes first, and then define TAZ population and employment densities to correlate to the transit route characteristics (See Section 3.2.3 for details).

Route alignments are created by choosing an origin and destination pair, finding the shortest distance path between those, and if needed, diverting the path to increase the indirectness. Origin destination pairs are chosen randomly from the set of valid pairs for that route. A pair is valid if it meets two conditions:

1. the Euclidean distance between the points is near the desired route geodesic distance, within a margin of error, and
2. the ratio between the Manhattan distance and the Euclidean distance is greater than the desired route indirectness.

The first condition ensures that the actual geodesic distance is close to the desired value, which is necessary due to the discrete nature of the road grid. The margin of error is

set to $b\sqrt{2}$, where b is the block size. This margin of error is the maximum Euclidean distance a point within a square with side length b can be from the square's vertices – using this margin of error is equivalent to making the assumption that, for any point, all four of the surrounding grid points are close enough approximations. The second condition ensures that it is possible to create a sufficiently direct route. The route must necessarily follow the roads, which are all horizontal or vertical and therefore the Manhattan distance is the shortest distance. If this is not short enough, then an origin-destination pair must be chosen that is more horizontal or more vertical.

The shortest path between the chosen origin and destination pair is an example of the shortest path problem found in graph theory, and solved using the *scipy.sparse.csgraph.shortest_path* function from the *scipy* software package. This function uses well defined algorithms such as Dijkstra's algorithm (Dijkstra, 1959).

The provisional route alignment created by Dijkstra's algorithm may not be indirect enough according to the input values from the route table. To increase indirectness, a diversion is added to increase the length of the route. Figure 2.13 shows an example of a route with a diversion added. To create the diversion, an arbitrary section from point a to b of the original route is selected to be removed. It is replaced by a longer section, found by offsetting the original route. A vector \vec{v} is found that is perpendicular to the vector \vec{u} , which goes from a to b . Vector \vec{v} is scaled to have a length of $k/2$, where k is the amount by which the route must be lengthened. Then the diversion section is created by finding the shortest path from a to $a + \vec{v}$ to $b + \vec{v}$ to b . In this way, the route is lengthened by approximately $2\|\vec{v}\| = k$. Implementing this process on the grid introduces a small unavoidable error.

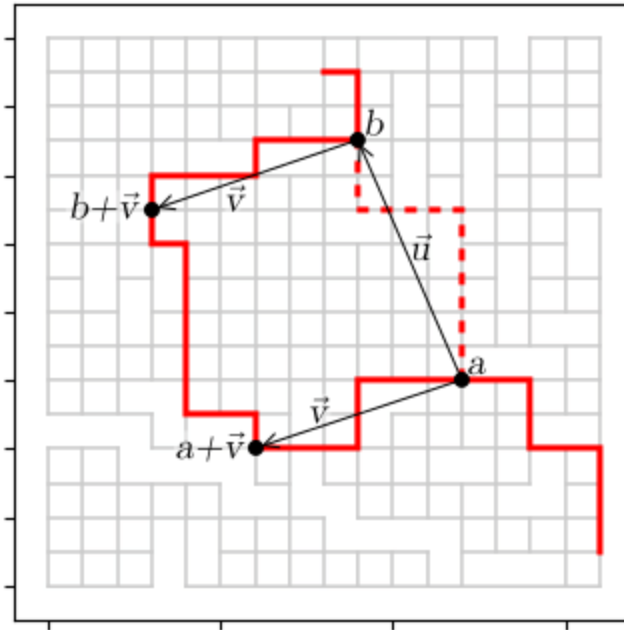


Figure 2.13: Example transit route alignment, with diversion.

2.2.3 Traffic Analysis Zone Properties

The outputs of the network generator model are used for the Four-Step Model. In particular, the trip generation step needs each zone to have a population and a number of jobs in that zone. Note that the jobs number is not the number of employed people living in the zone, but the number of employees that businesses and other employers located in the zone have.

The primary requirements of zone attribute assignment are that (1) high population and high employing areas should be served by a correspondingly high level of transit service, and (2) the amount of land at each population density level should be reasonable for a city of its size. To achieve the first goal, the placement of higher populations and higher number of jobs is biased towards zones with higher transit service. Further details are given in Section 2.2.3.2. To achieve the second goal, the population densities are drawn from real world data described in Section 2.2.3.1.

2.2.3.1 Distribution of Population Density

The population density dataset was obtained from the Canadian census. In order to include a variety of urban landscapes, a dataset was created using 12 of the 13 largest Census Metropolitan Areas (CMAs) in Canada:

| | |
|-----------|----------|
| Toronto | Québec |
| Montréal | Winnipeg |
| Vancouver | Hamilton |
| Calgary | Kitchen |
| Ottawa | London |
| Edmonton | Halifax |

These cities were chosen to capture most Canadian cities that are large enough for significant transit systems, as well as to ensure representation for the various regions of Canada (St. Catharines was excluded in favour of Halifax to include the Maritimes).

The Canadian census divides all of Canada into geographic regions on a range of different physical scales. The smallest region with data accessible to the public is the Dissemination Area (DA). The census defines a DA as a “small, relatively stable geographic unit composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons based on data from the previous Census of Population Program. It is the smallest standard geographic area for which all census data are disseminated. DAs cover all the territory of Canada.” (Dictionary, Census of Population, 2016). Additionally, DA boundaries follow roads as much as possible, but may also follow railways, water features, power transmission lines, and other features. They are designed to be uniform in population and to be compact in shape. Dissemination areas make up many larger geographic regions used by the census and by governments, such as census subdivisions (roughly equivalent to municipalities), census divisions (similar to counties, regional municipalities, and other first-level administrative divisions), and census metropolitan areas (adjacent municipalities with high economic and social integration).

The simplest way to obtain a distribution of population densities is to use the list of DAs. This creates a distribution in which each DA, regardless of population, area, population density, or any other attribute, counts the same. There are 28,654 DAs associated with the 12 CMAs included in the dataset. The distribution from this list of DAs is shown in Figure 2.14. Any zone (DA) with a population density above 30,000 people/km² is excluded, as those are extreme outliers that form only 1.13% of the data.

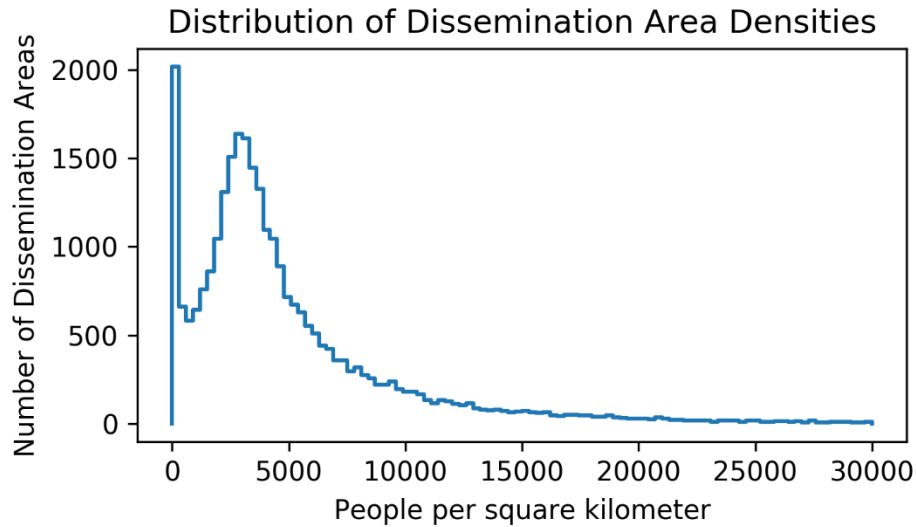


Figure 2.14: Distribution of dissemination area densities

However, this distribution is biased towards areas of high density. Due to the way that dissemination areas are defined (i.e. areas containing approximately 400 – 700 people), dissemination areas vary in terms of their size. For example a dissemination area can range from a single high-rise building, to two streets of suburban detached homes, to a large area of rural farms and homes. Conversely, the network generation model uses zones that have equal area but vary in terms of their population. Therefore, using the distribution above would result in more land being dedicated to high-density development than is actually found in the real world. In order to account for this, the census data from the dissemination areas was re-organized according to area as described below.

A sample of the raw data from the Population and Dwelling Count Highlight Tables of the 2016 census is shown in Table 2.10. Note that some of the columns have not been shown, such as the dwellings counts and the geographic codes for provinces, census divisions, and census subdivisions.

Table 2.10: Sample census data (Population and Dwelling Count Highlight Table)

| DA Code | Province | Population | Area km ² | Population Density people/km ² |
|----------|----------|------------|-------------------------|--|
| 24230057 | Quebec | 301.0 | 0.19 | 1616.5 |
| 24230633 | Quebec | 777.0 | 0.48 | 1615.0 |
| 24230860 | Quebec | 383.0 | 0.24 | 1610.6 |
| 35212143 | Ontario | 4682.0 | 2.59 | 1804.9 |
| 35210226 | Ontario | 635.0 | 0.35 | 1801.9 |

In order to produce a distribution that accounted for area equally, the DAs were put into bins according to population density. The bins were 10 people/km² wide. For each bin, the total land area was summed. An example of the output of this method, applied to the data above, is shown in Table 2.11.

Table 2.11: Amount of land by population density (example)

| Population Density Bin (people/km ²) | Dissemination Area (count) | Total Land Area (km ²) |
|--|----------------------------|------------------------------------|
| 1610-1620 | 3 | 0.91 |
| 1800-1810 | 2 | 2.94 |

Plotting the distribution of this data produces Figure 2.15. It is clear from this distribution that despite the dataset being composed of the largest CMAs in Canada and including over 20 million people, most of the land area encompassed by these 12 CMAs has very low population density. In fact, over 75% of the land area has a population density of 50 people/km² or lower. Some of this low density area is urban, such as parks, industrial areas, and transportation facilities. However, it also includes rural areas within municipal and CMA boundaries; these rural areas are not an area of interest for this research. The dataset clearly requires a filter to reduce the amount of low density area.

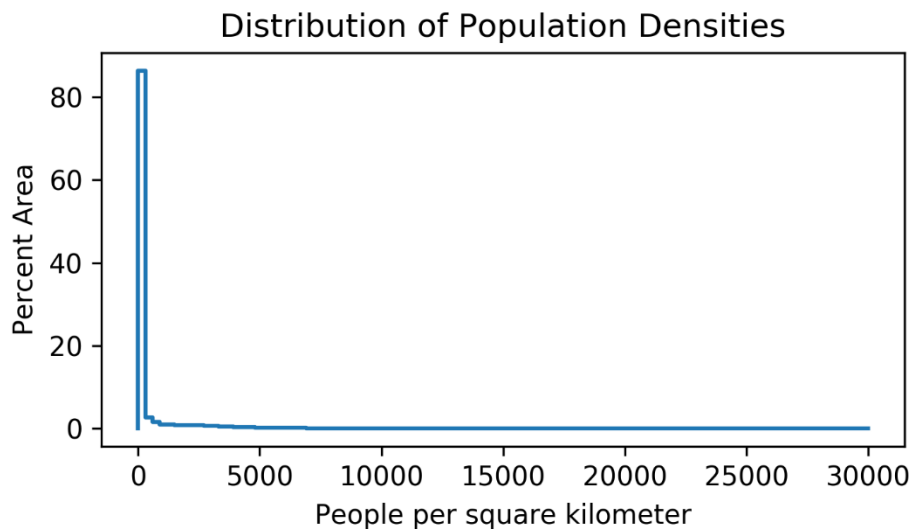


Figure 2.15: Distribution of population densities

In order to exclude rural land, the filter will need to operate on the scale of the geographic divisions that make up census metropolitan areas: census subdivisions and dissemination areas. Census subdivisions use the same boundaries as the municipalities

and municipality equivalents defined by provinces. These are administrative boundaries and therefore not designed to clearly delineate urban and rural areas. In fact, many municipalities in Canada include a ring of surrounding farmland. Therefore, the filter must operate on the DA area level.

The dataset has 2 attributes for each DA that are relevant for identifying low density areas: population and area. From these the population density can be obtained. Unfortunately, the nature of the dataset is such that the lowest population density threshold that eliminates the low density spike in the figure above is 500 people/km², which is too high to be sure that all suburban areas are included. As well, filtering out all low density DAs is not desirable regardless of the threshold as cities do contain unpopulated land, such as parks, industrial areas, and airports.

As a filter is not practical, a compromise is used: with respect to population, the network generation model is most concerned with having the population densities at the suburban level and above be in accordance with the distribution above. The amount of low density area, especially low density area that surrounds the city, is effectively immaterial. Therefore, the distribution from 0-300 people/km² was capped at the value of the distribution at 300 people/km² (effectively “cutting off” the spike). This allows for the model to include these densities without them overwhelming the density ranges that are of more interest. Figure 2.16 and Figure 2.17 show the resulting distribution.

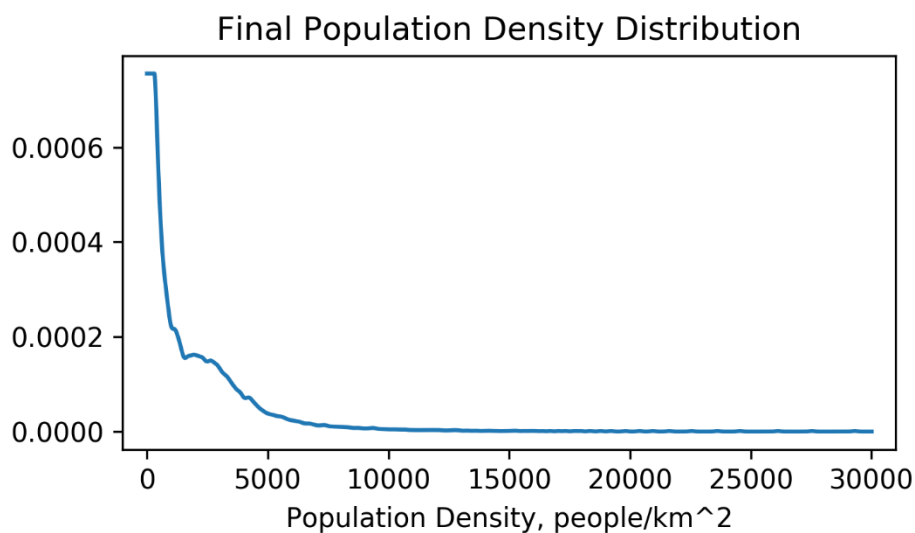


Figure 2.16: Final population density distribution

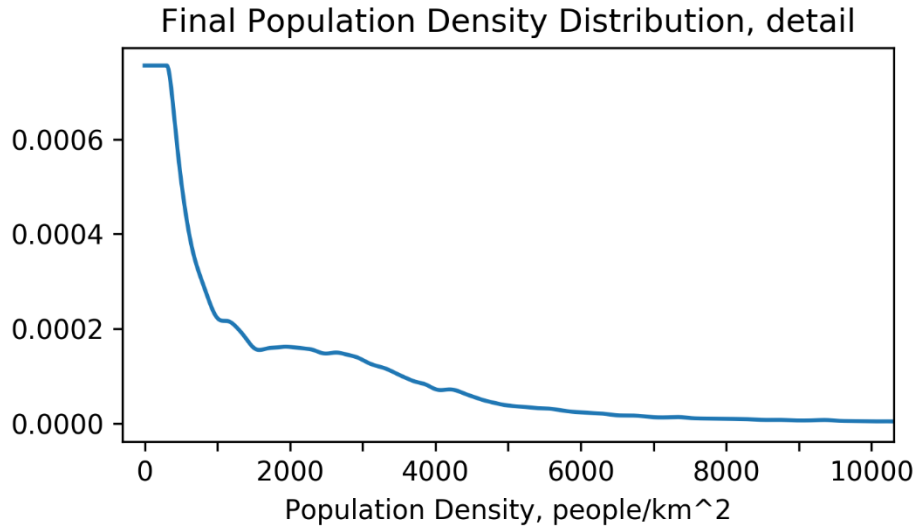


Figure 2.17: Final population density distribution, detail

2.2.3.2 Attribute assignment

Zone attribute assignment makes use of the following input data: the locations of the traffic analysis zones (TAZs); the number of TAZs n ; the paths and hourly capacity (passengers per hour) of the transit routes; a user specified target city population; a fill density (people/km²); an employment-to-population ratio; and the population density distribution. The target population, in people, is a user input to the entire modelling structure, and can be used to create different scenarios. The fill density is the population density of a typical rural area, and has a default value of 30 people/km². The output of zone attribute assignment process is a population and employment value (units of number of people) for each TAZ in the modelled urban area.

The algorithm itself consists of two stages. The first produces n population densities and n employment densities. The second takes those densities and assigns them as counts to the n TAZs.

To create the set of population densities, an initial sample vector $\vec{L} = \langle \ell_1, \ell_2, \dots, \ell_n \rangle$ is drawn from the population density distribution. This sample can easily be converted from population densities to population using the area of the TAZ, which are all equal. Equation (2.14) shows how the population of the city can be calculated from \vec{L} .

$$P_A = \sum_{i=1}^n \ell_i \cdot b^2 = nb^2 \sum_{i=1}^n \ell_i \quad (2.14)$$

where P_A is the actual population of the city, n is the number of TAZ, b is the block size (in km), and ℓ_i are the population densities of the TAZ.

It is likely that P_A is not going to be equal, or even close to, the target population P_T . The vector \vec{L} is then adjusted: either lowered by replacing some elements of \vec{L} with the fill density, or raised by replacing elements of \vec{L} with greater values also drawn from the census population density distribution. If P_A needs to be lowered, a new vector $\vec{M} = \langle \ell_1, \ell_2, \dots, \ell_k, f, \dots, f \rangle$ is created which consists of the first k elements of \vec{L} , and the remaining $n-k$ values are the fill density f . The value of k is chosen so that $nb^2 \cdot \text{sum}(\vec{M}) = P_A$ is minimized and so that $P_A \geq P_T$.

If P_A needs to be increased, then a secondary sample of $2n$ population densities, \vec{S} , is created. For each element s of \vec{S} , a random value ℓ of \vec{L} is selected that is less than s . Assuming such a value exists, ℓ is replaced by s . In this way the actual population $P_A = nb^2 \cdot \text{sum}(\vec{L})$ is increased. This is an iterative process, and once $P_A \geq P_T$, the process is finished. If the process never finishes, then the entire model is run again with a larger size n until the total population is greater than the goal population.

Once a vector \vec{L} of n population density values is acquired, the n population density values must be assigned to the n zones. For each zone i , the total transit capacity per hour C_i is summed from the transit routes that service the zone. \vec{L} is sorted from greatest to smallest, then assigned to zones in that order. Each zone is selected randomly, using a weight proportional to $C_i^{1.75}$. In this way, higher populations are more likely to be assigned to zones with more transit service.

The entire process is repeated independently to acquire employment values.

2.2.4 Road Network Properties

Each link in the network must have the properties necessary to model how the link travel time responds to traffic demands. This relationship is modelled through the BPR volume-delay function shown in equation (2.15), where t_0 is the free speed travel time on the link in minutes, v is the volume on the link in vehicles per hour, c is the capacity of the link in vehicles per hour, α and β are parameters, and t is the travel time in minutes on the link at volume v . Therefore, the result of this component of the network generator is that

the link table, which was created in the road grid component (Section 2.2.1), has the addition information required for equation (2.15).

$$t = t_0 \left(1 + \alpha \left(\frac{v}{c} \right)^\beta \right) \quad (2.15)$$

To simplify the assignment of these values to the road network, four profiles of different link types were created: highways, two tiers of arterials, and urban local roads. Each profile is a set of values for the uncongested travel time, the link capacity, the two parameters α and β , and the number of lanes. The capacity is given as vehicles per hour across all lanes, which makes the inclusion of the number of lanes superfluous, though useful to include as an aid in understanding what type of infrastructure each of the profiles represents. Separate profiles are needed for the automobile and bus transit modes, as these two modes have differing characteristics with respect to speed and acceleration. A separate profile is not needed for the AVRS mode, as it is assumed to behave similarly to human-driven automobiles. Additionally, in this research rail transit is assumed to operate on a separate right-of-way and therefore suffers no congestion effects. The automobile profiles for the link types are shown in Table 2.12, and the equivalent profiles for bus transit are derived in Section 0.

Table 2.12: Automobile-related parameters of link profiles

| Link Type | Link Capacity (veh/h) | Free Speed (km/h) | α_a | β_a | Lanes |
|-------------|-----------------------|-------------------|------------|-----------|-------|
| Highway | 6000 | 100 | 0.1500 | 5.000 | 3 |
| Arterial1 | 1568 | 52.51 | 0.9841 | 4.395 | 3 |
| Arterial2 | 984 | 47.58 | 0.8800 | 4.771 | 2 |
| Urban Local | 420 | 41.31 | 0.8160 | 4.590 | 1 |

It should be noted that a free speed is provided in place of a free speed travel time as speed is independent of block size.

The values in Table 2.12 for the highway link type were obtained from course notes from Dr. Fu's CIVE 640 course (Fu, 2016).

2.2.4.1 Derivation of Link Profile Parameters

The three non-highway link types are assumed to have a mix of signal-controlled and stop controlled intersections. For this type of road, the intersections have lower capacities than the midblock portions of the road, and as a result congestion occurs at the intersections. In order to reflect this, the parameters in Table 2.12 were obtained by fitting

equation (2.15) to the signalized intersection approach delay model found in the Canadian Capacity Guide (Teply, 2007, Section 4.8.1). This model, shown in equations (2.16) to (2.18), gives the average approach delay (d , in seconds) as a function of the approach volume λ , in passenger car units per lane per hour (pcu/l/h). Additional parameters are the approach green time g in seconds, the cycle time c in seconds, the saturation flow rate μ in pcu/l/h, the evaluation time period t_e in minutes, and a progression factor k which is dependent on other nearby intersections. Since the model is being used to represent wide range of situations, the progression factor k is assumed to be 1.

$$d_1 = \frac{c \left(1 - \frac{g}{c}\right)^2}{2 \left(1 - \frac{g}{c} \min\left(1, \frac{\lambda c}{\mu g}\right)\right)} \quad (2.16)$$

$$d_2 = 15t_e \left(\frac{\lambda c}{\mu g} - 1 + \sqrt{\left(\frac{\lambda c}{\mu g} - 1\right)^2 + \frac{240\lambda}{t_e} \left(\frac{c}{\mu g}\right)^2} \right) \quad (2.17)$$

$$d = kd_1 + d_2 = d_1 + d_2 \quad (2.18)$$

While the CCG model could have been used directly as the volume-delay function, its more complicated form led to problems in determining the values for bus transit mode. The derivation of the bus transit parameters is described in Section 0.

Equation (2.18) gives delay, but the BPR function (equation (2.15)) gives a travel time. To reconcile this, a free speed between intersections was assumed and used to find the travel time t between intersections in seconds, as shown in equation (2.19), where b is the block size in meters (see Table 2.1) and v_f is the free speed in km/h.

$$t = \frac{3.6b}{v_f} + d \quad (2.19)$$

The parameters for equations (2.16) to (2.19) were assumed as part of the link type definitions, and are shown in Table 2.13. The capacities in Table 2.12 were calculated as $n_l \mu g \div c$ (the capacity in the CCG model) where n_l is the number of lanes. The free speeds were found using the travel times at an approach volume of 0 pcu/l/h. Finally, α and β were calibrated using nonlinear least squares to values obtained from equation (2.19) on a domain of 0 pcu/l/h to $\mu g \div c$ pcu/l/h.

Table 2.13: Parameters for CCG signalized intersection approach delay model

| Parameter | Symbol | Arterial1 | Arterial2 | Urban Local |
|--------------------------------|--------|-----------|-----------|-------------|
| Block Size (m) | b | 1000 | 1000 | 1000 |
| Free Speed (km/h) | v_f | 80 | 60 | 50 |
| Free Speed Travel Time (s) | | 45 | 60 | 72 |
| Saturation Flow Rate (pcu/l/h) | μ | 1400 | 1200 | 1000 |
| Lanes | n_l | 3 | 2 | 1 |
| Cycle Time (s) | c | 120 | 90 | 90 |
| Lost Time per Cycle (s) | | 8 | 8 | 6 |
| Approach Green Time (%) | | 40 | 45 | 45 |
| Green Time (s) | g | 44.8 | 36.9 | 37.8 |
| Evaluation Time (min) | t_e | 30 | 30 | 30 |

2.2.4.2 Derivation of Bus Transit Parameters

To find the bus transit parameters, equation (2.15) can be rewritten as a function of the v/c ratio, which is common to both buses and automobiles:

$$t_m \left(\frac{v}{c} \right) = t_{0,m} \left(1 + \alpha_m \left(\frac{v}{c} \right)^{\beta_m} \right) \quad (2.20)$$

where m signifies mode and has a value of either a for automobile or b for bus. In order to solve for α_b and β_b , the following assumptions are used:

1. At a sufficiently high congestion level, the bus delay due to passenger loading and unloading becomes small enough relative to the delay due to congestion that buses and automobiles travel at effectively the same speed. Algebraically, this means that $t_b(x) = t_a(x)$ for all $x \geq M$. M is a constant greater than 1.
2. Buses respond to congestion in a manner mostly similar to that of automobiles, meaning that the general shapes of the automobile and bus BPR functions should be similar. This is achieved algebraically by ensuring that

$$\frac{t_a(F)}{t_{0,a}} = \frac{t_b(F)}{t_{0,b}}$$

for some value of F chosen to maintain the general shape.

3. The free speed travel time for buses should be longer due to the need to stop for passenger loading and unloading.

Using equation (2.20) with assumptions 1 and 2 creates a system of equations that can be used to solve for α_b and β_b :

$$\beta_b = \log_F \left[\frac{t_{0,b}(t_{0,a}(1 + \alpha_a F^{\beta_a}) - t_{0,a})}{t_{0,a}(t_{0,a}(1 + \alpha_a M^{\beta_a}) - t_{0,b})} \right] \quad (2.21)$$

$$\alpha_b = \frac{t_{0,a}(1 + \alpha_a M^{\beta_a}) - t_{0,b}}{t_{0,b} M^{\beta_b}} \quad (2.22)$$

Choosing both F and M is a somewhat subjective matter. A value of 0.5 was chosen for F to ensure that the automobile and bus functions grow at a similar rate for the v/c ratios that result in little congestion. For M , a different value was chosen for each link type: 2 for highway, 1.5 for Arterial 1, 1.65 for Arterial 2, 1.7 for Urban Local. These values were chosen because at these values, travel times are significantly increased (by a factor of roughly 5 from the free speed travel times) to the point that traffic is heavily congested and buses will be primarily boarding and alighting passengers when already stopped in a queue.

Finally, free speed travel times must be found for the four link types in Table 2.12. An average stop spacing (p , in km/stop), dwell time (d , in seconds), and reduction of travel speed taking into account acceleration and deceleration (r , in a unitless proportion) were assumed for each of the link types. These values were used to find the bus transit free speed travel time, $t_{0,b}$, using equation (2.23).

$$t_{0,b} = \frac{bd}{60000p} + \frac{t_{0,a}}{r} \quad (2.23)$$

where b is the block size in meters. The resulting link profiles are shown in Table 2.14, with a block size of 1000 m.

Table 2.14: Bus transit-related parameters of link profiles

| Link Type | r | p km/stop | d s | $t_{0,a}$ min | $t_{0,b}$ min | α_b | β_b |
|-------------|------|----------------|----------|------------------|------------------|------------|-----------|
| Highway | 0.95 | 4 | 60 | 0.600 | 0.8816 | 0.1332 | 4.468 |
| Arterial1 | 0.9 | 1.5 | 30 | 1.143 | 1.603 | 0.7598 | 4.022 |
| Arterial2 | 0.9 | 0.4 | 30 | 1.261 | 2.651 | 0.5326 | 4.047 |
| Urban Local | 0.9 | 0.4 | 30 | 1.452 | 2.863 | 0.5219 | 3.945 |

Plots of equation (2.20) with each of the 8 sets of $t_{0,m}$, α_m , and β_m parameters are shown in Figure 2.18.

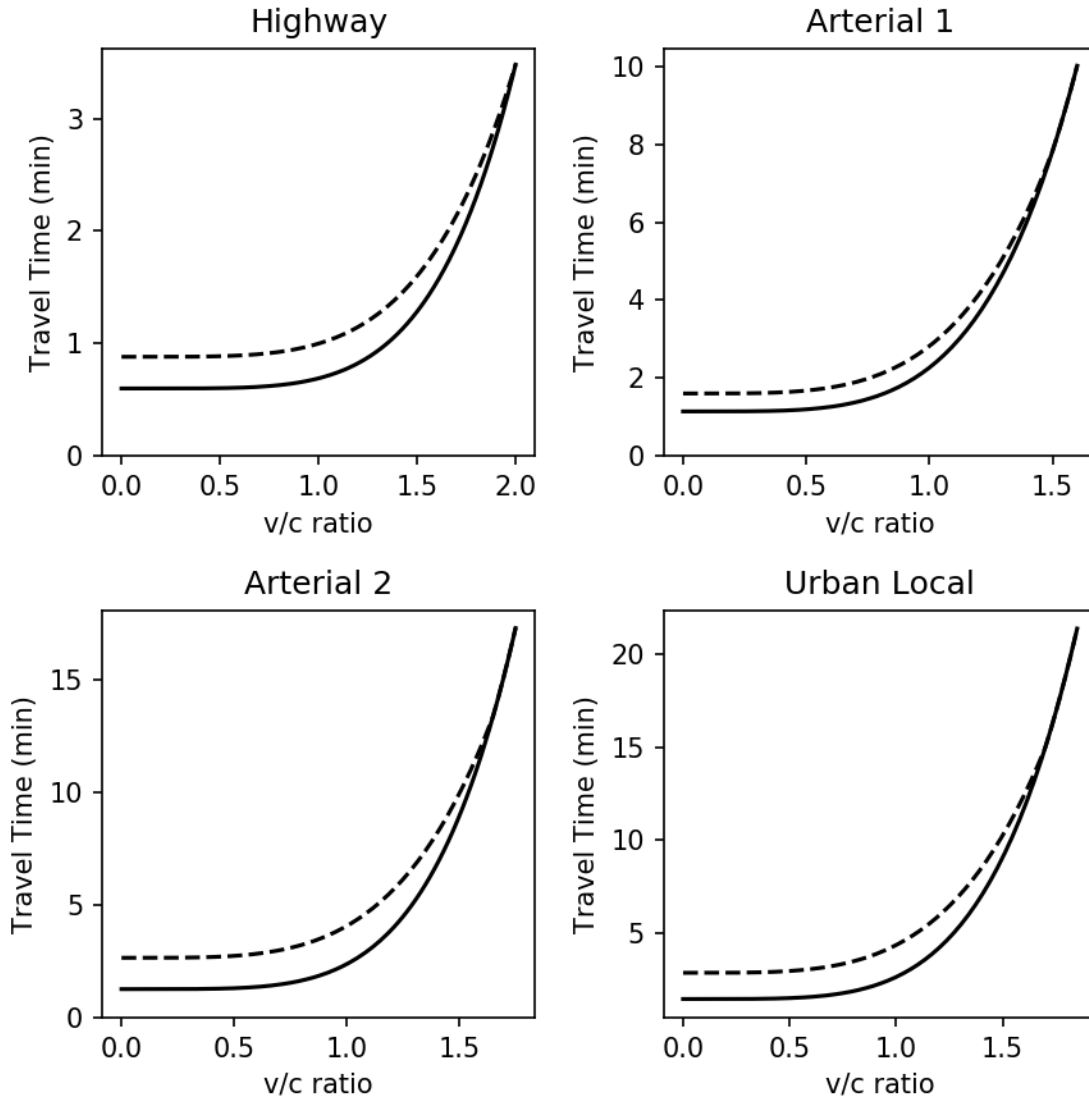


Figure 2.18: BPR functions for automobile and bus transit. The solid line is automobile, the dashed line is bus.

2.2.4.3 Assignment of Link Types to Links

The set of links belonging to each profile is determined in sequence, starting from highways and moving down through to the urban local roads. The highways are created following the spoke and arc structure common to Canadian cities. The two tiers of arterials are created using a grid structure, and the urban local roads are defined as any road that isn't a higher order. Both the highways and the arterials can be thought of a collection of paths – for the highways the paths are the individual spokes and the arc, and for the arterials the paths are the generally straight grid lines. The route of these paths is created

using a recursive depth-first search, which is an efficient way of finding a fairly direct path between a given starting point and a set of ending points. Vector fields are used to guide the depth-first search in creating the desired geometries: at each recursion, there is an option of what order to follow the links leading from the current location, and the order is determined by how well each link aligns with the vector field.

For the highways, 2-4 spokes are created starting from the centre of the city and travelling to the edge. These use vector fields that evaluate to the same vector at all locations, resulting in linear paths. The number of spokes is chosen by a uniform random variable, and the angle of the spokes are chosen to be roughly evenly spaced. The arc part of the highways can range from non-existent to a full circle, but the arc always starts and ends at an intersection with a spoke. Again, the length of the arc is chosen by a uniform random variable. To create the path of the arc, a vector field such as the one in Figure 2.19 is defined, centred on the centre of the city and with radius chosen with a random uniform variable ranging from 2000 meters to half the size of the city (meaning the arc would have a diameter equal to the width of the city).

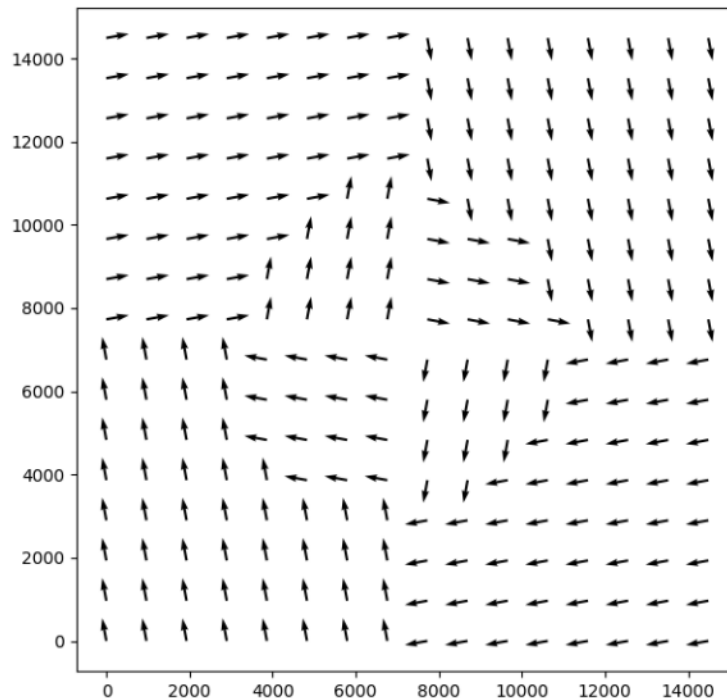


Figure 2.19: Vector field used to create a circular path.

Each tier of arterial has a grid that is tilted by a random angle. The two tiers share the same angle, but have different spacings: the default values are 7000 m for arterial 1 and 1000 m for arterial 2. The grids are created by defining appropriately spaced starting points along the edge of the city, and finding paths from the starting points to the edge of the city using the recursive depth-first search. The vector fields for the arterials evaluate to the same vector at all locations, giving the linear structure that makes up a grid.

An example of a city with the different road structures is shown in Figure 2.20.

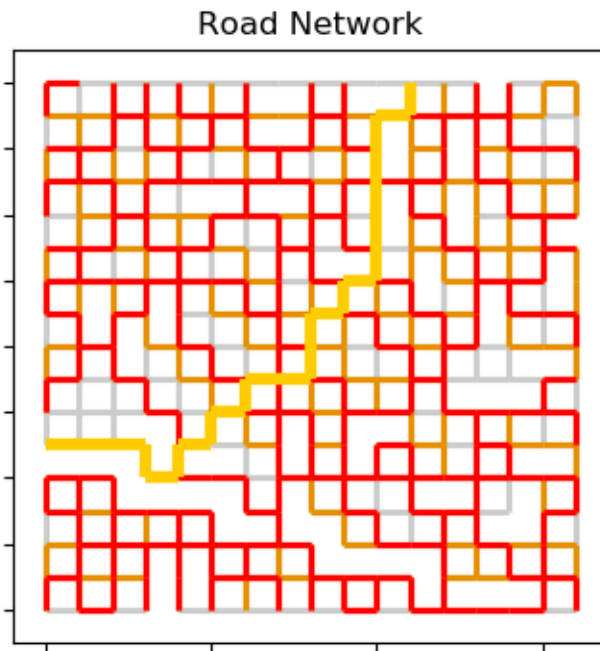


Figure 2.20: City map showing an example configuration of road types. The thick yellow roads are highways, the red is arterial1, the orange arterial2, and the grey urban local.

2.2.5 Origin-Destination Matrix Calculation

At the completion of the first four network generator components described in sections 2.2.1 to 2.2.4, the network is represented by a combination of three data structures: a zone table, a link table, and the transit routes. The transit routes are sequences of links in the order they are visited. The four-step model, described in section 2.3, requires data to be in the format of origin-destination matrices. An origin-destination matrix is an $n \times n$ matrix, where n is the number of TAZ, such that entry $A_{i,j}$ of the matrix is some measure of the trip from zone i to zone j . In particular, the four-step model requires an origin-destination matrix for the costs and for the 3 types of travel time. More explicitly,

in order to be used in the Mode Split stage of the four-step model each of the variables in the utility functions in equations (2.44)-(2.46) requires an origin-destination matrix.

The calculation of these matrices is done following the principle that travellers choose their path in order to minimize the time spent travelling. In other words, while a given origin-destination pair may have a large number of different paths, whichever path has the shortest total travel time is the path that will be used for the calculation of the origin-destination matrix. In order to find these shortest travel time paths, a weighted, directed graph is constructed for both automobile and transit travel. AVRS travel uses the same graph as automobile travel. The edge weights for the automobile graph are the automobile travel times along each link taken directly from the link table, and in fact the automobile graph is merely a reformatting of the link table. With such a graph structure, the fastest route between any origin and destination can be found using well established graph theory algorithms such as Dijkstra's algorithm.

The construction of the weighted, directed graph for transit travel is more complicated. It must include walking access and egress, transfers between transit routes, and wait times during the use of the transit services. These factors are accounted for by creating a graph which contains multiple graph nodes for a single TAZ: one node for the TAZ, and one node for each transit route that passes through that TAZ. An example of the structure of this graph is shown in Figure 2.21. In it, there are nodes belong to the TAZ in grey and nodes belonging to the Red and Blue routes in red and blue. The edges are coloured in the same manner, with the addition of dark red and dark blue edges that connect the transit route nodes to the TAZ nodes. (Not all the possible connecting edges are shown, only those included in the example trip. Not shown are the pair of connecting edges between each transit node and the corresponding TAZ node.) Each of the grey "road" edges has a travel time based on a walking speed of 1.15 m/s. The red and blue transit edges have travel times according to equation (2.20). The dark red and dark blue edge that are going to the TAZ nodes have a travel time of 0 minutes, and those going to the transit nodes have a travel time equal to the waiting of that transit route.

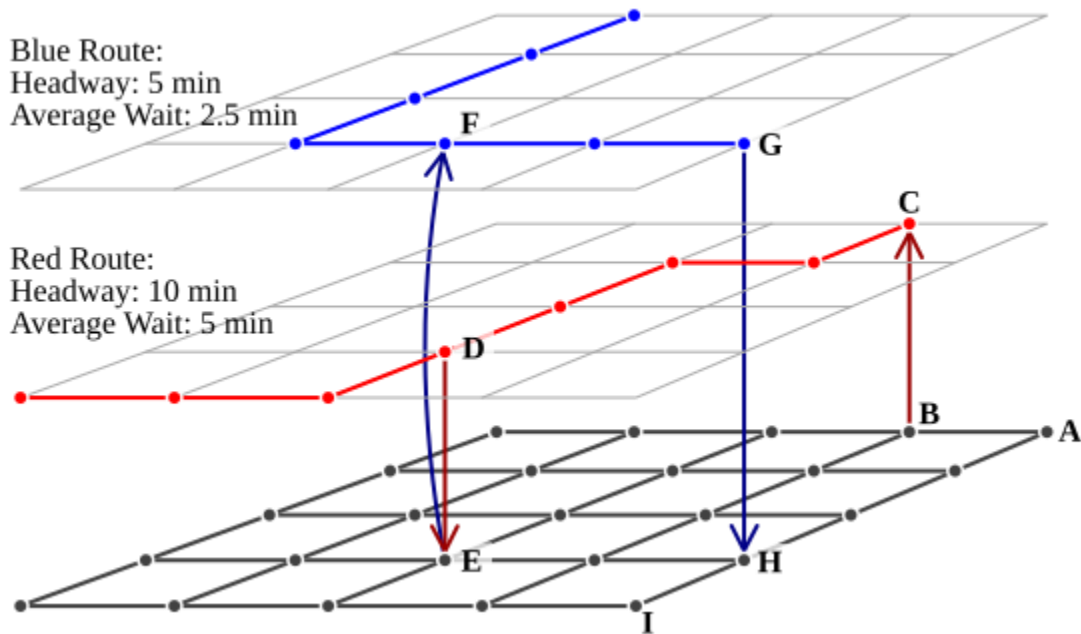


Figure 2.21: Structure of the graph used to calculate transit origin-destination matrices.

For example, consider the transit trip from A to I, assume a block size of 1 km, and an average transit speed of 24 km/h. Edges AB and HI have a weight of 14.5 minutes of walking time, BC has a weight of 5 minutes waiting time, EF has a weight of 2.5 minutes waiting time, edges DE and GH have weights of 0 minutes, and each edge along the red and blue routes (from C to D and from F to G) has a weight of 2.5 minutes in-vehicle time. With this graph structure, the shortest path according to travel time can be calculated including a breakdown (29 minutes walking, 15 minutes in-vehicle, and 7.5 minutes waiting), as well as the distance (2 km walking, 6 km in-vehicle).

The origin-destination matrices are calculated after the first four components of the network generator. They are also calculated using the updated travel time values from Traffic Assignment for iteration and analysis purposes.

2.3 Four-Step Model

In transportation demand planning, prediction of the transportation requirements of a city are commonly modelled using a four-step model, which can be most accurately thought of as a family of models that share the same large scale design structure.

Transportation demand modelling is an attempt to model a number of human decisions,

such as where to live, work, play, and how to access these locations, taking into account the transportation options available. This is a complicated set of decisions with many interacting factors. The distinctive aspect that all four-step models share is the division of these decisions into sub-models:

1. trip generation, which determines the number of trips leaving and entering each zone,
2. trip distribution, which matches trip origins and destinations according to travel proximity and destination attractiveness,
3. mode split, which chooses what mode is used to make each trip, mimicking human decisions, and
4. traffic assignment, which routes all the trips to find the traffic on each link.

These steps are expanded on in Sections 2.3.1-0.

To model travel demand, the city needs to be represented in data in such a way that the various components of the four-step model can use, manipulate, and add to the city data. Two key aspects of the city are modelled in this four-step model: the land and how it is used, and the transportation network. (Many four-step models also include households as a subdivision of modelling land use). The land is divided into small, geographically contiguous areas most commonly called traffic analysis zones (TAZs). The delineation of the TAZ is commonly done to create zones that are as homogeneous as possible, with respect to the households and land use types (commercial, residential, industrial, recreational, etc.). The TAZ are commonly stored and referred to as a table called a zone table, where each row corresponds to a zone and each column a measurable zone characteristic. Common characteristics include population, population that is employed, population that is students, floor area by purpose, and number of jobs located in the zone. The transportation network is represented as a graph (in the mathematical graph theory sense), where each link represents a road segment between two intersections. Each link has properties such as length, capacity, free flow speed, and response to congestion in the form of a volume-delay function (see Section 0). The links are commonly stored in a single table called a link table.

A four-step model is not the only way to model travel demand. Another grouping of models are called activity activity-based models, in which travel is seen as something that

enables various activities that people desire to engage in. Travel arises from the need to access specific locations at specific times. This is in contrast to trip-based models like the four-step model, which assigns a number of trips to each zone and then adds characteristics of the trip such as the destination, the mode, and the route in sequence and largely independently. For this project, the four-step model was chosen because it is well known, proven, and requires relatively little data. Additionally there are many resources to aid in implementation, such as textbooks and software options. In terms of data requirements, the four-step method of considering decisions in isolation reduces the need for complex disaggregate data (in time and space). A four-step model effectively only needs aggregate measures of population and land use. This is especially important for this project and its modelled city data, as aggregate data is easier to model accurately than highly disaggregate data.

2.3.1 Trip Generation

During the trip generation step, the number of trips that start from and end at each zone is determined. There are a variety of common ways to carry out this step, including zonal linear regression and household-based categorical models. The amount of data used can vary greatly both in resolution and the number of data fields used. In a full travel demand model carried out on a real-world city, it is advised to use a disaggregate household-based model, using attributes such as number of inhabitants, income, number of vehicles, access to transit passes and other transit discounts, number of student inhabitants, number of working inhabitants, etc. Categorical models are used to capture the complicated non-linear and multivariate relationships that often occur. This type of analysis requires a large amount of data and statistical work to ensure proper categorization. For a study carried out on a real city, this depth of analysis is possible and necessary to ensure that both the total amount of travel and the specific origins and destinations of travel are accurate. Different models are used for trips produced and trips attracted, as well as for the different types of trips such as home-based work, school, and discretionary, and non-home-based trips. For example, for trips attracted and non-home-based trips, zonal linear regression models are used as the household is not a valid unit of

analysis. For more on the different types of trips and models, see Ortúzar and Willumsen, 2011.

However, for this study the purpose of trip generation is somewhat different: since the city being used as input is not a real city, the priorities of trip generation have changed. Instead of seeking to have the number, spatial origin and destination of the trips be highly accurate to the city data, it is more important that the travel demand is in the right range for a city of its size, and that the travel is spread out through varied origins and destinations. To meet these requirements without introducing unnecessary data requirements, a zonal linear regression model was used to acquire the number of trips starting from and ending in each zone. These two models were used for all trip types. The linear relationships that were used are shown in equations (2.24) and (2.25):

$$O_i = dhP_i \tag{2.24}$$

$$D_i = dhE_i \tag{2.25}$$

where O_i is the number of trips with origin zone i , D_i is the number of trips with destination zone i , P_i is the number of inhabitants of zone i , and E_i is an abstract measure of the attraction strength of zone i (this would be related to number of jobs as well as other attractions such as retail, etc). d and h are coefficients for the number of trips per person per day and the proportion of daily trips that occur during peak hour, respectively.

In a typical use of a zonal linear model for trip generation, the calibration of the equations is carried out through linear regression on a set of collected travel data from the study region. Similar to the discussion above about the differences in nature between a typical trip generation model and the trip generation model required for this project, the calibration of this model isn't actually of the highest importance. Assuming a value of 1.8 for d (i.e. 1.8 trips per day) was sufficient for the model to meet its design goals. The value of 0.091 for h was obtained from the Highway Capacity Manual Exhibit 8-9, which gives 0.091 as the K-factor appropriate for urban settings (implying that 9.1% of daily trips occur during the peak hour).

2.3.2 Trip Distribution

Trip Distribution is the stage in which the trip origins and destinations are connected. The output of this step is an origin-destination matrix T with entries $T_{i,j}$ where $T_{i,j}$ is the number of trips from zone i to zone j . Therefore

$$D_j = \sum_i T_{i,j} \quad (2.26)$$

is the sum of column j of T , and is equal to the value from the trip generation step.

Similarly,

$$O_i = \sum_j T_{i,j} \quad (2.27)$$

is the sum of row i of T , and is equal to the value from the trip generation step. These matrix properties mean that the trip distribution step has to find a matrix with these predefined row and column sums. However, there are many matrices which fit this requirement, and so the model also takes into account zonal and network properties. This ensures that the matrix solution found reflects transportation concepts. Zonal and network properties are taken into account through the use of a gravity model in which the amount of travel between each pair of locations is influenced by a level of travel attraction between that pair of locations. Equation (2.28) shows the form of this model.

$$\lambda_{i,j} = O_i D_j c_{i,j}^a e^{bc_{i,j}} \quad (2.28)$$

where $\lambda_{i,j}$ is the travel attraction from zone i to j , O_i is the number of trips with origin zone i , D_j is the number of trips with origin zone j , $c_{i,j}$ is a generalized travel cost from zone i to j , and a and b are parameters. This model is called a gravity model because it is analogous to Newton's famous law of universal gravitation. In this equation there are effectively two key influences on the amount of travel between an origin and destination: the amount of travel at each end of the trip, and the travel impedance between the two locations. The amount of travel is captured by the product $O_i D_j$, and the impedance is captured by $c_{i,j}^a e^{bc_{i,j}}$. This function is called an impedance function, and is a decaying function that introduces a deliberate bias towards shorter trips. The value $c_{i,j}$ is a measure of the ease of travel from zone i to j . For this project, the automobile in-vehicle travel time is used as a proxy for this measure. AVRS travel uses the same travel time, and while transit systems can have a different travel time for a given origin and destination, on the whole the travel times are similar. Additionally, automobiles are the dominant mode in all of Canada.

In order to find the actual matrix that has the correct row and column sums and that uses the gravity model shown above, a doubly-constrained matrix balancing algorithm was used. A doubly-constrained model ensures that both the row and column sums are equal to

O_i and D_j , respectively, as discussed above. This algorithm makes use of equations (2.29)-(2.33).

$$B_{j,k} = \frac{1}{\sum_i A_{i,k-1} O_i c_{i,j}^a e^{bc_{i,j}}} \quad (2.29)$$

$$A_{i,k} = \frac{1}{\sum_j B_{j,k-1} D_j c_{i,j}^a e^{bc_{i,j}}} \quad (2.30)$$

$$T_{i,j,k} = A_{i,k} O_i B_{j,k} D_j c_{i,j}^a e^{bc_{i,j}} \quad (2.31)$$

$$\varepsilon_O = \max_i \left| O_i - \sum_j T_{i,j,k} \right| \quad (2.32)$$

$$\varepsilon_D = \max_j \left| D_j - \sum_i T_{i,j,k} \right| \quad (2.33)$$

In these equations, k is the iteration, and $A_{i,k}$ and $B_{j,k}$ are balancing factors used for iteration k . Note that $T_{i,j,k}$ is not equal to $T_{i,j}$ for all values of k . Pseudocode is shown below.

Set $A_{i,0} = 0$ for all i , and $B_{j,0} = 1$ for all j .

For iteration k (starting at $k = 1$):

Calculate $B_{j,k}$ for all j using equation (2.29).

Calculate $A_{i,k}$ for all i using equation (2.30).

Calculate $T_{i,j,k}$ for all i and j using equation (2.31).

Calculate ε_O and ε_D using equations (2.32) and (2.33).

If $\varepsilon_O < \text{tolerance}$ and $\varepsilon_D < \text{tolerance}$, then exit the for loop.

A value of 0.1 was used for the tolerance.

There are two parameters that need calibration in the gravity model: a and b from equation (2.28). Typically, these parameters are calibrated through the use of an origin-destination matrix obtained from a travel survey. The parameters a and b are chosen so that the root mean square difference between the origin-destination matrices obtained from the gravity model (using the row and column sums from the survey matrix as input) and from the survey is minimized. The survey data is described in section 2.3.3.1.

2.3.3 Mode Split

The third step of the Four-Step model is to assign modes to each of the trips being modeled. After the second step, there are a number of trips that have been generated, and starting and ending locations are chosen for each trip. Using just this information, and

information about the network in which the travel will occur, the mode split model assigns a mode to each trip by weighing the different modes according to their properties with respect to the particular origin and destination of the trip.

The most common mode split models are the family of *discrete choice* models, in which a utility (or a disutility) is calculated for each of the options. The utility is a measure of how valuable a given choice is to the person who may make that choice, compared to the other available choices. The utilities are calculated using functions that typically include properties of the trip and of the individual making the trip. There are a number of different types of models in this family such as multinomial logit models, nested multinomial logit models, multinomial probit models, mixed logit models, and others. The differences between these different models are generally found in how the utility functions are calculated and in how the utilities are used to make the choice. For more on discrete choice modeling in transportation modeling, see Ortúzar and Willumsen (2011).

For this research, the mode split is carried out using a multinomial logit model. There are a total of 3 possible transportation modes: personal automobile (including being a passenger in a personal automobile), public transit (including both bus and rail transit), and autonomous vehicle ride-sharing services (AVRS). It should be noted that walking and other active transportation are not explicitly modeled as their own mode, primarily because this research is more concerned with the interplay of transit and AVRS modes. However, the public transit model includes walking for access and egress, and some trips modeled as public transit do not actually involve a public transit vehicle. In other words, the model does include walking-only trips but classifies them as transit trips that never find it necessary to board a transit vehicle. The form of the utility functions for the model is shown in equation (2.34),

$$U_m = K_m + \alpha_{T,m}T_m + \alpha_{C,m}C_m + \alpha_{W,m}W_m + \alpha_{I,m}I_m \quad (2.34)$$

where U_m is the utility of mode m , K_m is the mode-specific constant for mode m , T_m

is the in-vehicle travel time in minutes for mode m , C_m is the out-of-pocket costs in 2019 CAD for mode m , W_m is the walk time in minute for mode m , I_m is the wait time in minutes for mode m , and $\alpha_{X,m}$ is the weight coefficient for variable X and mode m . The coefficient values were obtained through a combination of calibration methods and using values advised through literature review.

In a multinomial logit model, the probability of making a given choice is given by

$$P_{i,m} = \frac{e^{U_{i,m}}}{\sum_{k \in M} e^{U_{i,k}}} \quad (2.35)$$

where $P_{i,m}$ is the probability of individual i choosing mode m and M is the set of all modes. These probabilities then become the proportion of all individuals traveling between that origin and destination that use mode m .

The multinomial logit model was chosen for its simplicity and relatively low requirements for data. More complicated and potentially more accurate models such as mixed logit models would require more complicated information about the characteristics of the 3 modes and the relationships between them. This type of information is difficult to obtain even without considering that one of the modes being modeled (AVRS) does not yet exist and its characteristics are not well understood. In contrast, the multinomial logit model only requires coefficients which weigh the relative value of money and time spent in various activities. These values have been studied in a number of contexts and are easily adapted between different modeling situations. As well, the multinomial logit model can handle complexities such as characteristics of the individual trip maker and mode-specific coefficients.

The utility functions are structured as multivariate linear functions. This functional form is both commonly used for utility functions and is mathematically simple. The variables are all cost and time variables, two categories which are traditionally the most important factors for transportation choices. Time is broken down into different activities – walking, waiting, and in-vehicle – not only because there is evidence that people perceive time differently during these activities but also because doing so highlights the differences between the 3 modes involved in this study. Compared to personal automobiles, the necessity of waiting for and of walking to and from the transit access locations (the last mile problem) has played a major role in the frequent disutility of transit relative to automobile. A mode such as the AVRS in this project could solve the last mile problem and reduce wait times, and therefore this study seeks to understand what impact walking and waiting have on the competition of public transit with AVRS. This justifies their inclusion in the mode split model.

2.3.3.1 Calibration

Multinomial logit models require the calibration of the utility function coefficients from equation (2.34): $\alpha_{X,m}$ and K_m , for all values of X and m . Though the coefficients $\alpha_{X,m}$ could be specified for each mode, for this research they have been chosen such that $\alpha_{X,m} = \alpha_X$ for all modes m and variables X . This requires the assumption that people perceive the various times and costs the same regardless of the mode. This assumption is made partly because the AVRS mode is not well understood, and therefore it is unknown how exactly people will respond to it. However, it shares properties with the other modes, which are well understood. For example, in-vehicle travel time using AVRS can be expected to be very similar to being a passenger in a personal automobile, and is comparable to being the driver as well (the key difference being that time spent as a passenger can be used for another purpose and isn't necessarily lost, while time spent as a driver cannot really be used for other purposes). Additionally, money spent on AVRS can be expected to be perceived very similar to money spent on transit, as AVRS can be reasonably expected to use the same kind of payment methods and times as public transit.

Applying the function form from equation (2.34) to the three modes results in the following set of utility functions:

$$U_{auto} = K_{auto} + \alpha_T T_{auto} + \alpha_C C_{auto} \quad (2.36)$$

$$U_{transit} = \alpha_T T_{transit} + \alpha_C C_{transit} + \alpha_W W_{transit} + \alpha_I I_{transit} \quad (2.37)$$

$$U_{avrs} = \alpha_T T_{avrs} + \alpha_C C_{avrs} + \alpha_I I_{avrs} \quad (2.38)$$

Walk time for the automobile and AVRS modes has a value of 0 for all trips, as does waiting time for the automobile mode. Those terms have been omitted for clarity. In a standard multinomial logit model with three modes, two mode-specific constants (i.e. K_m) are required. The constant represents all other properties of the mode that are not otherwise included, part of which is the trip-makers' attitudes towards the mode. Values for the coefficients in the utility functions, including the mode specific constants, are normally calibrated using observed mode choice data. In this study, AVRS is a projected future mode and therefore it is not possible to calibrate a mode specific constant for AVRS using observed data. Furthermore, the goal of this project was to compare AVRS and transit based on the key mode attributes (i.e. cost and travel time) already included in the

equations. For these reasons, the constant for AVRS was set equal to 0 (the same as the mode specific constant for transit).

Calibration of multinomial logit models is typically done using the Maximum Likelihood method. This methodology is required since the output of the model is the probability of making a certain choice, and any data will by necessity only include the choice that was actually made. Maximum Likelihood evaluates a set of parameters by the likelihood of those parameters to predict the calibration data – the probability of the model to predict the observed data. Maximum Likelihood calibration was carried out on a data set obtained from the Transportation Tomorrow Survey and the Regional Municipality of Waterloo, and done using the *Biogeme* software package (Bierlaire 2018).

The Transportation Tomorrow Survey (TTS) is a travel survey carried out by the University of Toronto in conjunction with the governmental agencies of the regions included in the study. It is carried out every 5 years and includes information on trip behaviour of participants. Due to the partnership with the regional governments, the data collected is not quite identical for each region. This project used the 2011 data from the Regional Municipality of Waterloo, a census metropolitan area with a 2011 population of 496,383 people, which was divided into 578 traffic analysis zones (TAZ) for the purposes of the survey.

The TTS data was reformatted as either origin-destination (OD) data in a matrix, or trip data. The OD data consisted of the automobile and the transit travel time between each origin and destination TAZ, although some OD pairs were missing transit data due to transit being unavailable. Each trip in the dataset had characteristics such as the origin and destination zones TAZ, the distance between those TAZ, the mode used to make the trip, the number of transit routes used (if transit was used at all), the departure time of the trip, and the trip purpose. Trip-maker characteristics were also available, including the number of vehicles in the household, possession of a licence by the trip maker, possession of a licence by anyone in the household, age, and employment/student status. Additional information was available; the data mentioned here is the most relevant to this project. The OD data was combined with the trip data to give expected transit and automobile travel time of each trip.

The utility functions in equations (2.36) to (2.38) require the in-vehicle travel time, the wait time, the walking time, and the cost for each trip. Note that despite calibration data being available for trip-maker characteristics, these were not included in the utility functions as the network generation model describe in section 2.2 does not provide individual characteristics. For the data that was not directly available, the wait time was calculated by multiplying the number of transit routes by an average wait time of 7 minutes. The transit cost was a flat fare of \$3.25 and the auto cost was the distance multiplied by 0.64 dollars per km. These parameters are the same as those used in Table 3.1. The walking time was not available and was omitted from the functions for the calibration. Due to the literature-based approach described in section 2.3.3.2, this was not a major challenge. The TTS data included more modes than this project: private automobile driver, private automobile passenger, motorcycle, public transit, walking, bicycling, school bus, taxi, and rideshare. The auto mode for this project consisted of the drivers and passengers of private automobiles and motorcycles.

Before using this data set, three filters were applied:

1. Some of the OD pairs had no transit travel time information. Trips between these OD pairs were removed.
2. Some of the OD pairs never had a transit trip made between them. This means that those trips also had no information about the transit routes used, and therefore couldn't have wait times calculated. Trips between these OD pairs were removed.
3. Some of the households had no vehicles, or had vehicles but no one with a driver's licence. These households were considered transit captives since they had no real choice; in a transit versus auto situation they had to take transit. The mode split model is meant to capture the choices, meaning including these captives clouds the data.

The application of these filters reduced the dataset from 49,549 trips down to 4,599. This is a dramatic reduction, but it was found that without this degree of filtering, the mode split model resulted in over 99% of trips being auto. Table 2.14 is a summary of the set of trips used for mode split calibration.

Table 2.15: Summary of TTS Trip Data

| Measure | unit | Auto | Transit | All Trips |
|---------------------|---------|-------|---------|-----------|
| Number of Trips | | 3,104 | 1,495 | 4,599 |
| Percent of Trips | | 67.5 | 32.5 | 100 |
| Average Distance | km | 4.326 | 6.571 | 5.056 |
| Average Travel Time | minutes | 6.45 | 20.60 | 11.05 |
| Average Wait Time | minutes | 0 | 9.20 | 2.99 |

This calibration is not guaranteed to produce results that are suitable for use in the full modeling context. There are a number of theoretical checks that can be used to evaluate the model. Ortúzar and Willumsen (2011) suggest ensuring that the sign matches with expectations and that the variables are statistically significant according to the t-test, and assessing these two with respect to the importance of the variable in question. They divide variables into two groups: policy variables and other variables. Policy variables are those that are included because there is strong theoretical evidence that they are relevant to the decision, or they are part of the goal of the modeling. For this project, the in-vehicle travel time and the out-of-pocket cost are both certainly policy variables, and the walk and wait times, while not crucial, are very valuable due to way that they distinguish between the different modes. This leads to the categorization of all the variables in equations (2.36)-(2.38) as policy variables. Therefore, they should be included in some way regardless of the results of the calibration.

However, the calibration results were not very meaningful. It is expected that both increasing the cost and increasing the time should make a trip less desirable, but in every configuration of the utility functions that was tried, there were sign issues. Additionally, many configurations had variables that were insignificant, but no consistent lack of significance was observed that would suggest not including one of the variables. In cases such as this, Ortúzar and Willumsen (2011) suggest reconsidering the model structure or turning to accepted values from other studies.

2.3.3.2 Values of Time in Literature

One of the challenges of finding calibration values from literature is that each study has its own goals, requirements, context, and scope. To mitigate these issues, a value of time was used instead of seeking to directly find the coefficients shown in equations (2.36)-(2.38). Values of time have primarily been studied in the context of evaluating the

economic value of an improvement to transportation that results in less time spent in travel. The value of time is effectively the answer to the question of how many dollars a traveller would be willing to spend in order to reduce the travel time of a trip by 1 hour, and this is where a connection to utility functions can be made. By setting the change in utility resulting from an increase of 60 minutes of in-vehicle travel time equal to the change in utility resulting from an increase of \$1 of out-of-pocket costs, a relationship can be found between the value of time and the utility function coefficients:

$$VOT = \frac{60\alpha_{T,m}}{\alpha_{C,m}} \quad (2.39)$$

Reports describing methods for computing value of time and citing values of time have been published by various countries including the USA (US DOT, 2016), United Kingdom (UK DFT, 2009), Denmark (Fosgerau et al., 2007), France (Boiteux and Baumstark, 2001), Norway (Ramjerdi et al., 1997), Sweden (Algers et al., 1995), Switzerland (Axhausen et al., 2006) and others. Of these the most valuable is the US report, since the United States has the most similarities to Canada with respect to transportation, infrastructure, and culture.

The US DOT report gives value of travel time recommendations for various types of travel. The major classification is between business and personal travel. Business travel includes all travel done by employees while they are being paid by their employers, meaning it excludes commuting. Special consideration is taken for vehicle operators such as truck drivers, bus drivers, transit rail operators, locomotive engineers, and airline pilots and engineers. Travel is also categorized by distance (local and intercity) and mode (surface, except high-speed rail; and air and high-speed rail). For this research, the relevant context is local surface travel. There is also a recommended aggregation method for all purpose travel: a weighted average of the personal and general business values of travel time. The weights are 0.954 for personal and 0.046 for business.

The value of personal travel time is 50% of the hourly median household income. The 2016 Canadian census gives a value of \$70,336 (2015 CAD), which is \$75,695.46 (2019 CAD). This is \$36.39/hr, using the assumption (also made in US DOT (2016)) that there are 2080 working hours in a year. The value of business travel time is equal to the hourly median individual total compensation. Due to differences in Canadian and American labour

statistics, the American value of \$25.40/hr (2015 USD) was used. This becomes \$35.60/hr (2019 CAD). This method results in a value of travel time of \$19.00/hr (2019 CAD). Full calculations are shown in equation (2.40).

$$VOT = 0.954 \left(\frac{0.5 \cdot \$75,695.46/year}{2080 \text{ hours/year}} \right) + 0.046(\$35.60/hour) \quad (2.40)$$

$$= \$19.00/hour$$

For the walking time and waiting time coefficients, it is sufficient to know the ratios between their values of time and the in-vehicle value of time. A number of values for these ratio have been suggested: Gwilliam (1997) suggests 1.5 for walking and waiting, Boiteux (2001) suggests 2 for walking and waiting, and both Zhang et al. (2005) and UK Department for Transport (2009) suggest 2 for walking and 2.5 for waiting. Taking these suggestions into account, a value of 2 is used for both walking and waiting. Incorporating the value of time and equation (2.39), the α parameters from equations (2.36)-(2.38) for the in-vehicle, walking, and waiting time can be found in terms of α_C . This results in equations (2.41)-(2.43) below.

$$U_{auto} = K_{auto} + \left(\frac{\$19.00}{60 \text{ min}} \right) \alpha_C T_{auto} + \alpha_C C_{auto} \quad (2.41)$$

$$U_{transit} = \left(\frac{\$19.00}{60 \text{ min}} \right) \alpha_C T_{transit} + \alpha_C C_{transit} \quad (2.42)$$

$$+ \left(\frac{2 \cdot \$19.00}{60 \text{ min}} \right) \alpha_C W_{transit} + \left(\frac{2 \cdot \$19.00}{60 \text{ min}} \right) \alpha_C I_{transit}$$

$$U_{avrs} = \left(\frac{\$19.00}{60 \text{ min}} \right) \alpha_C T_{avrs} + \alpha_C C_{avrs} + \left(\frac{2 \cdot \$19.00}{60 \text{ min}} \right) \alpha_C I_{avrs} \quad (2.43)$$

These equations have only two values that need to be calibrated: α_C and K_{auto} . To find the value of K_{auto} , the calibration process discussed in section 2.3.3.1 was used in conjunction with testing the performance of the mode split model when applied to idealized networks and the calibration data. The best values of α_C and K_{auto} were found to be 1 and -6.62398, respectively. A negative value of K_{auto} indicates a bias against using auto and likely occurs because socio-economic characteristics (such as household income, car ownership, possession of a driver's licence) of the trip makers are not included as part of the model. The final utility functions are shown in equations (2.44)-(2.46). Unlike the calibration attempts discussed in section 2.3.3.1, the new model has no insignificant coefficients or sign issues.

$$U_{auto} = -6.62398 - 0.316673 T_{auto} - C_{auto} \quad (2.44)$$

$$U_{transit} = -0.316673 T_{transit} - C_{transit} \quad (2.45)$$

$$U_{avrs} = -0.316673 T_{avrs} - C_{avrs} - 0.633347 I_{transit} \quad (2.46)$$

2.3.4 Traffic Assignment

With the number of trips calculated for each mode and for every origin and destination pair, the last remaining step of the four-step model serves this travel demand by finding a route for each trip. The most important principle for how to assign these routes is that of user equilibrium: no trip can be made in a shorter time by selecting an alternative route. Another way of saying this is that all routes used for a given origin and destination have the same travel time, and all unused routes for that origin and destination have a longer travel time. In both of these definitions a generalized cost could be used in place of the travel time, but travel time is typically the leading factor in user route choice, and so will be used for this project.

At the completion of this step, a wide range of metrics can be calculated for analysis. The base metric is the volume on each link. From this, metrics such as the speed on links, the travel time on links, commonly used routes, travel time between any origin and destination, and more can be calculated. Of particular interest for analysis in this project are the travel times of links and between origins and destinations, as they can be used to recalculate the mode split and therefore the effects of the changes in transit usage.

Congestion is a key phenomenon modelled during traffic assignment. Congestion is modelled through the use of volume demand functions, which relate the number of vehicles using a link to the travel time required to traverse that link. The specific volume demand function used was the BPR function, shown in equation (2.15). In section 2.2.4, the inputs into this equation such as the capacity and the α and β parameters were assigned to the road grid. These properties are then used as inputs into equation (2.15).

Traffic assignment was carried out using Visum software. More specifically, the LUCE algorithm for static user equilibrium was used. For more on the LUCE algorithm, see Gentile, 2014.

2.3.5 Iteration

The four-step model makes a major simplification for the sake of effective modelling: the human decisions in the model are divided into four and carried out in sequence. This is not really accurate to how travel decisions are actually made. In reality people generally take the many competing and complicating factors into account simultaneously, then make a decision once. While dividing this single human decision-making process into a sequence of four separate decisions is beneficial for developing models with clear relationships, it also introduces a weakness: the earlier decisions (whether to travel, where to go) can't be made using the information gained once the later decisions (how to get to the destination) are made. A common strategy to address this flaw is iterating between the four steps, so that the outputs of later steps can be used as inputs for earlier steps.

Depending on the goals of the study and the exact formulation of each of the four sub-models, the iteration can take on a number of different forms. For example, if the effects of induced demand are a topic of study, the trip generation model could be adjusted to use measures such as travel speed or amount of delay time as an input. Iterations carried out on all four steps would then allow for the model to take into account effects such as the addition of lanes on a highway inducing more demand as more people find using that facility to be tolerable.

The introduction of an AVRS mode will likely have an impact on all the decisions that make up the four-step model. For example, the introduction of an easier form of transportation has historically increased travel. This can be seen across nearly all types of transportation from the introduction of new technologies such as trains, automobiles, and airplanes to the completion of specific transit projects in cities. AVRS could also have an effect on trip distribution decisions as those decisions are heavily dependent on the amount of congestion in the city. Finally, as AVRS represents the introduction of a new mode, it will definitely have an impact on the mode split decisions. All of these effects could be modelled through iteration. However, as AVRS is an unknown and not in existence, it was decided that trying to model induced changes in city-wide demand was too large a problem to include in this research. For this reason, iteration was done over the last three

steps of the four-step model (i.e. the total number of trips remains constant between the without AVRS and the with AVRS scenarios).

The results of the traffic assignment step (see Section 0) is that the link table has updated travel times. These travel times can be used to calculate new origin-destination matrices for costs and for the various travel times. These are the same matrices that are used as inputs into the trip distribution and mode split models, making for natural iteration. A diagram of the four-step model, with iteration, is shown in Figure 2.22.

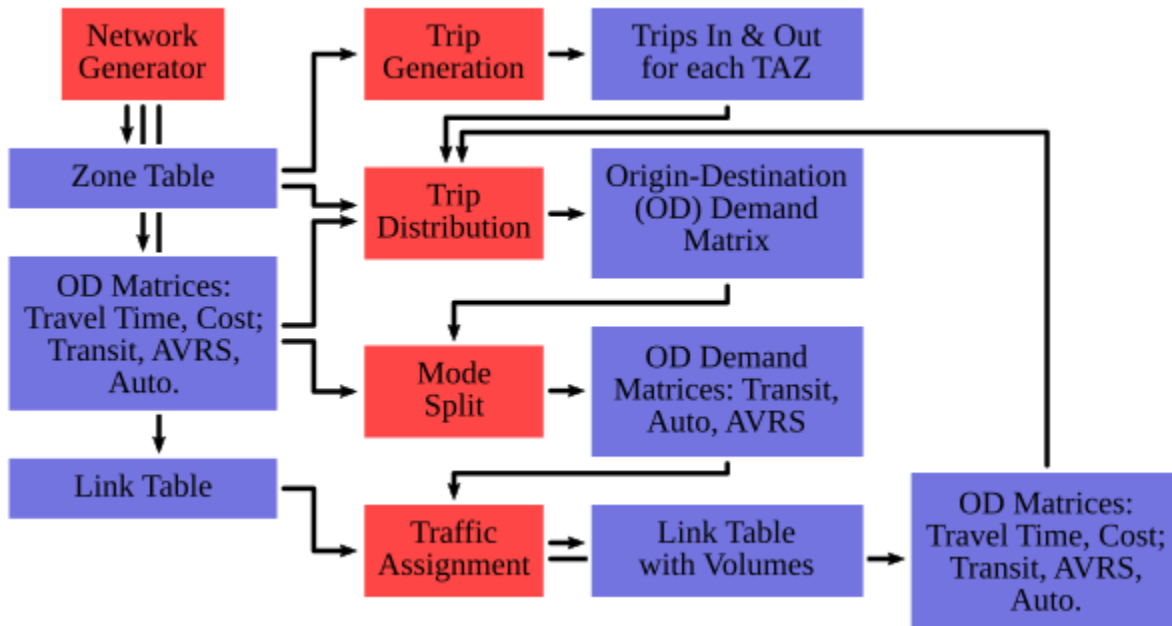


Figure 2.22: Four Step Model components with iteration.

The convergence measurement was the maximum entry-wise absolute difference between the current and previous origin-destination demand matrix. The origin-destination demand matrix is the result of the trip distribution, as discussed in Section 2.3.2. Convergence was considered to have occurred if this difference was less than 5 trips, as shown in equation (2.47), where C is the convergence measurement, T is the origin-destination demand matrix, i and j are matrix indices, and k is a counter of the iterations.

$$C_k = \max_{i,j} |T_{i,j}^k - T_{i,j}^{k-1}| \leq 5 \quad (2.47)$$

Not every run converged, and there was a tendency for larger cities with more traffic to converge more slowly or not at all. However, the general pattern was that either convergence occurred within the first 12 iterations or not at all. For this reason, the model

was allowed to time-out after 12 iterations. Runs that did not converge were not used in the results.

3 Results

3.1 Experimental Set-Up

The impact of AVRS on traditional transit and on cities in general was studied by comparing control simulations against simulations which included AVRS as a transportation mode. The route and network generators were used to create cities and transit systems. The four-step model was applied to these cities under two different modal conditions: once with the traditionally dominant modes of private automobile and public transit, and once with AVRS alongside the traditional modes. The parameters for the cities were chosen to create cities comparable to the cities and transit systems that currently exist in Canada. Parameter choices are given in Table 3.1. The work done to produce realistic transit route characteristics (Section 2.1), transit alignments (Section 2.2.2), population distributions (Section 2.2.3), and road networks (Sections 2.2.1 and 2.2.4) informed the selection of many of these parameters. The AVRS cost was obtained from the work done by Bösch et al. (2018), who calculated the price (for the user) of a privately-owned vehicle to be 0.47 CHF/km and the price for an AVRS service to be 0.43 CHF/km. This price calculation for AVRS includes business overhead, fuel, maintenance, depreciation, cleaning, parking and tolls, tax, insurance, interest, and a profit margin of 3%. These prices were applied to a Canadian context by assuming a constant ratio of automobile to AVRS price.

Table 3.1: Parameters used to create hypothetical networks

| Parameter | Value |
|-------------------------|--------------|
| Block Size | 1000 |
| Fill Density | 30 |
| Bridge Gap | 1.75 |
| Bridge Deviation | 1.25 |
| Remove Percent | 7 |
| Trips per Person | 1.8 |
| Peak Hour Trips per Day | 0.091 |
| Impedance a | 0 |
| Impedance b | -0.2 |
| Automobile Cost | 0.6458 |
| AVRS Cost | 0.5908 |
| AVRS Wait Time | 7 |
| Transit Fare | 3.25 |
| Walking Speed | 1.15 |

In the running of the model, 6 cities were created. High-level characteristics of these cities are shown in Table 3.2. Note that the populations range from 0.5 million to 2 million. The original scope included cities up to about 8 million population, which would cover nearly all urban areas in Canada. However, simulation of larger cities required too much computation time to be practically analysed, and larger cities rarely converged. Consequently, the original scope was revised to model cities up to 2 million population. The number of transit routes and the city sizes were selected to be similar to the sizes of the transit systems observed during the examination of Canadian transit data (see Section 2.1.1.3). The “City Size” column refers to which of the transit route datasets (Small vs Large cities, for bus routes) were used to generate transit routes. Note that the model is pseudo stochastic, and both Population and Population Density are outcomes, not inputs. Consequently, the values listed in the table are the values obtained for those particular runs.

Table 3.2: Modelled City Characteristics

| City | City Size | Bus Routes | Rail Routes | Population | Population Density (people/km ²) |
|------|-----------|------------|-------------|------------|---|
| A | Small | 50 | 0 | 501,380 | 1,388.9 |
| B | Small | 50 | 0 | 500,135 | 1,134.1 |
| C | Small | 100 | 2 | 1,000,580 | 1,737.1 |
| D | Small | 100 | 2 | 1,009,945 | 2,086.7 |
| E | Large | 150 | 4 | 2,000,420 | 2,744.1 |
| F | Large | 150 | 4 | 2,000,335 | 1,836.9 |

Additional tables of data characterizing the modelled cities are included in the appendix (Table 4.4 and Table 4.5).

3.2 Impact of Autonomous Vehicle Ridesharing

One of the main goals of this research was to examine the impacts of the introduction of AVRS on different characterizations of transit modes. To accomplish this, the relationship between transit route characteristics and the change in passengers was analyzed. Route characteristics were previously discussed in Section 2.1.1. The key route characteristics used were headway, length, geodesic distance, operating speed, route indirectness, and mode. The comparison of cases with and without AVRS allows for initial passengers of the route (i.e. the usage without the AVRS mode present) to be another route characteristic.

The demand (which is effectively the usage) for each route was measured in passenger-kilometers (pkm). Correspondingly, the change in demand that results from comparing the two cases is measured in change in passenger-kilometers (Δp) and the relative change in passenger-kilometers (Δr). The calculation for relative change in passenger-kilometers is shown in equation (3.1) where r_0 and r_1 are the passenger-kilometers for the case without AVRS and with AVRS, respectively, and Δr is the relative change in passenger-kilometers.

$$\Delta r = \frac{r_1 - r_0}{r_0} \quad (3.1)$$

To characterize transit routes according to how many passengers they lose, the 6 quantitative route characteristics listed above were compared against the change and relative change in passenger-kilometers.

The set of data available for analysis consisted of the set of routes from each of the 6 city sizes that were modelled. As per Table 3.2, this provided results for 600 bus routes and 12 rail routes. However, a review of the results indicated a few issues and consequently some of these results were removed from the data set before analysis as follows:

1. There are 75 routes (of 612) that had 0 initial passengers. The process of network generation, including transit route locations, service frequency, and population densities is stochastic and therefore sometimes transit routes are created that are not competitive relative to auto and/or traverse zone for which there are no or few trips. Whenever Δr is used in analysis, these routes are excluded to avoid division by zero.
2. Due to the stochastic nature of the modelling process there is the potential for extreme values (outliers) to occur within the output. Filters were applied to each of the Δr and Δp analyses output to identify and remove these extreme values. It was observed that most of the outliers occurred during simulation runs which experienced issues with convergence.
 - a. Seven routes had increases in passenger-kilometers after the introduction of AVRS or greater than or equal to 200,000. The results from these routes were excluded from Δp analysis.
 - b. Two routes that satisfy $\Delta r > 3 \vee (\Delta r > 0 \wedge r_0 < 10)$ were excluded from Δr analysis.

A correlation analysis was conducted (coefficients of correlation are shown in Table 3.3) with the remaining data. Most of the route characteristics exhibit weak evidence of a relationship at best, with the exceptions being the initial passengers (correlates with Δp) and the headway (correlates with Δr).

Table 3.3: Correlation coefficients for quantitative route characteristics

| | Correlation Coefficient | |
|--------------------|---|--|
| | Change in passenger-kilometers (Δp) | Relative change in passenger-kilometers (Δr) |
| Headway | 0.20 | -0.55 |
| Length | -0.24 | 0.24 |
| Geodesic Length | -0.15 | 0.29 |
| Indirectness | -0.11 | -0.08 |
| Operating Speed | -0.28 | 0.28 |
| Initial Passengers | -0.97 | 0.29 |

The initial passenger-kilometers correlates strongly with change in passenger-kilometers. Figure 3.1 shows this relationship. The linear coefficient of -0.4875 indicates that it can be reasonably expected that most transit routes could lose up to about half of their passenger-kilometers after the introduction of AVRS. The intercept was not statistically different from 0 and was therefore fixed to 0. The statistics associated with the regression shown in Figure 3.1 suggest that the relationship is meaningful: the p-value for the linear coefficient was 0, and the relationship explains 93.5% of the variance. This was the strongest relationship found using the route characteristics in Table 3.3.

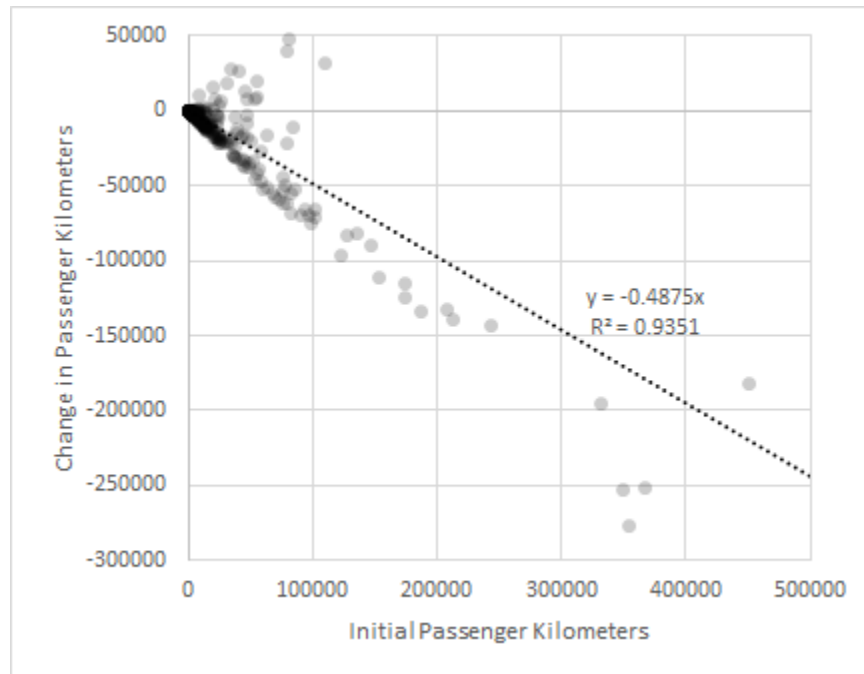


Figure 3.1: Relationship between initial passenger kilometers and change in passenger kilometers. Note that the axes limits have been chosen to show the detail of the majority of routes at the cost of excluding outliers. Full figure can be found in the appendix, Figure 4.2.

The relationship between headway and relative change in passengers is best modelled as the nonlinear relationship

$$\Delta r = 15.2h^{-1.55} - 1 \quad (3.2)$$

where Δr is the relative change in passenger-kilometers and h is the headway in minutes. This relationship was found by applying linear regression to the independent variable $\ln(h)$ and the dependent variable $\ln(\Delta r + 1)$, and is shown in Figure 3.2. The statistical significance of the regression is confirmed through the p-values of 6.1×10^{-14} and 2.7×10^{-33} for the value of the coefficient 15.2 and power -1.55 respectively, and the F-test p-value of 2.7×10^{-33} . The R^2 value is 0.238. The modelled relationship suggests that headway can have a large effect on the competitiveness of transit against AVRS. Additionally, the benefits of decreased headway are felt most strongly at headways below about 15 minutes, whereas once headways become larger than 20 minutes, additional increases had a reduced impact.

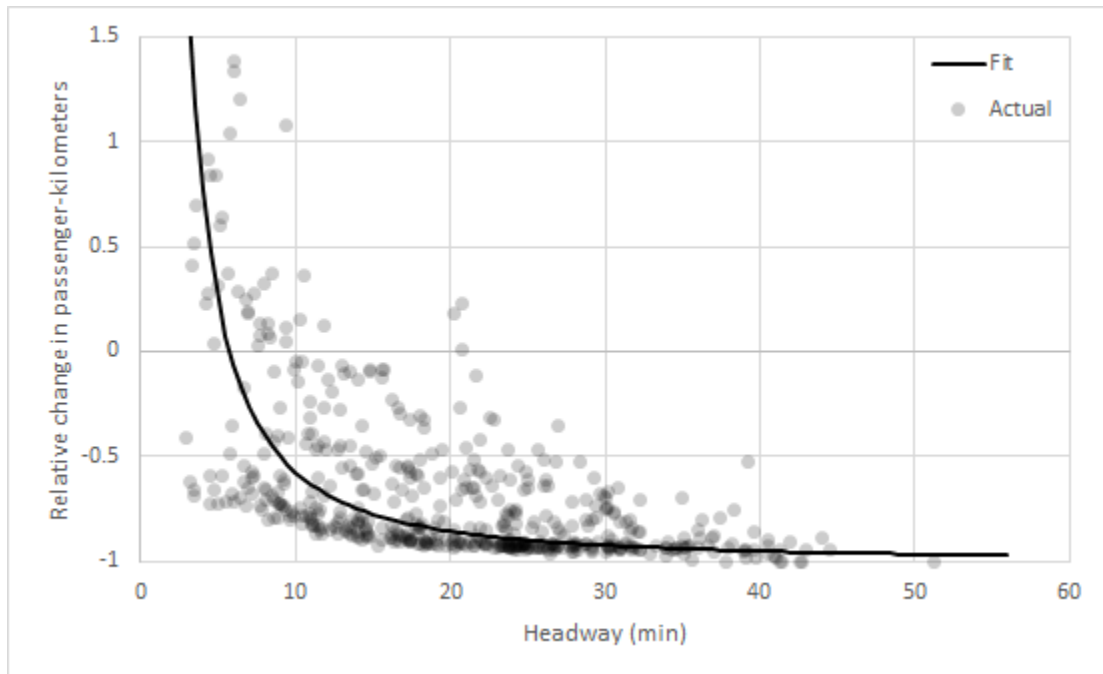


Figure 3.2: Relationship between route headway and the relative change in passenger-kilometers.

That headway and initial passengers have the strongest impact on future passengers is not a large surprise, especially in the context of the utility functions used to determine mode split. Comparing the transit and AVRS utility functions, both take in-vehicle travel time, waiting time, and cost into account, but transit also considers walking time. The effect

of walk time gives AVRS an inherent advantage over transit because AVRS has zero walk time, and this is reflected in the large number of passenger-kilometers lost by the transit routes. Depending on the length of the trip, the cost can give an advantage to either mode. Finally, the in-vehicle travel time for bus transit is largely the same as the in-vehicle travel time for AVRS, and when they differ the bus time is longer. This leaves only two opportunities for transit to consistently compete: lowering wait times and including more rail routes (i.e. routes that are not impacted by road congestion). For wait times, the assumed constant 7-minute wait time for all AVRS trips means that lowering the transit headways below 7 minutes provides transit with a competitive advantage. Rail routes are assumed to operate on separate rights-of-way than on-street modes and so there is a possibility to have increased operating speeds and therefore a lower in-vehicle travel time than AVRS. However, the inherent disadvantages of transit in the form of walk times and high in-vehicle travel times causes most routes to lose roughly half their passenger-kilometers.

What is somewhat more surprising is that the operating speed doesn't seem to have a large effect on transits' competitiveness with AVRS. This is likely due to the operating speed of bus transit being dependent on the amount of congestion on the roadways used. Therefore, any bus transit route that has a faster operating speed is going to be competing against an AVRS trip with a corresponding higher speed. The only exception to this are the rail routes, of which there are only 12 in the data set, likely too few to show a clear relationship.

The possibility of a difference in the effect of AVRS on rail and bus transit was investigated using Welch's t-test. Since the rail routes resulting from this model have higher initial passenger-kilometers than the bus routes, the t-test compared the relative change in passenger-kilometers between rail and bus routes. The t-test found that the mean relative change in passenger-kilometers was significantly different ($t=-3.65$, $p=0.0038$) for rail and bus, with the rail routes losing fewer passengers.

The introduction of AVRS had an effect on the simulated cities and their transportation beyond just affecting the transit routes. Many trips that were formerly made using transit were instead made using AVRS, causing an increase in the number of automobiles on the road. This had further effects such as longer travel times, shorter trips,

and increased congestion. Other research suggests that there is some expectation that AVRS will increase the capacities of existing infrastructure and reduce collisions, as a result of autonomous vehicles being able to coordinate their movements. This study does not include these sorts of effects and therefore it cannot speak to whether they will be stronger than the effects caused by the decrease in transit usage. However, this study does demonstrate that there is reason to believe that AVRS could increase congestion.

4 Limitations and Recommendations

As has been discussed throughout previous sections, there are many limitations of the modelling structure and their results. Since they have already been covered to some degree, this section will address large limitations and those that affect multiple components of the research. One limitation already discussed were that challenges with convergence limiting the study's scope to medium-sized cities (sections 2.3.5 and 3.1).

The model assumes that AVRS has been in existence for a sufficiently long time for the general population to accept it as a legitimate transportation option and to start using the availability of AVRS as a factor in making long-term decisions such as whether or not to own a vehicle, as well as the location of their home and location of their work (i.e. their trip origins and destination). This is not a short-term time scale and so the results should not be taken as an indication of the immediate future.

Several recommendations are suggested to address specific limitations of the model. The first involves the mode split model, which is focussed primarily on the choice between transit and AVRS. While the study uses all steps of the Four-Step model, the mode split component is of particular importance as the goal of the study is to examine the impact of a new mode (AVRS) on an existing mode (transit). Future research could develop a more comprehensive mode split model by including trip-maker characteristics such as wealth, location of home and work and other destinations, access to public transit, willingness to use new technology, and whether or not the trip-maker is an auto captive, transit captive, or neither. This could also allow for investigations to address different planning horizons by having some of the population unwilling or unable to use the new mode.

Another limitation to be addressed is that the model does not limit capacity on transit vehicles. As a result, the model outputs include some routes that have more passengers than the route could actually handle. A more complicated model could be constructed to limit the number of transit passengers specific to the type of transit vehicle and the number of passengers already using that transit vehicle, but this would require a less macroscopic model and could produce other unintended effects.

Traffic assignment models account for the effect of traffic volume on travel times through the use of volume-delay functions. In this research, delays at intersections (which

are not relevant for the highway link type) were modelled via the BPR function calibrated to represent the delay function for signalized intersections. An alternative approach would be to model the link travel time and the intersection delays separately. This could be accomplished by using the BPR function to model the link travel time and then add an additional travel time for the intersection delay. The intersection delay could also be a function of the type of turning movement being made (e.g. left turn vs right turn vs through movement). This approach would likely produce more realistic estimates of travel times.

A final major limitation of the model is that AVRS could have effects on the behavior of highways, roads, intersections, parking based on expected communication among autonomous vehicles and the volume of autonomous vehicles without passengers. Other researchers have investigated some of these areas and these findings could be included in subsequent models.

In addition to these suggestions for improving the model, another avenue of further research could investigate additional questions beyond the scope of this research. One area would be to investigate the research questions using network data from existing cities, which would eliminate the simplifications involved in using modelled networks and provide higher resolution data. Another possibility would be to use existing transportation software to take advantage of more complicated model structures, such as micro- or mesoscopic model structures. The experimental design of this research was a with-and-without study focussed on the introduction of AVRS. These impacts could also be studied with other experimental designs. For example, the model could be used to compare two scenarios, both with AVRS, that had small differences in the transit systems to isolate the impact of AVRS on specific pieces of transit systems. This could be done by comparing two scenarios, the second of which would have rail transit routes in the place of some of the bus transit routes of the first. Transit routes could also be added and removed. Finally, in addition to the seven route characteristics in the model, many other characteristics could be investigated. Of particular interest would be route characteristics related to the surroundings of the route such as land use measurements (population density within walking distance of transit routes, density of trip attractors within walking distance of transit routes), and land use types (residential, commercial, education, industrial, mixed).

References

- Algers, S., J. L. Dillén, Transek consultancy, and S. Widlert. (1995). "The National Swedish Value of Time Study." Östersund, Sweden: Swedish Institute for Transport and Communications Analysis (SIKA).
www.trafikdage.dk/td/papers/papers95/metode/algers/algers.pdf.
- Axhausen, K.W., A. König, G. Abay, J.J. Bates, and M. Bierlaire. (2006). "Swiss value of travel time savings." Zürich: ETH, Eidgenössische Technische Hochschule, IVT, Institut für Verkehrsplanung und Transportsysteme. <http://e-collection.ethbib.ethz.ch/view/eth:28966>.
- Batley, Richard, John Bates, Michiel Bliemer, Maria Börjesson, Jeremy Bourdon, Manuel Ojeda Cabral, Phani Kumar Chintakayala, et al. "New Appraisal Values of Travel Time Saving and Reliability in Great Britain." *Transportation* 46, no. 3 (April 2017): 583–621. <https://doi.org/10.1007/s11116-017-9798-7>.
- Bierlaire, Michel. "PandasBiogeme: a short introduction." Technical report TRANSP-OR 181219. Transport and Mobility Laboratory, ENAC, EPFL. (2019).
- Boiteux, M., and L. Baumstark. (2001). "Transports: choix des investissements et coût des nuisances." France: Commissariat général du plan. <http://www.ladocumentationfrancaise.fr/rapports-publics/014000434/index.shtml>.
- Bösch, Patrick M., Felix Becker, Henrik Becker, and Kay W. Axhausen. "Cost-Based Analysis of Autonomous Mobility Services." *Transport Policy* 64 (May 2018): 76–91. <https://doi.org/10.1016/j.tranpol.2017.09.005>.
- "Calgary Transit Scheduling Data | Open Calgary" City of Calgary. Accessed July 2018. <https://data.calgary.ca/Transportation-Transit/Calgary-Transit-Scheduling-Data/npk7-z3bj>
- "Developers | Société de transport de Montréal" Société de transport de Montréal. Accessed July 2018. <http://www.stm.info/en/about/developers>.
- Dictionary, Census of Population, 2016*. Ottawa, Ontario: Statistics Canada, 2018.
- Dijkstra, E. W. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1, no. 1 (1959): 269–71. <https://doi.org/10.1007/bf01386390>.
- Fagnant, Daniel J., and Kara M. Kockelman. "The Travel and Environmental Implications of Shared Autonomous Vehicles, Using Agent-Based Model Scenarios." *Transportation Research Part C: Emerging Technologies* 40 (March 2014): 1–13. <https://doi.org/10.1016/j.trc.2013.12.001>.
- Fagnant, Daniel J., Kara M. Kockelman, and Prateek Bansal. "Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market." *Transportation Research*

- Record: Journal of the Transportation Research Board* 2563, no. 1 (2016): 98–106.
<https://doi.org/10.3141/2536-12>.
- Fosgerau, M., K. Hjorth, S.V. Lyk-Jensen. 2007. "The Danish Value of Time Study: Final report (Report 5)." Lyngby, Denmark: Danish Transport Research Institute.
https://orbit.dtu.dk/fedora/objects/orbit:81617/datastreams/file_4049265/content.
- Fu, Liping. *Modelling Travel Demand for Urban Transportation Planning: CIVE 640*. University of Waterloo, 2016.
- Gentile, Guido. "Local User Cost Equilibrium: a Bush-Based Algorithm for Traffic Assignment." *Transportmetrica A: Transport Science* 10, no. 1 (2012): 15–54.
<https://doi.org/10.1080/18128602.2012.691911>.
- Ghiasi, Amir, Omar Hussain, Zhen (Sean) Qian, and Xiaopeng Li. "A Mixed Traffic Capacity Analysis and Lane Management Model for Connected Automated Vehicles: A Markov Chain Method." *Transportation Research Part B: Methodological* 106 (December 2017): 266–92. <https://doi.org/10.1016/j.trb.2017.09.022>.
- "GTFS Data" Translink. Accessed July 2018.
<https://developer.translink.ca/ServicesGtfs/GtfsData>.
- "GTFS Static Overview | Static Transit | Google Developers." Google. Accessed July 2018.
<https://developers.google.com/transit/gtfs>.
- Gwilliam, Kenneth M. 1997. "The value of time in economic evaluation of transport projects (English)". Infrastructure notes; no. OT-5. Washington DC; World Bank. (1997).
<http://documents.worldbank.org/curated/en/759371468153286766/The-value-of-time-in-economic-evaluation-of-transport-projects>.
- Hanna, Josiah P., Michael Albert, Donna Chen, and Peter Stone. "Minimum Cost Matching for Autonomous Carsharing." *IFAC-PapersOnLine* 49, no. 15 (2016): 254–59.
<https://doi.org/10.1016/j.ifacol.2016.07.757>.
- "HSR Transit Feed | Open Hamilton" City of Hamilton. Accessed July 2018.
<http://open.hamilton.ca/datasets/6eeccf172c824c2db0484aea54ed7fe4>.
- Levin, Michael W., and Stephen D. Boyles. "Effects of Autonomous Vehicle Ownership on Trip, Mode, and Route Choice." *Transportation Research Record: Journal of the Transportation Research Board* 2493, no. 1 (2015): 29–38.
<https://doi.org/10.3141/2493-04>.
- Liang, Xiao, Gonçalo Homem De Almeida Correia, and Bart Van Arem. "Optimizing the Service Area and Trip Selection of an Electric Automated Taxi System Used for the Last Mile of Train Trips." *Transportation Research Part E: Logistics and Transportation Review* 93 (September 2016): 115–29.
<https://doi.org/10.1016/j.tre.2016.05.006>.

- Ma, Jiaqi, Xiaopeng Li, Fang Zhou, and Wei Hao. "Designing Optimal Autonomous Vehicle Sharing and Reservation Systems: A Linear Programming Approach." *Transportation Research Part C: Emerging Technologies* 84 (November 2017): 124–41.
<https://doi.org/10.1016/j.trc.2017.08.022>.
- Ohnemus, Michael, and Anthony Perl. "Shared Autonomous Vehicles: Catalyst of New Mobility for the Last Mile?" *Built Environment* 42, no. 4 (2016): 589–602.
<https://doi.org/10.2148/benv.42.4.589>.
- "Open data - Grand River Transit" Grand River Transit. Accessed July 2018.
<https://www.grt.ca/en/about-grt/open-data.aspx>.
- "Open Data - London Transit Commission" London Transit Commission. Accessed July 2018. <http://www.londontransit.ca/open-data/>.
- Ortúzar S., Juan De Dios., and Luis G. Willumsen. *Modelling Transport*. 4th ed. Oxford: Wiley-Blackwell, 2011.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (October 2011): 2825–30.
- Ramjerdi, F., L. Rand, I-A. F. Sætermo, and K. Saelensminde. (December 1997). "The Norwegian Value of Time Study." Oslo: Transportøkonomisk institutt.
<http://www.toi.no/article18334-29.html>.
- Teply, S., and J. W. Gough. *Canadian Capacity Guide for Signalized Intersections*. 3rd ed. Washington, DC: Institute of Transportation Engineers, District 7 - Canada, 2007.
- Transportation Research Board. (2000). *Highway Capacity Manual 4th Edition*. National Academies of Science, USA.
- "TTC Routes and Schedules - City of Toronto Open Data Portal" City of Toronto. Accessed July 2018. <https://open.toronto.ca/dataset/ttc-routes-and-schedules/>.
- UK Department for Transport. (2009). *Values of Time and Operating Costs: TAG Unit 3.5.6*. UK Department for Transport, Transport Analysis Guidance (TAG).
<http://www.dft.gov.uk/webtag/documents/expert/unit3.5.6.php>.
- US Department of Transportation. (2016). *The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations*.
- VanderPlas, Jake. *Python Data Science Handbook: Essential Tools for Working with Data*. Sebastopol, CA: O'Reilly, 2016.
- Vuchic, Vukan R. *Urban Transit Systems and Technology*. Hoboken, N.J.: J. Wiley, 2007.
- Winter, Konstanze, Oded Cats, Gonçalo Homem De Almeida Correia, and Bart Van Arem. "Designing an Automated Demand-Responsive Transport System: Fleet Size and Performance Analysis for a Campus–Train Station Service." *Transportation Research*

Record: Journal of the Transportation Research Board 2542, no. 1 (2016): 75–83.
<https://doi.org/10.3141/2542-09>.

Winter, Konstanze, Oded Cats, Bart Van Arem, and Karel Martens. “Impact of Relocation Strategies for a Fleet of Shared Automated Vehicles on Service Efficiency, Effectiveness and Externalities.” *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2017.
<https://doi.org/10.1109/mtits.2017.8005630>.

Zhang, A., A.E. Boardman, D. Gillen, and W.G. Waters II. (2005). “Towards Estimating the Social and Environmental Costs of Transportation in Canada.” Centre for Transportation Studies, University of British Columbia, Vancouver BC.
<http://www.tc.gc.ca/pol/en/report/fullcostinvestigation/transmodal/t003/t003.pdf>.

Appendix

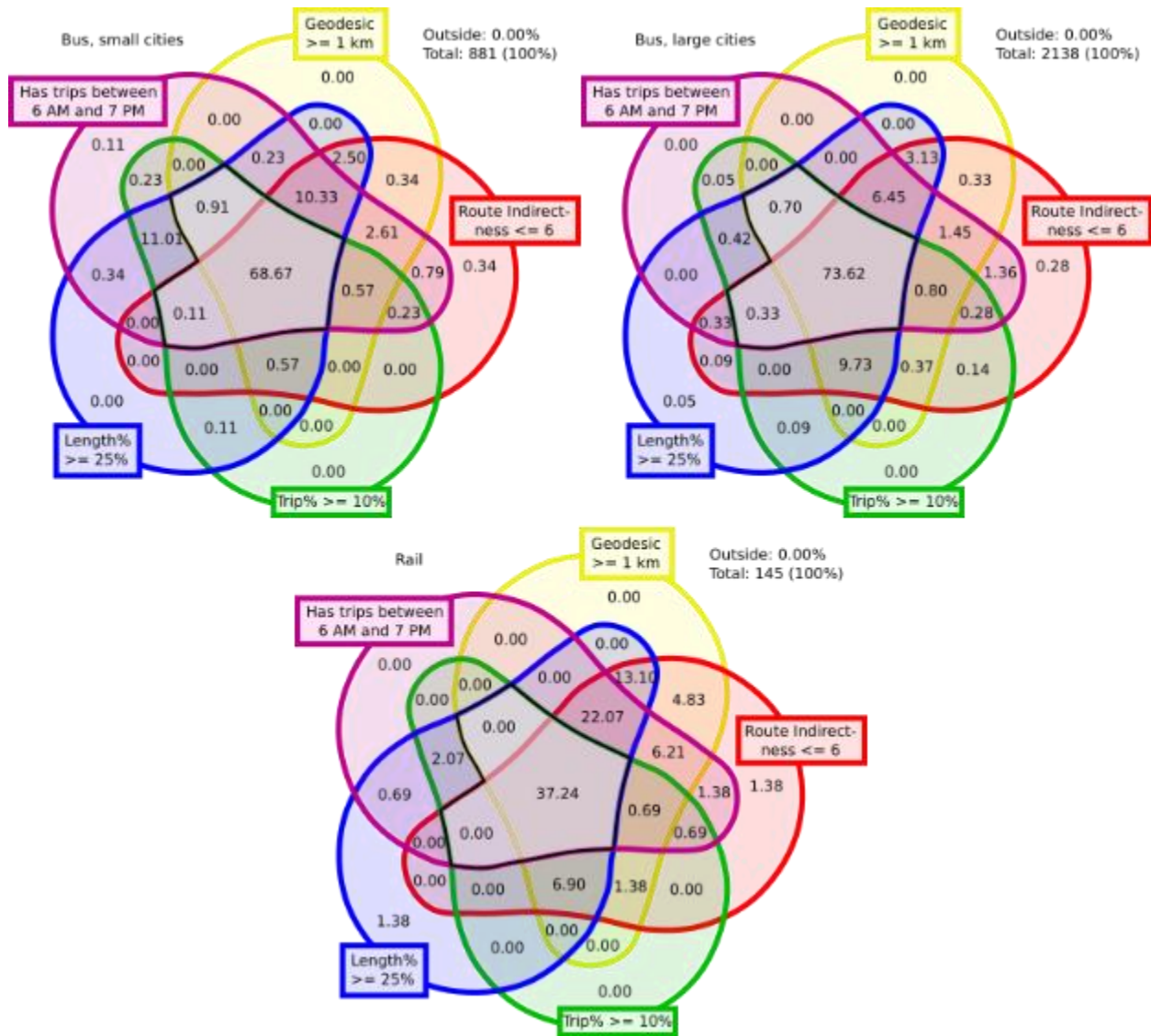


Figure 4.1: Venn diagrams showing filtering logic. Numbers in diagram represent percent of total subroutes for that data set. The accepted subroutes are shown in black-outlined shape in the center. See section 2.1.1.2.

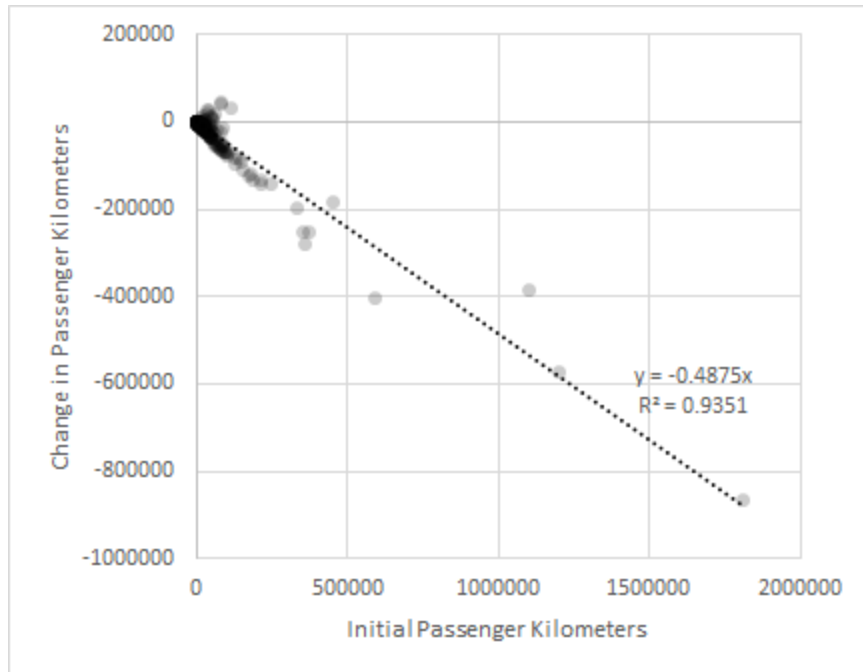


Figure 4.2: Extended view of Figure 3.1 including all data points.

Table 4.1: Transit route dataset summary for bus routes in small cities

| Aggregation | Trip Count | Route Length (km) | Geodesic (km) | Total Travel Time (min) | Headway (min) | Operating Speed (km/h) | Wait Time Percentage | Route Indirectness |
|---------------|------------|-------------------|---------------|-------------------------|---------------|------------------------|----------------------|--------------------|
| Percentiles: | | | | | | | | |
| 0 (Minimum) | 2.0 | 1.77 | 0.99 | 6.0 | 5.2 | 11.8 | 4.74 | 1.03 |
| 5 | 2.0 | 4.68 | 2.26 | 12.0 | 11.3 | 16.5 | 9.10 | 1.21 |
| 25 | 9.0 | 8.65 | 4.21 | 22.0 | 18.4 | 20.7 | 19.45 | 1.45 |
| 50 (Median) | 27.0 | 12.52 | 7.09 | 31.4 | 24.2 | 23.3 | 33.65 | 1.73 |
| 75 | 41.0 | 17.37 | 9.74 | 44.0 | 30.2 | 27.3 | 64.98 | 2.16 |
| 95 | 68.0 | 25.54 | 16.57 | 63.9 | 85.0 | 33.4 | 150.00 | 3.43 |
| 100 (Maximum) | 147.0 | 39.56 | 23.4 | 87.3 | 90.0 | 47.7 | 281.25 | 14.66 |
| Mean | 28.2 | 13.43 | 7.57 | 34.1 | 29.2 | 24.2 | 51.71 | 2.01 |
| Median | 27.0 | 12.52 | 7.09 | 31.4 | 24.2 | 23.3 | 33.65 | 1.73 |
| Mode | 2.0 | 1.77 | 4.72 | 30.0 | 30.0 | 11.8 | 106.25 | 1.03 |

Table 4.2: Transit route dataset summary for bus routes in large cities

| Aggregation | Trip Count | Route Length (km) | Geodesic (km) | Total Travel Time (min) | Headway (min) | Operating Speed (km/h) | Wait Time Percentage | Route Indirectness |
|---------------|------------|-------------------|---------------|-------------------------|---------------|------------------------|----------------------|--------------------|
| Percentiles: | | | | | | | | |
| 0 (Minimum) | 2.0 | 0.91 | 0.41 | 5.0 | 2.0 | 8.1 | 1.75 | 1.00 |
| 5 | 5.0 | 3.62 | 2.01 | 11.8 | 5.4 | 12.6 | 4.40 | 1.07 |
| 25 | 17.0 | 7.02 | 4.12 | 22.7 | 9.5 | 16.5 | 9.90 | 1.24 |
| 50 (Median) | 39.0 | 10.73 | 6.46 | 32.0 | 17.1 | 20.4 | 21.24 | 1.45 |
| 75 | 63.0 | 15.25 | 10.77 | 45.0 | 27.8 | 24.1 | 39.25 | 1.81 |
| 95 | 127.2 | 24.28 | 18.06 | 62.5 | 42.2 | 32.2 | 114.58 | 3.22 |
| 100 (Maximum) | 306.0 | 46.91 | 31.26 | 93.8 | 90.0 | 68.6 | 312.45 | 9.21 |
| Mean | 47.3 | 11.93 | 8.02 | 34.0 | 19.9 | 21.1 | 33.30 | 1.72 |
| Median | 39.0 | 10.73 | 6.46 | 32.0 | 17.1 | 20.4 | 21.24 | 1.45 |
| Mode | 5.0 | 3.31 | 6.19 | 15.0 | 30.0 | 8.1 | 125.00 | 1.04 |

Table 4.3: Transit route dataset summary for rail routes

| Aggregation | Trip Count | Route Length (km) | Geodesic (km) | Total Travel Time (min) | Headway (min) | Operating Speed (km/h) | Wait Time Percentage | Route Indirectness |
|---------------|------------|-------------------|---------------|-------------------------|---------------|------------------------|----------------------|--------------------|
| Percentiles: | | | | | | | | |
| 0 (Minimum) | 2.0 | 3.81 | 2.95 | 5.5 | 1.6 | 11.9 | 1.44 | 1.00 |
| 5 | 15.0 | 3.94 | 3.32 | 8.4 | 2.5 | 12.4 | 1.74 | 1.02 |
| 25 | 140.0 | 7.39 | 5.82 | 25.0 | 4.0 | 14.0 | 2.99 | 1.11 |
| 50 (Median) | 183.0 | 15.09 | 8.81 | 40.0 | 5.0 | 31.2 | 4.72 | 1.20 |
| 75 | 224.2 | 26.18 | 16.62 | 55.6 | 6.3 | 37.8 | 12.82 | 1.37 |
| 95 | 297.5 | 34.15 | 24.23 | 75.4 | 11.6 | 42.1 | 34.43 | 4.87 |
| 100 (Maximum) | 357.0 | 38.90 | 28.72 | 78.8 | 12.0 | 42.5 | 41.67 | 5.23 |
| Mean | 170.1 | 17.35 | 11.59 | 39.5 | 5.4 | 27.9 | 9.70 | 1.70 |
| Median | 183.0 | 15.09 | 8.81 | 40.0 | 5.0 | 31.2 | 4.72 | 1.20 |
| Mode | 140.0 | 3.81 | 3.21 | 25.0 | 5.9 | 11.9 | 1.44 | 1.11 |

Table 4.4: Mode usage for modelled cities

| City | Mode | Without AVRS | | With AVRS | |
|------|---------|-------------------|--------------|-------------------|--------------|
| | | Trips (number) | Trips (%) | Trips (number) | Trips (%) |
| A | Auto | 71,867 | 87.5 | 5,648 | 6.9 |
| A | Transit | 10,259 | 12.5 | 2,939 | 3.6 |
| A | AVRS | | | 73,539 | 89.5 |
| B | Auto | 62,924 | 76.8 | 5,157 | 6.3 |
| B | Transit | 18,998 | 23.2 | 5,934 | 7.2 |
| B | AVRS | | | 70,831 | 86.5 |
| C | Auto | 125,695 | 76.7 | 10,170 | 6.2 |
| C | Transit | 38,199 | 23.3 | 18,731 | 11.4 |
| C | AVRS | | | 134,992 | 82.4 |
| D | Auto | 84,359 | 51.0 | 8,233 | 5.0 |
| D | Transit | 81,070 | 49.0 | 48,010 | 29.0 |
| D | AVRS | | | 109,186 | 66.0 |
| E | Auto | 200,070 | 61.1 | 19,543 | 6.0 |
| E | Transit | 127,598 | 38.9 | 58,751 | 17.9 |
| E | AVRS | | | 249,373 | 76.1 |
| F | Auto | 184,395 | 56.3 | 18,118 | 5.5 |
| F | Transit | 143,258 | 43.7 | 66,052 | 20.2 |
| F | AVRS | | | 243,480 | 74.3 |

Table 4.5: Trip Characteristics for modelled cities

| City | Mode | Without AVRS | | | | With AVRS | | | |
|------|---------|-------------------------|----------------------------|------------------|---------------------|-------------------------|----------------------------|------------------|---------------------|
| | | Average IVTT (min/trip) | Average Distance (km/trip) | Total IVTT (min) | Total Distance (km) | Average IVTT (min/trip) | Average Distance (km/trip) | Total IVTT (min) | Total Distance (km) |
| A | Auto | 8.22 | 7.30 | 590,787 | 524,761 | 7.04 | 5.99 | 39,746 | 33,817 |
| A | Transit | 7.83 | 3.25 | 80,377 | 33,388 | 5.94 | 2.45 | 17,470 | 7,212 |
| A | AVRS | | | | | 7.99 | 7.01 | 587,580 | 515,316 |
| B | Auto | 8.34 | 8.66 | 524,980 | 544,645 | 6.97 | 6.76 | 35,960 | 34,838 |
| B | Transit | 12.10 | 5.66 | 229,858 | 107,528 | 10.87 | 5.00 | 64,525 | 29,685 |
| B | AVRS | | | | | 8.12 | 8.22 | 575,186 | 582,157 |
| C | Auto | 8.97 | 7.85 | 1,127,188 | 986,811 | 7.38 | 6.07 | 75,049 | 61,757 |
| C | Transit | 7.88 | 4.11 | 300,853 | 157,143 | 7.05 | 4.08 | 131,997 | 76,429 |
| C | AVRS | | | | | 8.49 | 7.26 | 1,145,940 | 979,686 |
| D | Auto | 8.17 | 7.70 | 689,339 | 649,628 | 7.06 | 6.09 | 58,142 | 50,161 |
| D | Transit | 12.13 | 7.52 | 983,666 | 609,298 | 11.11 | 7.34 | 533,561 | 352,500 |
| D | AVRS | | | | | 8.00 | 7.16 | 873,759 | 781,718 |
| E | Auto | 9.18 | 7.76 | 1,836,474 | 1,553,093 | 6.83 | 5.35 | 133,447 | 104,541 |
| E | Transit | 12.14 | 5.85 | 1,548,948 | 745,928 | 8.81 | 4.43 | 517,784 | 260,158 |
| E | AVRS | | | | | 7.79 | 6.36 | 1,943,856 | 1,585,346 |
| F | Auto | 8.98 | 8.19 | 1,655,903 | 1,510,434 | 7.53 | 6.33 | 136,494 | 114,651 |
| F | Transit | 14.78 | 8.21 | 2,116,854 | 1,176,061 | 13.77 | 7.95 | 909,305 | 525,305 |
| F | AVRS | | | | | 8.80 | 7.66 | 2,142,626 | 1,866,032 |