Abstract

Accessibility Beyond the Schedule

Nathan Wessel
Doctor of Philosophy
Graduate Department of Geography & Planning
University of Toronto
2019

The study of accessibility - the ease of reaching destinations - by public transport has made huge advances thanks to the availability of standardized, routable transit schedule data. The General Transit Feed Specification (GTFS) has provided researchers with a vast trove of machine-readable data allowing for highly detailed spatio-temporal modelling of scheduled transit operations. Yet it is well established that in the real-world schedules are imperfect - vehicles often run late, get bunched, miss transfers, arrive too full for anyone to board, and otherwise behave in predictably unpredictable ways. Schedule data alone cannot possibly account for this distinctly stochastic component of much transit service, which to date has been considered separate from accessibility analysis under the umbrella of “reliability”. This dissertation takes the perspective that transit service is reliably unreliable and will continue to be so until humans are taken out of the equation. Detailed observations of actual service can be used to construct more realistic models for estimating travel times and thus accessibility via transit.

Chapter 2 introduces a novel method of converting a detailed GPS record of transit fleet locations into a retrospective GTFS package. This backward-looking “schedule” format allows the same tools developed for schedule-based GTFS analysis to be applied in Chapter 3 to a more accurate depiction of actual transit accessibility. The findings indicate that models of transit accessibility based on schedule data alone tend to produce substantial overestimates of accessibility and systematic spatial errors by failing to account for normal irregularities in service provision. Chapter 4 points toward a way of better suiting available GTFS analysis tools to actual transit service by addressing the problem of imperfect information in modelled route choice. The travel time implications for a large minority of trips are shown to be substantial.

Transit accessibility research has come a long way in the last decade and has a long way yet to go. Models based on schedule data alone should give way in many cases to models based on service as actually provided, acknowledging that schedules may guide but rarely constrain the transit services that passengers actually use every day.
To everyone on the bus - may we get there in time.

Acknowledgements

While many take this opportunity to thank the important people in their lives, I hope such thanks to mine have already been sufficiently expressed. Rather, I would like to thank those who I have thanked too rarely or not at all: the people who drive the buses, plan the routes, come to the public hearings, scrape the gum off the platforms, write the nerdy blogs, and otherwise work to advance the quality of public transit in their cities.
Preface

This dissertation can be read as a complete and coherent document, however Chapters 2, 3, & 4 were originally written as separate articles to be published in peer-reviewed academic journals. As such there is necessarily some redundancy in the introductory material, literature reviews, and methods sections of each. As each of the articles provides some of its own context, the goal of Chapter 1 of this dissertation is to frame the issues motivating this research as a coherent body of work. Similarly Chapter 5 serves to tie together the results of each article into broader implications for the study of transit accessibility and to suggest future lines of research.

As of this writing, Chapter 2 has been published in the Journal of Transport Geography as Wessel, Allen, and Farber (2017). Chapter 3 was just accepted by the Journal of Transport and Landuse, and is available online as a pre-proof (Wessel and Farber, 2019a). Chapter 4 is also available as a preproof (Wessel and Farber, 2019b) and has been submitted for review as a coauthored paper with Steven Farber. The bulk of the work on these chapters was my own, though I would like to thank Jeff Allen for his technical assistance on the analysis in Chapter 2 and Steven Farber who, as my advisor, contributed extensive comments and reviews to all parts of this work.
## Contents

**Abstract** ii  
**Acknowledgements** iii  
**Preface** iv  

### 1 Introduction 1

1.1 From Trip Planning to Accessibility Analysis ................................. 3  
1.1.1 Development of the GTFS ............................................. 3  
1.1.2 GTFS Tools Applied to Accessibility Analysis .......................... 3  
1.1.3 Current State of the Art .............................................. 4  
1.2 Utility of the Current Transit Accessibility Model ......................... 5  
1.2.1 Pre-GTFS Accessibility Studies ...................................... 5  
1.2.2 Applications and Extensions .......................................... 5  
1.3 Means of Improving the Accessibility Models ................................. 6  
1.3.1 Retrospective Data as a Forecast of Future Stochastic Service ......... 6  
1.3.2 Route-choice Without Omniscience ................................... 7  

### 2 Retrospective GTFS 9

2.1 Abstract ................................................................. 9  
2.2 Introduction ........................................................... 9  
2.3 Required Data ......................................................... 10  
2.4 Algorithm ..................................................................... 11  
2.4.1 Outline ................................................................. 11  
2.4.2 Collecting and storing the data ........................................ 11  
2.4.3 Delimiting Trips and Blocks ......................................... 12  
2.4.4 Spatial Matching and Positional Error Handling ....................... 12  
2.4.5 Determining Stop Times .............................................. 14  
2.4.6 Constructing the Retrospective GTFS Package ......................... 15  
2.5 Accessibility Case Study .................................................. 15  
2.6 Discussion ................................................................... 17  

### 3 The Accuracy of Schedule Data for Accessibility 19

3.1 Abstract ..................................................................... 19  
3.2 Introduction .................................................................. 19
Chapter 1

Introduction

This research seeks to advance the development of models of public transit used for the study/analysis of accessibility, where accessibility is defined as a measure of the relative ease of reaching destinations. There has been a profusion of work on transit accessibility in the last decade or so, due in large part to a sudden influx of publicly available schedule data in a machine-readable format known as the General Transit Feed Specification or GTFS. While such data has led to some incredibly useful and interesting applications, its standard format and the suite of tools developed around it have made it easy to apply a model of transit suited for one purpose to another for which it is troublingly flawed. Specifically, GTFS was designed to allow computers to guide passengers in the selection of optimal routes based on schedules that were often too complex for people to manage on their own without difficulty. Once online trip-planning tools became normal, it was an easy conceptual leap from asking the computer for one optimal future transit trip and estimated travel time - to asking the computer for all of them. Accessibility studies based on such methods proliferated with few acknowledging that the methods and data suited to route planning are not the methods and data best suited to creating a practical accessibility measure.

Prior to this boom in accessibility research, transport engineers had for years been developing highly detailed and increasingly realistic models of public transport for applications such as demand modelling, land-use planning and operations management (e.g. Cats, 2011). Such models however are typically developed within metropolitan planning organizations and built around idiosyncratic local data not available to the general public or to outside researchers. Thus the wide availability of detailed and standardized transit schedule data in the form of GTFS represented a real opportunity for transport researchers who could for the first time collaborate with open-source tools to conduct analysis across borders and between regions. This massive advance in data availability however came at the cost of data quality and model validity. At a conceptual level, the fundamental problem with current accessibility models is that they produce estimates of travel times that are theoretically achievable under fully predictable conditions. Transit service however often remains stubbornly unpredictable and the people using it are not rational actors making decisions with perfect information. Any practical study of accessibility should have an interest in what is practically possible given the way things are and a model that shows us only what is literally possible in sterile conditions will not take us all the way there.

In this sense, current models of transit accessibility may be much like early models of automobile accessibility. When large graphs of street networks first became available for analysis in GIS there were numerous studies of accessibility based only on shortest-path network distance and posted speed limits.
While perhaps accurate in some rural places, such a measure would be optimistic in any city with more than a few stop lights and absurd in a congested metropolis. Models of automobile accessibility have developed a lot since then and the better ones now include edge-level estimates of traffic congestion by time of day (e.g. Moya-Gómez and García-Palomares, 2015) and even day of week. Such estimates are based on huge volumes of GPS probe data, available thanks to the fact that many people now carry an internet connected GPS device. Just so, research on transit accessibility is still in its early days, reacting with enthusiasm to a big new source of data and using largely off-the-shelf algorithms adapted from structurally similar, but conceptually different applications. To achieve a more realistic measure of transit accessibility, our methods must be refined with new sources of data and new ways of working with it.

This dissertation outlines a middle path between the extreme availability but poor validity of GTFS-based models and the better validity but poor availability of much more narrowly tailored applied transport models. It does this by making use of so-called “real-time” transit datasets which are themselves designed, like GTFS, to aid passenger navigation and manage expectations about travel and wait times. These real-time data are based on live Internet access to Automatic Vehicle Location (AVL) systems and often include precise locations of all active vehicles in an agency’s fleet. By monitoring these online data feeds and archiving their contents it is possible to create an accurate database of actual historical fleet movements. Such datasets have often been used to produce measures of reliability at the level of stops and routes (e.g. Nguyen, Diab, and Shalaby, 2018; W. X. Hu and Shalaby, 2016), though in the past these have typically been supplied to researchers only on request and in non-standard formats. Such an AVL location database can be coerced into a format structurally comparable to GTFS and thus amenable to the kind of standardized network-level travel time analysis enabled by standard accessibility tools. In the limiting case where all operations are perfectly on-schedule such a dataset, termed a “retrospective GTFS”, would be functionally equivalent to a schedule-based GTFS dataset for a given time period. In all other cases however, it would implicitly incorporate data on reliability relative to the schedule and model performance as it actually occurs.

This research outlines two key ways that accessibility models can be improved by the use of such data and makes an empirical case that those changes contribute to a more realistic standard measure of accessibility. First, I argue that it is not sufficient to use schedule data alone to model a transit system because it is actually quite rare for an agency to adhere closely to its own schedules. Current models treat transit systems like a piece of clockwork with precise timings and predetermined and foreknowable outcomes. Second, because transit does not adhere to a strict schedule, passengers often experience transit as a fundamentally stochastic phenomenon. This means that they do not have the ability to optimize their journeys in the same way as a computer algorithm which has perfect knowledge of a schedule. Instead, passengers rely on an assortment of heuristic methods for optimizing travel time - among other criteria such as stress, risk aversion, and cognitive load.

Before getting into the details of this research, it is helpful to understand something about the way current models came to be and where research on transit accessibility may be heading. The following section covers the growth and development of the GTFS standard, the trip-planning applications which developed around it, and the eventual use of these applications for accessibility analysis. Section 1.2 addresses the advance these models represented over earlier methods and discusses some of the interesting

---

1 Some researchers are now looking at modelling the time needed to find parking as well, something few car trips are without (e.g. Tallis, 2014).
Chapter 1. Introduction

1.1 From Trip Planning to Accessibility Analysis

1.1.1 Development of the GTFS

The General Transit Feed Specification (GTFS) was originally developed in 2005 as a means of enabling automatic end-to-end trip planning for public transit networks. Popular online trip-planning applications like MapQuest and Google Maps were already handling millions of requests for driving directions and there was interest in accommodating transit routing in the same tools (McHugh, 2013). Some agencies had built their own online trip-planners by this time though these were relatively rare and idiosyncratic (Ferris, Watkins, and Borning, 2009), sometimes difficult for new passengers to find and use. Google Transit, when it launched, integrated transit directions into a single, already-popular application. Instead of making their own trip-planning application agencies could now just supply the data and let someone else take care of it for them. A few major agencies quickly did just that (Antrim, Barbeau, et al., 2013), leading to the rapid consolidation of GTFS (originally: Google Transit Feed Specification) as the de facto standard for supplying schedule data to online trip-planning applications.

Around the same time, a general movement toward open civic data was taking root and many agencies followed the lead of Portland’s TriMet in releasing their GTFS data to the public under relatively permissive licenses. Over the last decade more than a thousand transit agencies around the world have published GTFS schedule data online (Zervaas, 2018) and as of 2019 it is actually unusual in many countries for any moderately sized transit agency not to publish their schedule data in GTFS format and make it freely available online. It is important to remember though that the primary impetus for this rapid growth is that once published, the data becomes available to passengers through a range of intermediary trip-planning applications.

1.1.2 GTFS Tools Applied to Accessibility Analysis

Only a few years after the initial development of GTFS, the academic literature on transit accessibility was beginning to take notice of the new kinds of analysis made possible by the availability of detailed digital schedule data. Ferris, Watkins, and Borning (2009) were some of the first to use GTFS data for applied accessibility analysis, discussing how they built a web-based trip-planning application which was able to identify points of interest within a given travel time and display an isochrone map to the user. Lei and Church (2010) address the measurement of transit accessibility systematically, and though they do not actually mention that they made use of GTFS data in their study, they propose a model of accessibility based essentially on the kind of schedule-based trip-planning software that was becoming increasingly common at the time. McGurrin and Greczner (2011) outline an open-source toolchain based on OpenTripPlanner and using data from OpenStreetMap and GTFS. OpenTripPlanner was originally developed by Portland’s TriMet as a transit trip-planning tool, providing an open-source alternative to the proprietary Google Maps platform. McGurrin and Greczner however propose that this software can be adapted to the purposes of accessibility analysis essentially by making thousands of requests to the trip-planning API in order to retrieve travel time estimates. Following shortly after these methodological
developments, Tomer et al. (2011) report using an entirely different software and an astonishing 371 GTFS datasets to measure access to jobs by transit across 100 major US metro areas. It is not clear whether this software was purpose built with accessibility analysis in mind, or if, like McGurrin and Greczner (2011), the authors adapted a trip-planning application for another purpose. The broad scope of the study however clearly illustrates the rapid growth of GTFS, which by this point had become standard for even mid-sized US transit agencies.

These initial studies showed a great deal of promise and by 2013, OpenTripPlanner was in the process of being adapted to function as an analysis tool, with the introduction of the “analyst extension” (Antrim, Barbeau, et al., 2013). This placed the expectation for analysis capabilities on exactly the same back-end software responsible for route planning. Around the same time, Melinda Morang at ESRI was developing what would eventually become a suite of tools for transit analysis which were beginning to be used for accessibility analysis (Tallis, 2014; Farber, Morang, and Widener, 2014). These tools, like OpenTripPlanner were based essentially on a trip-planning framework, suggesting optimal routes and estimated travel times based only on the GTFS schedule data.

These developments are very much in line with the suggestions of Lei and Church (2010), who explicitly critique previous models of transit travel time estimation up to that point and lay out a vision for improved models of accessibility based on the integration of detailed spatio-temporal schedule data into temporally-aware GIS applications.

1.1.3 Current State of the Art

While not much time has passed since the first transit accessibility studies entered the literature, there has been some methodological development since Tomer et al. (2011), beside ongoing development of the route-planning software itself. Conway, Byrd, and Linden (2017) and Conway, Byrd, and Eggermond (2018) for example address the problem of unfixed departure times for headway-based schedules, especially in the context of proposed development scenarios for transport planning. Others address themselves to the computational complexity of travel time estimation with Stepniak et al. (2019) measuring the implications of alternate temporal sampling resolutions, finding that the increasingly standard method of estimating minutely departures (e.g. Farber and L. Fu, 2017) may be more than is required for many practical applications. Similarly, Owen and Murphy (2018b) discuss the issue of sampling frequencies, noting that harmonics can develop between evenly spaced scheduled headways and evenly spaced sampling intervals. Thus the issue of sampling and specification has been a concern, with relatively little attention given to the validity of the underlying model, especially as regards issues of reliability.

The most interesting methodological development in the context of this work is the dissertation by Stewart (2017a), which makes use of a GTFS-structured dataset based on observed vehicle locations rather than schedule data. Stewart’s work was essentially contemporaneous with Chapter 2 of this dissertation (i.e. Wessel, Allen, and Farber, 2017), though it did not make use of the same software and the data used in that analysis was also not retrieved through open means but was supplied by the agency on request in an unspecified format. In his analysis Stewart otherwise uses standard modelling assumptions, but bases an applied accessibility analysis in two major cities on this non-schedule-based data.

\(^2\)They used “Traffic Analyst”, a proprietary ArcMap extension apparently capable of modelling transit journeys from GTFS.
1.2 Utility of the Current Transit Accessibility Model

The current standard way of modelling transit accessibility on a dynamic spatio-temporal, schedule-based network, using trip-planning software has obvious advantages over the more simplistic methods that came before. This modelling paradigm was advanced explicitly by (Lei and Church, 2010) and they, and the research that followed, make a convincing argument for its utility.

1.2.1 Pre-GTFS Accessibility Studies

The earliest studies of accessibility involving transit treated stops and stations themselves as a destination to be accessed via non-transit modes such as walking or driving (e.g. Shaw, 1991). A thorough review of such studies can be found in Y. Lee et al. (2001) and Lei and Church (2010), though these will not be addressed here as they did not measure accessibility to destinations via transit. Studies of accessibility via transit before the advent of GTFS were relatively rare and relied on data from a variety of sources. In one of the first papers to address the topic Cervero, Rood, and Appleyard (1999) sidestepped the issue entirely, saying:

...the analysis focuses on job accessibility via highways. No attempt was made to incorporate distances via mass transit into the analysis because, given the region’s predominantly bus-based transit system, most regional transit trips are made over the same highway network.

From a modern perspective, such an approach has obvious and remediable flaws. Thus, when Shen (2001) was able to make use of a transit-specific travel time matrix provided by a metropolitan planning organization, we seem to see an improvement in the attention given to transit as a distinct mode, even if the researcher never specifies how this matrix was compiled. Thompson (2001) uses a similar dataset, in this case borrowed from a regional transport model of unspecified provenance. A few years later, (Grengs, 2004) indicated that “transit maps and schedules” were used to estimate travel times, but again did not indicate how exactly this was accomplished or even what time of day was being modelled. Given the smaller scale of the study it is likely that travel times between points were estimated manually using paper maps and schedules.

1.2.2 Applications and Extensions

Emerging from this context, the current schedule-based modelling paradigm is a major advance. For example, the explicit modelling of departure times and transfers allows for a highly detailed analysis of variation in transit service through time (Farber, Morang, and Widener, 2014; Fransen et al., 2015). The standardization of data and thus methods allows the direct comparison of large numbers of transit agencies and the integration of whole regions (Tomer et al., 2011; Owen and Levinson, 2016; Allen and Farber, 2018a). It also readily accommodates accessibility forecasting related to proposed schedule/service changes (Conway, Byrd, and Linden, 2017).

Perhaps the biggest advantage however is that the model, better than previous models, readily maps onto people’s actual experience of using transit by explicitly modelling all legs of the trip including walking, waiting, transfers, even boarding times etc. and allows routes to be constrained, just as it can be in trip-planning software to a reasonable complexity or cost or other minimizable criteria.

Such features have led to the use of transit accessibility modelling in a variety of domains including the study of spatial equity (Farber, Morang, and Widener, 2014; Pereira et al., 2018; Fransen et al., 2015;
El-Geneidy, Buiung, et al., 2016), comparative planning forecasts (Ma and Jan-Knaap, 2014; Farber and Grandez, 2016; J. Lee and H. J. Miller, 2018), and inputs to various urban activity models (Allen and Farber, 2018b; Owen and Levinson, 2015).

1.3 Means of Improving the Accessibility Models

The purpose of this research is not to critique the current paradigm out of existence, but to suggest ways of transcending certain limitations now that the benefits of a detailed spatio-temporal model of transit networks have been well demonstrated in the context of accessibility analysis. The data and computing resources are now available to construct much more realistic and thus more practically useful models. I begin by making the case that schedule data is a poor substitute for actual observations of transit service and provide in Chapter 2 a method for using standardized automatic vehicle location (AVL) data to construct a routable retrospective GTFS-formatted dataset. Chapter 3 applies standard accessibility tools to a comparative analysis of such retrospective and schedule-based GTFS data, offering evidence that the two are substantially different in a variety of contexts. Finally, Chapter 4 addresses the problem of modelling route-choice in fundamentally stochastic networks, giving evidence that the optimal path returned by trip-planning software may have substantially different travel time outcomes than the path a passenger is likely to choose in reality. Taken together these suggestions encompass both supply-side and demand-side improvements to the standard accessibility model.

1.3.1 Retrospective Data as a Forecast of Future Stochastic Service

It is difficult to overstate the fact that transit services do not always quite adhere to their posted schedules. This is born out by an extensive literature on reliability (e.g. W. X. Hu and Shalaby, 2016) and by the widespread use of real-time passenger information systems (e.g. Watkins et al., 2011). That schedules have dominated transit accessibility models does not so much reflect the theoretical validity of their use in that context but the ready availability of their substance. Automatic vehicle location systems (AVLs) provide a potentially rich source of data on actual fleet operations and are commonly used in dispatching vehicles and updating passenger information systems in real-time. Though AVLs are widely implemented the data generated by these systems has generally only been stored by the agency managing the system and has not been made available outside of local research communities if at all. Such datasets have been used primarily for applied research on reliability and operations management strategies (e.g. Nguyen, Diab, and Shalaby, 2018).

Chapter 2 develops a method of deriving a standard archive of AVL data from public-facing Application Programming Interfaces (APIs) and of converting this data (sequential GPS point observations with some route and vehicle metadata) into a format suitable for network-level routing analysis. This is accomplished by map-matching the relatively sparse GPS point data to a street/rail network and interpolating stop times at all scheduled stops along a given route. Sequences of stops are labelled as trips and trips sequenced into blocks in a format ultimately mimicking the GTFS standard quite closely. This produces what I term a retrospective GTFS dataset, where each trip in the “schedule” actually happened at a particular date and time and every day is different - not because the schedule differed, but because that is the way things were observed to have happened.

While such a dataset might be used to account for past operational characteristics at some particular moment, what Chapter 2 proposes is that it can be used to model future transit performance as well with
an assumption that the future will be essentially like the past, or at least more like the past than like
the schedule. While this is obviously not true in the minute details, the broad contours of accessibility,
the statistical distributions of travel times, are expected to be consistent from one month or year to the
next. Conceptually the difference between this and the schedule-based model is similar to the distinction
in automotive accessibility models between using “free-flow” travel times versus estimates of congestion
based on reliable, cyclic patterns in travel behaviour. Such historically informed congested models are
increasingly being used in automotive trip-planning applications as well as studies of accessibility (e.g.
Moya-Gómez and García-Palomares, 2015; Pritchard, Stepniak, and Geurs, 2019).

Chapter 3 puts this data to an empirical test, taking week-long retrospective observations of four ma-
jor North American transit agencies and measuring the degree to which the retrospective data conforms
to the standard schedule-based model in terms of measured accessibility outcomes. The results show sub-
stantial differences both in time and space and even in aggregate measures across both dimensions. It is
shown for example that errors in origin-based accessibility even when averaged over one-hour periods can
vary widely between locations. Origins with lower scheduled accessibility tend to produce less reliable es-
timates with more variability from hour to hour in real accessibility, while higher accessibility zones seem
to converge on an estimate 5-15% lower than the schedule predicts. Such over- and under-predictions
exhibit strong spatial patterns which should be of concern especially to those using accessibility metrics
in statistical models or measures of spatial disparity. Momentary measures of accessibility are briefly
discussed and found to be weakly related to momentary changes in real accessibility.

1.3.2 Route-choice Without Omniscience

Chapter 4 takes on the legacy of trip-planning applications as the foundation of accessibility studies. A
trip-planning application is designed to produce an optimal route or set of optimal alternatives per some
criteria which can be minimized such as travel times and/or number of transfers (See Delling, Pajor,
and Werneck, 2014). While in theory we may surmise that real passengers do try to optimize their route
selection based on some quantifiable criteria, the very use we have for trip planning applications points
out that this is a difficult task in practice. Even if one were to suppose that real passengers religiously
consulted - and obeyed - well-tuned route-selection algorithms before departing, the stochastic nature of
real transit services would prevent even these passengers from achieving optimal travel times for every
departure.

It is possible in theory to know every detail of a schedule before departing, making an optimal
path a somewhat defensible basis for route selection when everything is assumed to go according to
schedule. Once the validity of schedule data is questioned however (as in Chapters 2 & 3) it needs to be
acknowledged that no reasonable model of travel behaviour can allow a traveller to optimize their choice
with much accuracy at any moment on the basis of what can be known only probabilistically. Fortunately
the problem of modelling realistic passenger route-choice has been worked out quite thoroughly under
the guise of “route assignment” for transport demand modelling (Gentile et al., 2016; Y. Liu, Bunker,
and Ferreira, 2010; Q. Fu, R. Liu, and Hess, 2012). Starting from this empirically-oriented literature on
passenger route-choice behaviour, Chapter 4 takes real passenger behaviour as the basis of two heuristic
route selection strategies. These allow a model passenger to optimize their route selection on the basis
of only imperfect or probabilistic information, resulting in sub-optimal itinerary selections for many
departures with longer average travel times as a result.

These strategies are applied to a model of the Toronto Transit Commission (TTC) based on a
week of retrospective GTFS data and it is shown that the implications for estimated travel times are substantial enough to warrant further development of heuristic models of route choice in accessibility studies. While this limitation of current models had been noted before, no one had yet measured the impact this assumption has on travel time estimates or to my knowledge suggested or implemented a better alternative.
Chapter 2

Constructing a Routable Retrospective Transit Timetable from a Real-time Vehicle Location Feed and GTFS

2.1 Abstract

We describe a method for retroactively improving the accuracy of a General Transit Feed Specification (GTFS) package by using a real-time vehicle location dataset provided by the transit agency. Once modified, the GTFS package contains the observed rather than the scheduled transit operations and can be used in research assessing network performance, reliability and accessibility. We offer a case study using data from the Toronto Transit Commission and find that substantial aggregate accessibility differences exist between scheduled and observed services. This 'error' in the scheduled GTFS data may have implications for many types of measurements commonly derived from GTFS data.

2.2 Introduction

Over the last ten years, the General Transit Feed Specification (GTFS) has emerged as an industry standard for publishing data about transit operations. Data in this format has issued from more than a thousand transit agencies around the world and that data has been incorporated into just as many user-facing routing applications. GTFS data defines transit schedule information in a format that is essentially a routable spatiotemporal network graph with stops as nodes, scheduled travel between stops as edges, and estimated travel times as the cost. This not only allows people to find their way from A to B, but due to the open nature of the standard, has allowed researchers to ask interesting questions and have them answered with a degree of accuracy and scope that would have been impossible before GTFS. Such questions, still very much under active research, include measures of disparities in service provision (Farber, Ritter, and L. Fu, 2016; Fransen et al., 2015), temporal variability (Farber, Morang, and Widener, 2014), the role of relative travel times and costs in mode choice (Owen and Levinson,
2015; Salonen and Toivonen, 2013), the degree of accessibility offered by competing transit development plans (Farber and Grandez, 2016), and many others. Yet such research using GTFS is subject to a serious criticism: it is based entirely on schedules, which are expectations about services, rather than observations of them. It is common knowledge that transit does not run precisely as scheduled, and that it often differs substantially from the schedule. Occasionally there are major unscheduled disruptions due to severe congestion, vehicle breakdowns, or signal malfunctions. These disruptions are a fact of life for most transit users and well acknowledged by transit agencies themselves.

One way that agencies have acknowledged service delays and disruptions is to issue live updates to their transit schedules. While for some smaller agencies, these take the form of hastily posted signs and twitter notifications, for many larger operations, the effort to update their customers has become perpetual. We refer here to the “real-time” data sources that contain constantly updated arrival predictions as well as live vehicle location reports based on Geographic Positioning Systems (GPS). These algorithmically produced updates have proven quite useful to many transit users (Watkins et al., 2011; Tang and Thakuriah, 2012; Brakewood, Macfurlane, and Watkins, 2015), and the automatic vehicle location (AVL) systems that enable them have proven useful to transit researchers who have primarily used them to assess reliability at the level of stops or lines (e.g. Tribone et al., 2016; El-Geneidy, Horning, and Krizek, 2011) or to propose ways for transit operators to improve reliability or prediction accuracy. Yet transportation researchers have not yet to our knowledge made use of such data to answer research questions that involve routing across the transit network, a type of query enabled by GTFS data.

In this chapter, we propose a novel way of making this “real-time” data available to answer precisely the same sorts of research questions that people have already been asking of and answering with GTFS. We do this by using it to update an existing, schedule-based GTFS package with observed trips and arrival times, yielding a GTFS package based on observation rather than schedule. Of course, such data cannot be used directly for routing actual passengers since the events it describes will already have transpired. Yet, most research questions currently answered by GTFS data might be better directed not at a schedule but at a measure of average performance which could be derived from past events, or at the events of some particular day or time in the past.

In the remainder of this chapter we will describe our method for creating this retrospective GTFS package from GTFS and real-time data. We will then undertake a brief case study of the Toronto Transit Commission to illustrate the potential utility of this approach. Because real-time data is currently not well standardized, we will have to describe our algorithm to some extent as it was designed around a particular standard - the data available from the NextBus company’s API (NextBus, 2016) which is used in Toronto and many other cities. The code we developed is available online\(^1\), and we encourage others to use and contribute to the project. Work to extend the code to accommodate other real-time data standards is underway, and we will try to describe the project here with a degree of abstraction sufficient to allow the reader to see how the technique can be applied to any real-time data standard.

### 2.3 Required Data

Our method relies upon three datasets:

1. A current GTFS package, which will be used primarily to provide the locations of stops and stations.

---

\(^1\)https://github.com/SAUSy-Lab
2. A real-time feed of vehicle locations which are preferably associated with some route information.

3. A routable road and/or rail network dataset including all ways used by transit vehicles.

The method we have developed is based on the information provided by a basic GTFS package, and the real-time information provided by the API delivered by the NextBus company, which currently serves data for about 50 agencies around the world (NextBus, 2016). Other real-time data formats exist, and real-time data has not yet standardized to the same degree as schedule data; some agencies provide different information in their feeds. We believe our algorithm should be generalizable to most real-time data feeds, so long as they provide accurate vehicle locations and update them at a sufficient frequency.

2.4 Algorithm

2.4.1 Outline

The basic ontological units of a GTFS package are routes, trips, blocks, and stops. Routes are sets of typical service patterns grouped under a single name, usually a number. Trips are particular occasions when a vehicle goes from one end of a route to the other, serving an ordered set of stops. Blocks are sets of trips in the order of their performance, operated continuously by the same vehicle. Stops are the locations where passengers can access a trip.

Real-time data often does not describe the transit system in the same terms as does GTFS; it is not designed to provide a routable network, but to give updates on particular vehicles and lines. The task of our algorithm then must be to translate the description provided by the real-time data into the terms/units used by GTFS. In the real-time data provided by NextBus for example, there is no explicit concept of trips or blocks and these must be inferred. The focus of the NextBus data is on vehicles and stops.

Our basic approach is to monitor vehicles in real-time as their locations are updated, keeping track of where they are and when. When we observe that a vehicle seems to have completed a trip, we process the data associated with that trip and begin building up a new trip for the vehicle if it is still in service. Finished trips are associated with stops from the schedule data and stop times are derived from the timestamps of vehicles passing those stops. Next, trips are assigned to blocks and routes and the data is stored in a database. At this point, it is straightforward to extract the data in the text-based GTFS format. Much of the hard work of the algorithm is in error handling and data cleaning as reported vehicle locations are always a bit fuzzy and occasionally spurious.

2.4.2 Collecting and storing the data

We developed a program in Python to collect real-time data from the NextBus API, which is used by the Toronto Transit Commission. The NextBus API is a publicly available web service designed to serve real-time transit data primarily to mobile phone or web applications. One of its functions is to report the latest locations for all operating vehicles in the fleet, the idea being that these will be displayed to users on a mobile web map. For each vehicle, the API returns information on the vehicle’s ID, route, heading, last known location, and the time it reported that location to the server. Vehicles update their location every 20 seconds on average.

\[^2\) Passengers can often stay on the vehicle as it transitions between trips.\]
Our program requests all updated vehicle locations every ten seconds to ensure that all new data is collected and then stores the locations in a PostGIS database along with a timestamp and the other associated information.

2.4.3 Delimiting Trips and Blocks

We assume that each vehicle reported to be in service is operating a trip and that that trip belongs to a schedule block. Each vehicle under observation must be assigned a unique trip ID and a unique block ID. A block ends only when the vehicle goes out of service, defined as failing to report a location for three or more minutes, though some other threshold could be used. The NextBus API did not explicitly report when a vehicle was going out of service. The end of a schedule block implies the end of any ongoing trip, but a trip may also end in a number of other conditions. For the data available to us from NextBus, we defined these as

1. The vehicle reporting a different route.
2. The vehicle reporting a different headsign.

Headsigns typically indicate the direction of travel, i.e. “501 E” and a change to “501 W” implies that the terminal station has been reached and the return trip begun in the opposite direction. With the NextBus data we were fortunate that the headsign and route changes were a good indicator of trip endings. Other real-time formats such as GTFS-realtime could use simpler techniques for determining when trips are ending. For pure historical GPS data, it may be necessary to determine when a vehicle turns back on its route to go the other way.

Once a trip has ended, all information associated with that trip is pulled from the database and processed, as the rest of this section will describe. If the vehicle is still in service, a new trip is started for that vehicle, and new location reports begin to accumulate in the database associated with the new trip ID.

2.4.4 Spatial Matching and Positional Error Handling

Visual inspection of the data provided by the NextBus API showed normal, expected levels of GPS position error for most points, but also some extreme errors associated with trips either starting or ending in a bus garage, where vehicles reported positions many kilometres from the last reported location before going out of service. The extreme errors were easily dealt with by measuring the speed (distance/time) between location reports. Segments over 120 km/h indicated obvious errors and the offending points were removed accordingly. As these were almost always associated with vehicles entering or leaving service, this was not a major problem.

Ordinary GPS measurement error did present a more serious issue when coupled with the relatively low sampling rate of the data (about every 20 seconds). There were often times when no location would be reported near a stop as the vehicle must have passed that stop quickly. This was particularly problematic where the stop was positioned just before or after a turn in the route, as can be seen in Figure 2.1.

To improve the spatial resolution of the data throughout the duration of the observed trip, and to catch any substantial positional errors that remained, we matched the sampled locations of the trip to a plausible route on the network. Map-matching is a common task in modern on-board navigation systems
GPS traces not within 20m of stop

Figure 2.1: Map showing GPS tracks overlaid on street centerlines. Positional error and low sampling rates mean the GPS tracks may be poorly related to the actual path taken by the vehicle.

Chapter 2. Retrospective GTFS

and the problem has been well researched (e.g. Newson and Krumm, 2009; Lou et al., 2009; Krumm, Horvitz, and Letchner, 2007).

We matched our trips to a combination of roads and streetcar tracks with detailed data from OpenStreetMap using the Open Source Routing Machine (OSRM), an open-source routing software. An example of the matched result compared to an original GPS input can be seen in Figure 2.2.

OSRM accepts a series of points and (optionally) timestamps and positional error estimates which it uses to match the route to the provided network. It also allows for custom vehicle profiles; we developed one for transit vehicles which gave us the ability to match to streetcar tracks and agency-managed service roads, as well as to adjust expectations about realistic expected travel speeds. For each returned match, OSRM also estimates the probability that the returned route is the correct one based on a number of factors using a built-in naive Bayes classifier. Matches were stratified according to their assigned match probability and overlaid on a base map; visual inspection indicated that these match confidence estimates were generally very effective at picking out erroneous matchings produced by positional or network data errors.

At first, as many as 10% of our trips were able to find only very poor matches \( (P < 0.5) \). Many of these were attempting to pass through intersections tagged with general turn restrictions which applied to private vehicles but not to transit vehicles. Once the various exceptions for public transit vehicles were manually added to the OpenStreetMap data however the error rate (again, \( P < 0.5 \)) fell to around 1.5% of all trips. These badly mismatched trips appear to be mostly the result of GPS noise in urban

---

3http://www.openstreetmap.org
4http://project-osrm.org/
canyons, and for the purpose of this analysis such trips were discarded. Further refinement of either the routing profile or the street network dataset could almost certainly increase the match rate further, though we found the 1.5% error rate acceptable for the purpose of demonstrating the technique here and do not see reason to suspect that it substantially influenced the case study which follows. Removed trips were not overly concentrated on any one line or area of the city\textsuperscript{5} though the fact that not all data could be used is a limitation of this research.

\subsection*{2.4.5 Determining Stop Times}

Once we have a trip matched to the network, with timestamps associated with matched vehicle locations on that route, the next step is to determine which stops were passed and estimate their times of arrival and departure. From the trips’ reported heading and route ID, we can find a plausible set of stops from the schedule-based GTFS package. This is especially easy if the metadata associated with vehicles in the real-time data provides trip IDs that match those in the schedule-based GTFS, as would be the case for the GTFS-Realtime standard developed by Google. In our case, the set of stops being serviced came from the NextBus API’s routeConfig command\textsuperscript{6}, though these data were identical to that provided in the GTFS package. From these stops, only those within 20 meters of the matched route of the vehicle are used and these are then matched to the nearest point on that route. This is an important step, as many vehicles may not operate their full route, either by taking a detour around a blockage or by

\textsuperscript{5}Clusters of poorly matched trips were used to identify problematic areas in the street network data and such clusters could generally be resolved with an edit to OpenStreetMap. This process was repeated until there did not appear to be any strong pattern to the removed trips.

\textsuperscript{6}https://www.nextbus.com/xmlFeedDocs/NextBusXMLFeed.pdf
making a short-turn to reduce bunching (Nguyen, Diab, and Shalaby, 2018). Thus if a vehicle made a short turn stops in the portion of the route not serviced would not be assigned times.

The time for the stop can then be estimated by linear interpolation from the surrounding vehicle reports (Figure 2.3). Ideally, it should be possible also to estimate dwell times if there are two or more location reports in the vicinity of the stop. These would then be encoded as different arrival and departure times for the stop. In practice, we found this difficult to implement; for most stops as long as the departure time is more or less correct the dwell time would make little difference to any routing on the network. Transfer stations with significant actual dwell time are an exception to this and further work is needed to accurately represent arrival and departure times in such cases.

Figure 2.3: Diagram showing interpolation method

2.4.6 Constructing the Retrospective GTFS Package

GTFS is a text-based format with a specified table structure\(^7\) and comma-delimited files. Up to this point, we have not specified how data should be stored as it is processed. In our implementation, all data is stored in a PostGIS database, where some of the spatial operations are executed. This database is structured in essentially the same way as GTFS data with a table each for trips, stop times, routes, stops, and service dates. Since our data is essentially already in a tabular format, we need only select all required fields and format them according to the GTFS documentation. A distinct service ID is generated for each day of observed data. All trips with more than two observed stops are included.

2.5 Accessibility Case Study

We now demonstrate the potential utility of the retrospective GTFS by comparing a simple measure of average accessibility to jobs in Toronto between scheduled and retrospective datasets for the same period. The Toronto Transit Commission (TTC) operates four rapid transit lines as well as a grid network of high frequency surface bus and streetcar routes. Employment in Toronto is concentrated

\(^7\)https://developers.google.com/transit/gtfs/reference/
in the Central Business District (CBD) and there is high residential density near the CBD, with lower
density residential neighbourhoods in the rest of the city. Public transit mode share in Toronto for
commuting trips is 36.55% (Statistics Canada, 2011).

Accessibility to jobs via transit is a measure that has recently been computed using scheduled GTFS
packages (Owen and Levinson, 2015; El-Geneidy, Buliung, et al., 2016; Farber and L. Fu, 2017). For
our case study, we computed travel times from all census dissemination areas (DAs) in Toronto to
all traffic analysis zones (TAZs). This was done using OpenTripPlanner\(^8\), an open-source multimodal
trip-planning application, which computed centroid-to-centroid shortest-path travel times inclusive of
walking, waiting, in-vehicle, and transfer times. From each DA we calculated the total number of jobs
(aggregated to TAZs) accessible within 45 minutes of travel time, for every minute in the morning rush-
hour period (7am-9am), over two days of both observed and scheduled data. This resulted in a set of
time-varying accessibility scores for each DA, which we then averaged. The difference between the average
accessibility between the retrospective and scheduled datasets is shown in Figure 2.4. Unfortunately,
real-time data were not available in a compatible format for the TTCs four rapid transit lines (black
lines in Figure 2.4). To include these important services in the analysis we had to copy the scheduled
GTFS data for these routes directly into the retrospective GTFS package.

\[100 \times \left( \frac{A_{\text{real}} - A_{\text{sched}}}{A_{\text{sched}}} \right)\]

Figure 2.4: Relative (percent change) differences between scheduled and real-time accessibility scores.

Negative values in Figure 2.4 (red) denote locations where accessibility scores using the real-time
network are lower than those using the scheduled network, and indicate where observed service culmi-
nated in lower than scheduled levels of accessibility to jobs. Positive values (blue) denote the opposite;
these are areas where observed accessibility levels are higher than those obtained by using the scheduled
network.

It may perhaps come as a surprise to some readers that actual operations could result in greater
average local accessibility than strict adherence to the schedule. This is likely due to conservative
schedules, expecting delays that were not actually encountered (Wessel and Widener, 2016). It is also
worth noting that in Toronto’s case, many lines are not actually operated with the goal of adhering to
a published schedule; rather the operational goal is to maintain adequate headways between vehicles,
especially during peak hours. In this case, the speed of vehicles across the route is not explicitly defined
and drivers will likely go as fast as traffic permits while still maintaining space from the leading vehicle.
Such operational practices (headway maintenance versus schedule adherence) can be indicated in GTFS

\(^8\)http://www.opentripplanner.org/
schedules\textsuperscript{9}, yet many agencies do not make use of these optional features.

The two accessibility scores obtain similar values nearer to the city center and along the rapid transit lines, largely because the information on these lines is identical between the two packages. Differences in accessibility are clustered in certain neighbourhoods and along specific routes where transit operations appear to differ substantially from the published schedule. This is further visualized in Figure 2.5 with two comparative minute-by-minute travel time plots, each from a residential neighbourhood (The Beaches and Eglinton West) to Toronto’s Central Business District (CBD).

![Figure 2.5: Scheduled and real-time travel times resulting in accessibility differences.](image)

The left plot is an example of where transit operates with greater headways and results in longer trips compared to the published schedule. Presumably there were less transit vehicles in operation than scheduled during this period or there was severe vehicle bunching. Conversely, the plot on the right is an example of where travel times from the real-time network are on average lower than those from the scheduled network. This could be because the agency overestimated operational delays when scheduling their service.

Interestingly, our case study finds that the aggregate population-weighted accessibility scores for the entire city are almost equal when comparing the two networks. For the 45 minute travel time threshold shown in Figure 2.4, the scheduled dataset had an overall average accessibility score of 268,100 jobs, while the real-time dataset resulted in a score of 269,022 jobs; a difference of only 0.34%. Similar results were found when testing different travel time thresholds.

2.6 Discussion

What is interesting in this case study is not that there is substantial deviation from the schedule, but that this deviation seems to exhibit a spatial pattern. Our measure of cumulative accessibility, even when averaged over four hours of peak period service from two days of operation showed what appears to be systematic spatial variation that affects neighborhoods differently. This means that local measures of accessibility, travel time, etc. derived from GTFS data may not be a good estimate of actual values even when averaged. Alternatively, given the relatively short period of observation, it may be that such patterns vary randomly over time or differ by time of day or by season in ways not revealed by schedule-based data alone.

\textsuperscript{9}https://developers.google.com/transit/gtfs/reference/frequencies-file
It would not be fair to characterize such non-adherence to a GTFS schedule as an operational failure. Transit systems were operating before GTFS data came into being and many of them are still operating in the same way. For example, scheduled transit is often operated by timepoints, usually a small subset of stops at which strict schedule adherence is desirable. Arrival times for stops between these are undefined in practice yet GTFS demands that a precise arrival and departure time be given for these stops. In reality, drivers may make no attempt to adhere to these times and may not even be aware of them. Similarly, we should be careful before interpreting deviations from GTFS schedules with a spatial equity lens. GTFS data is provided by agencies primarily to enable customer-facing trip-planning applications, not to enable researchers to conduct various kinds of accessibility analysis with a high degree of accuracy. The goal of our work is to provide a dataset more suitable for such research applications.

As researchers continue to make use of the wealth of data available via GTFS, they should be aware that such data may differ systematically from actual transit operations. Measures derived from scheduled GTFS data alone may be subject to bias if there is a mismatch between the data and the operational reality.
Chapter 3

On the Accuracy of Schedule-Based GTFS for Measuring Accessibility

3.1 Abstract

In this chapter we assess the accuracy with which General Transit Feed Specification (GTFS) schedule data can be used to measure accessibility by public transit as it varies over space and time. We use archived Automatic Vehicle Location (AVL) data from four North American transit agencies to produce a detailed reconstruction of actual transit vehicle movements over the course of five days in a format that allows for travel time estimation directly comparable to schedule-based GTFS. With travel times estimated on both schedule-based and retrospective networks, we compute and compare a variety of accessibility measures. We find that origin-based accessibility even when averaged over one-hour periods can vary widely between locations. Origins with lower scheduled accessibility tend to produce less reliable estimates with more variability from hour to hour in real accessibility, while higher accessibility zones seem to converge on an estimate 5-15% lower than the schedule predicts. Such over- and under-predictions exhibit strong spatial patterns which should be of concern to those using accessibility metrics in statistical models. Momentary measures of accessibility are briefly discussed and found to be weakly related to momentary changes in real accessibility. These findings bring into question the validity of some recent applications of GTFS data and point the way toward more robust methods for calculating accessibility.

3.2 Introduction

Over the last decade, the General Transit Feed Specification (GTFS) has been established as a standard format for exchanging information on scheduled transit operations. GTFS was designed to enable point-to-point routing applications for transit users, and around a thousand (Zervaas, 2018) transit agencies around the world are now engaged in creating and updating GTFS datasets, primarily for the role these play in enabling popular applications such as Google Transit (Antrim, Barbeau, et al., 2013).

Transport researchers however have been quick to see other applications for this rich new source of data; the same format that allows efficient point-to-point route-finding in user-facing applications, also allows for the efficient estimation of transit travel times between many points all across a city or even a
Chapter 3. The Accuracy of Schedule Data for Accessibility

The large travel time datasets thus derived have been used as the basis for studies of accessibility by transit in a number of different contexts. These applications range from a concern with an equitable distribution of accessibility (e.g. Pereira et al., 2018; Widener, Farber, et al., 2015) to accessibility as an input to mode-share models (e.g. Owen and Levinson, 2015; Boisjoly and El-Geneidy, 2016) to accessibility as a metric for use in transport planning decisions (e.g. Farber and Grandez, 2016; Stewart, 2017b; Farber and L. Fu, 2017).

It is important to remember however that GTFS data is only a schedule, an expectation for future transit service, and not necessarily a realistic description of service as it actually happens. Transit vehicles often run late, get stuck in traffic, depart the station early, and require detours around obstacles. Thus there is potentially a large gap between the accessibility we would expect based only on a schedule and the accessibility that people actually experience in the real world. If this is so then our measures of accessibility may have substantial error, or more likely, are systematically biased. This brings into question the validity and accuracy of numerous studies that make use of GTFS-based travel time calculations and accessibility scores, and may suggest the need either to explicitly acknowledge the limitations of schedule-based analyses or find a more realistic way of measuring accessibility. While there have been numerous studies assessing schedule adherence as a general performance metric (e.g. El-Geneidy, Horning, and Krizek, 2011; Bertini and El-Geneidy, 2003), very few have yet considered the network effects of schedule non-adherence on door-to-door travel time estimates and the aggregate accessibility measures derived from them.

This study takes four North American transit agencies as case studies:

- the Toronto Transit Commission (TTC)
- the Jacksonville Transportation Authority (JTA)
- the Massachusetts Bay Transportation Authority (MBTA)
- the San Francisco Municipal Transportation Agency (SF Muni)

For each agency, we use archived Automatic Vehicle Location (AVL) data to construct a routable ground-truth dataset to which the schedule-based GTFS data can be compared directly. Our assumption is that observations of the transit fleet based on AVL systems, while imperfect, are likely much more accurate representations of what happens on the ground than schedules produced before the events actually take place. These paired datasets, which we refer to as schedule-based and retrospective, are then each used to estimate travel times and accessibility scores across the four agencies at every minute of the morning and evening peak periods over the course of five weekdays. Our analytic strategy is to calculate identical accessibility scores with the schedule-based and retrospective datasets and compare these to understand the nature of any systematic and/or random differences between them.

The first goal of this chapter is to uncover any systematic bias present in schedule data. Schedule-based accessibility measures that are systematically too high or too low could bias a comparison of transit accessibility with other modes or between agencies and regions. A second goal is to find out how much observed levels of accessibility vary around these typical values, and whether that variation exhibits strong spatial patterns. This should give researchers some idea how confident they should be when using schedule data to estimate accessibility levels at any particular time or place.
3.3 Background

3.3.1 Review of GTFS Accessibility Studies

Travel time estimates have long been used as a way of understanding transport accessibility. With the rise of GTFS data, detailed time-based accessibility studies for public transit were made possible at a large scale and there has been a great deal of recent work making use of schedule-based transit accessibility measures. At their core these studies use the cost of travel, usually measured only in total travel time\(^1\), as the basis for assessing the comparative utility of transit from different places and times. It would be impossible to mention every application, but our purpose in this section is to review some of the more common themes and to survey the methods employed.

One of the most common applications of GTFS in the literature is to assess urban accessibility in terms of environmental justice. For example, Farber, Morang, and Widener (2014) use GTFS to look at small-scale temporal variability in transit access to supermarkets in Cincinnati, finding that few of the city’s low income residents have adequate access to healthy food. Pereira et al. (2018) look at the social distribution of change in access to schools and jobs in Rio de Janeiro after a major restructuring of the transit system. Fransen et al. (2015) combine socio-demographic data and a GTFS accessibility metric to look for gaps in service provision in Belgium where households without cars are not collocated with high levels of transit accessibility. And El-Geneidy, Buliung, et al. (2016) compare the spatial distribution of transit accessibility to the relative social advantage of neighborhoods in Toronto.

Another common theme is the use of GTFS to project changes in accessibility from the current state to a future planned state. Researchers can do this by taking a published GTFS package as the status quo and then simply add in some proposed transit line while holding other service more or less constant. For example, Ma and Jan-Knaap (2014) make use of this technique to demonstrate an expected change in access to jobs that would result from the development of a proposed light rail project in Maryland. Farber and Grandez (2016) explore competing transit development schemes in the Greater Toronto area. And J. Lee and H. J. Miller (2018) compare accessibility outcomes before and after a proposed bus rapid transit project in Columbus, Ohio. Conway, Byrd, and Eggermond (2018) discuss some of the practical problems with fabricating GTFS schedule data before a project is built and suggest strategies for estimating accessibility with a variety of alternative schedules for proposed services.

Other applications of GTFS accessibility analysis include assessing historical levels of accessibility over time within a single city (Farber and L. Fu, 2017), estimation of block-group level transit mode share at a metropolitan scale (Owen and Levinson, 2015), and measuring students access to campus as related to activity participation (Allen and Farber, 2018b).

Such applications make use of a wide range of techniques for calculating accessibility. As transit vehicles arrive and depart at discrete times, transit travel times can vary widely from moment to moment (Anderson, Owen, and Levinson, 2012) and there has been some disagreement on how large a temporal sample is necessary to generate a representative travel time and thus accessibility metric. Estimating travel times from GTFS can be a computationally expensive process and many are reluctant to take a larger sample than they feel is necessary for their purposes (Stepniak et al., 2019). At one extreme, researchers have picked just one time as representative (e.g. 8am for the morning commute) and estimated travel times from that moment only (e.g. Ma and Jan-Knaap, 2014; Widener, Farber, et al., 2015).

\(^1\)For examples of other metrics applied to accessibility, see El-Geneidy, Levinson, et al. (2016) and Cui and Levinson (2018).
Boisjoly and El-Geneidy (2016) conducted a comparative analysis of time-sensitive transit accessibility measures, finding them to be generally correlated and appropriate at least for a mode share regression model. They suggest that a single representative time may be enough for such applications. Others have used regular sampling at hourly intervals over a single day (e.g. El-Geneidy, Buliung, et al., 2016), though Owen and Murphy (2018b) point out the dangers of sampling error and of regular sampling in particular which may interact in undesirable ways with repeating service patterns in the schedule data. At another extreme, some have exhaustively calculated travel times at every minute of an entire day to look at fine-grained variability in accessibility over time (e.g. Farber and L. Fu, 2017; Anderson, Owen, and Levinson, 2012). Though Stepniak et al. (2019) suggest that this last approach may require more intensive calculation than is necessary for many applications, it seems at present to be the most common and defensible method of estimating travel times.

When minutely times are calculated, the typical method of estimating transit accessibility is to compute momentary accessibility measures which are averaged together into a single score. This is the approach taken in a series of reports produced by Owen and Murphy (2018a) providing detailed statistics on average levels of transit access to jobs in 49 major US metropolitan areas over the course of three years. Conway, Byrd, and Eggermond (2018) call this approach average instantaneous accessibility as it is the arithmetic average of a series of moment to moment accessibility measures. Many others have used the same approach with minutely averages over time windows of one (e.g. Widener, Minaker, et al., 2017; Allen and Farber, 2018b) or two (Allen and Farber, 2018a; Farber and L. Fu, 2017; Owen and Levinson, 2015) hours.

All of the studies of transit accessibility mentioned so far have however used only schedule-based GTFS data provided by transit agencies, some with minor modifications, while only two publications to date have attempted to measure the accuracy of schedule data for estimating accessibility. Both use an archived AVL dataset to construct a ground-truth measure of accessibility and compare this to a schedule-based alternative. Wessel, Allen, and Farber (2017) describe the technique and put forward a software application to enable the construction of a retrospective GTFS package, applying their method only to a small case study in Toronto. Stewart (2017a) however uses essentially the same method more extensively, calculating measures of accessibility to jobs for low skill workers in London England and to health care centers in Boston, MA. This chapter develops on this work by using the same methods on a broader sample of agencies, allowing the generalization of findings to the general suitability of schedule-based GTFS data for accessibility analysis in a variety of contexts.

### 3.3.2 Reasons to be critical of schedule-based GTFS data

In order to model urban processes at a metropolitan scale, it is often necessary to rely on many simplifying assumptions. In any kind of statistical analysis, the hope must be that such assumptions introduce only statistical noise and fail to cause any systematic bias in the results of the study.

All of the studies mentioned in the previous section implicitly assume that transit systems operate precisely according to their schedule, as defined in an agency’s published GTFS package. No transit agency however can really come close to achieving this, and numerous studies have been done assessing the degree to which certain agencies do or do not adhere to their own published schedules (e.g. Mandelzys and Hellinga, 2010; Y. Lee et al., 2001). Indeed, the widespread provision of real-time transit data is itself tacit acknowledgement that schedules are not enough by themselves for transit users to rely on. For the study of transit accessibility, we might hope that schedule non-adherence is random in such a
Chapter 3. The Accuracy of Schedule Data for Accessibility

The ability of a GTFS schedule to accurately describe transit operations is not only a matter of schedule adherence. While on-time performance issues alone might warrant a look into the accuracy of GTFS schedules for certain applications, there are bigger problems with the format and the way it has been implemented by transit agencies. GTFS provides a model of transit operations but that model may be missing key structures that would make it behave like the actual transit system it represents. GTFS has several optional fields designed to allow the standard to better reflect operational practices, though these are not implemented by many agencies that actually use those practices. For example, the Toronto Transit Commission attempts to operate some high-frequency lines at a consistent headway, yet their GTFS data give a discrete list of trips with fixed departure times at every stop rather than using the optional frequencies table allowed for in the specification (Google, 2019).

Most transit agencies schedule routes by way of “timepoints”, a subset of stops along the route at which they try to be precisely on-schedule. Between timepoints, departure times are undefined in practice yet are typically given to the second in the GTFS schedule, presumably through some kind of interpolation. The standard allows stops which serve as timepoints to be optionally indicated in the data (Google, 2019), but it can be observed that agencies using timepoints often fail to include this data in their public GTFS data.

Many agencies do not model dwell time at all, though this is implicit in the GTFS format, which uses separate fields for arrival and departure times at every stop. In almost every case, the arrival and departure times are identical. This implies that agencies may be applying schedule padding or recovery time arbitrarily along a route rather than at layover points (Wessel and Widener, 2016). The TTC for example gives little or no scheduled layover time at stations, at least according to its GTFS schedule, but does use stations as layover points in reality. Such a practice might significantly effect estimated travel times from schedule-based GTFS as most vehicles may actually arrive at their terminal station earlier than scheduled.

Another issue is that agencies may not even be including major service changes in their GTFS schedules. Again, the TTC, for example, has been observed to not include planned subway closures in its published schedule, though the bridging of these segments by buses predictably causes major delays for travellers.

Still, the biggest theoretical problem is just that agencies do not and cannot adhere precisely to their schedules. Bunching for example is a major problem on frequent routes (Nguyen, Diab, and Shalaby, 2018) and the perfectly even vehicle spacing typically shown in the schedule is rarely observed in reality, as much as operators may try for it. In-vehicle travel times too can vary widely due to traffic, passenger crowding issues, traffic light timing and so on (Wessel, 2015). This inherent variability is acknowledged everywhere but in the schedule which typically repeats like clockwork. Indeed, how to appropriately incorporate temporal variability into the GTFS standard would be an interesting study in its own right. At present, variability is unaccounted for by the GTFS standard.

Accessibility researchers may hope that these and other issues with the accuracy of GTFS data will cancel each other out or have only minor impacts on aggregate travel time and accessibility measures. However it seems just as likely that these issues could compound one on another to produce large net

---

2The Southwest Ohio Regional Transit Authority for example can clearly be seen operating with timepoints, often making vehicles wait for several minutes before passing one, while their GTFS data includes no information on this. The method they use to estimate times between timepoints is not provided to users of the data.
differences; without some empirical measurement, we are left to speculate.

### 3.3.3 Implications for Research

There would be substantial implications for research on transit accessibility if it were determined that schedule-based measures were generally not adequately representative of the degree of accessibility that people actually experience. For example, any systematic global bias toward higher accessibility scores would problematize the comparison of accessibility between transit and other transport modes (e.g. walking or cycling), making transit appear relatively more useful than it actually is. This could be an issue in studies looking at regional mode share, or attempting to use estimated travel costs in a trip distribution model. This would be a concern especially for planning applications forecasting net changes in accessibility that would result from policy decisions.

Any consistent pattern in the location of over- or under-estimated accessibility would have implications for the study of accessibility as a spatial phenomenon. Regional studies of social equity in the distribution of transit services for example rely entirely on the spatial distribution of accessibility being accurately measured. And any research attempting to use such spatial accessibility scores as inputs to a statistical model would have to contend with this omitted spatial pattern (see Anselin and Griffith, 1988).

### 3.4 Methods & Data

This research seeks to discover whether schedule-based GTFS accessibility calculations generally over or underestimate real accessibility, in what circumstances, and by how much. The overall strategy is to compare accessibility measurements derived from a published GTFS dataset to those derived from a detailed reconstruction of what actually happened on the ground based on fleet-wide Automatic Vehicle Location (AVL) data. That ground-truth dataset, though it inevitably has its own flaws, is taken as a more accurate representation of reality than the schedule and allows us to measure differences between the schedule and reality, or what we will term error in the schedule-based accessibility metrics. We will explore the magnitude and distribution of this error as it varies over space, over time, and between different types of transit agencies.

#### 3.4.1 The Transit Agencies

We take as case studies four very different transit agencies over the course of a single five day work-week in November 2017: the Toronto Transit Commission (TTC), the San Francisco Municipal Transportation Agency (SF Muni), the Massachusetts Bay Transportation Authority (MBTA), and the Jacksonville Transportation Authority (JTA) in Florida. These agencies were chosen for their size and diversity and because they all had the necessary data available: a current schedule-based GFTS package and AVL data for most of the fleet available through the NextBus API. A map of each agency is provided in Section 3.5.4.

The TTC serves the City of Toronto, Ontario and is the third largest transit agency in North America with 1.7 million passenger trips daily. It is situated in a larger urban region where surrounding municipalities operate their own transit services, and fare integration between these agencies is minimal. The TTC sits at the center of this region and operates a grid of high-frequency services across the city
ranging from grade-separated rapid transit to streetcars and buses operating in mixed traffic (see Figure 3.5). Most service is operated (during the day at least) with the goal of maintaining adequate headways and minimizing major delays, though some smaller routes do attempt to adhere to a regular schedule where service is relatively infrequent.

At the other extreme, the JTA serves roughly 42 thousand daily passenger trips across the entire Jacksonville, Florida metropolitan area. The system is largely radial, with most lines radiating out from a single downtown transit station (see Figure 3.8). All services operate in mixed traffic, and most operate at hourly or half-hourly frequencies. A timed-transfer system dictates that schedule adherence is often of critical importance for passengers making connections between lines.

In between these extremes are the MBTA and the SF Muni. The MBTA operates all public transit services in the entire Boston, Massachusetts metro area, serving about 1.3 million weekday trips. Much of this ridership is concentrated on the rapid transit lines which radiate out from downtown, while about third of all daily trips are served by surface-running buses and trolleys which fill in the network and make connections to subway stations. While the MBTA operates an extensive commuter rail system, we will ignore it in this analysis and focus on the more central parts of the region (Figure 3.6).

The San Francisco (SF) Muni by contrast serves only the dense core city of San Francisco, providing about 600,000 daily trips via buses, trolley buses and trams running mostly in mixed traffic. Like the TTC, the network follows a grid street network, though Market Street provides a diagonal connection to downtown where many services operate in a transit-only right of way. (See Figure 3.7). The SF Muni serves a relatively small but dense and central part of a large region, and surrounding municipalities have their own connecting transit services. Notably, the Bay Area Rapid Transit (BART) is a commuter rail system with many stops in San Francisco, though we do not consider it or other agencies in this analysis.

### 3.4.2 Retrospective GTFS

We implement the method described in Chapter 2 for constructing a routable retrospective GTFS package based on historical vehicle locations; we will briefly outline the method here, but refer readers to that chapter for a more detailed explanation.

We start by scraping a web-API which provides live GPS-based vehicle locations for all vehicles in a transit agency’s fleet. In the case of this study, we use the NextBus API, which covers about forty agencies around North America from which the above agencies were selected. Vehicle locations along with a timestamp, vehicle ID, the route ID, and other attributes were collected and stored. Sequential observations for the same vehicle are then grouped into trips and trips into blocks to mimic the format used in schedule-based GTFS. Spatial precision is increased by map-matching trips to detailed street network data from OpenStreetMap and the matched geometry is in turn intersected with transit stops from the schedule for the route. Arrival times at each stop along the route are interpolated from the timestamps of the vehicle location records. This procedure effectively reproduces a GTFS format but where each trip is unique to a time and date and based on actual GPS observations of a particular vehicle. The code used to produce the retrospective GTFS is available at https://github.com/SAUSy-Lab/retro-gtfs.

For this study, we constructed a one work-week retrospective GTFS network for all four agencies with data collected from Monday November 6th through Friday the 10th of 2017. This week was chosen because it did not have any major holidays, festivals, or exceptional weather events that would have impacted transit service in any of these cities. While a longer period of observation would be desirable,
a five day period across four agencies already was a strain on our computational resources. All agencies had schedules with uniform weekday service, meaning that the data contains five real observations of a standard scheduled level of weekday service.

Each of the four agencies, to the extent possible, was treated in isolation from any surrounding or overlapping transit services. While this may in some cases produce unrealistic absolute estimates of accessibility, it would be a bigger issue for this study to confound the results by including scheduled services from other agencies in both the retrospective and schedule-based GTFS packages. Commuter rail services for example are not included for the TTC, MBTA, or Muni, and any neighboring agencies such as AC Transit in the Bay Area or York Region Transit in Greater Toronto are excluded as well.

Where AVL data was missing entirely for a few key routes within an agency however, it was necessary to copy these routes over from the schedule-based GTFS in order to have a complete network. For the TTC and MBTA it was necessary to copy all rapid transit lines (see Figures 3.5 & 3.6) from the schedule and for Muni, three tourist-oriented cable car lines. The omissions are likely due to differences between the AVL systems available for different vehicle types (ie. bus, rail, trolley). The JTA did not require any substitutions from the schedule. Because of the substitutions though we should expect some degree of confounding for the TTC and MBTA data, and to a very minor degree for the SF Muni. That is, differences from the schedule data will appear smaller than they actually are because some parts of the schedule are used in both datasets.

### 3.4.3 Calculating Travel Times

It has become conventional in many studies of transit accessibility to demonstrate a new method or dataset by showing its application to the measure of residents’ access to jobs in a region. The use of such simplistic measures in applied studies is becoming less common however as researchers complicate the jobs-accessibility problem with notions of competition (Merlin and L. Hu, 2017) or skill-appropriateness (J. Lee and H. J. Miller, 2018). Many other researchers have focused on measuring access to phenomenon with very different spatial distributions such as medical services (Zygo, 2017), grocery stores (Farber, Morang, and Widener, 2014), or schools (Pereira et al., 2018). We hope this work will be relevant to such studies; to that end, instead of weighting our origin and destination points toward a region’s population distribution (as census geometries tend to do) or toward space (as a regular grid would) we wanted to synthesize a set of points weighted to the distribution of transit service and placed along the transit network itself. This has the benefit of preventing many essentially redundant calculations in areas with little or no service, and reduces somewhat the role of time spent walking to and from stops in total travel times.

To achieve this, we randomly selected a subset of stops from the GTFS data for each agency and used an iterative, centroidal Voronoi tessellation\(^3\) to disperse these points along the network. Centroids were weighted by the number of arrivals scheduled at each stop. The result is a semi-stochastic set of points dispersed along the transit network and generally placed within easy walking distance of stops. The resulting sampling distribution for each agency can be seen in the maps in Section 3.5.4. Roughly 300 points per agency was chosen as a balance between coverage and computational feasibility. A handful of points that failed to properly snap to the street network were removed.

Travel time matrices were computed with OpenTripPlanner (version 1.2.0), one of several applications commonly used for this purpose and default settings were used for all calculations. Travel times were

\(^3\)For background on centroidal Voronoi tessellations, see Du, Faber, and Gunzburger (1999).
calculated between all points at every minute of the morning and evening peak periods, defined as 6:00-10:00am and 4:00-7:00pm respectively. Due to an error with the server the Wednesday morning peak period for the JTA was missing data and had to be omitted from the analysis of that agency. For the retrospective data, travel times were calculated for each of the five days, while for the schedule data only one day was required since scheduled service for all agencies was invariant between weekdays. This gives a total of 35 hours \((4 + 3) \times 5\) of estimated minutely travel times for each agency with the exception of the JTA which has 31 hours. Scheduled and retrospective travel times were arranged in 3-dimensional matrices, respectively \(T_{odt}^{sched}\) and \(T_{odt}^{retro}\) where \(o\), \(d\), and \(t\) index origins, destinations and departure times.

### 3.4.4 Measuring Accessibility

In transport planning, accessibility can be defined as a measure of the ease with which some set of destinations distributed in space can be reached. As this is a spatial problem, accessibility is always tied to a particular location. And as accessibility by transit in particular can vary greatly from moment to moment compared to other transport modes, accessibility metrics for public transit have typically been tied to a particular time as well. While many different ways of measuring accessibility have been proposed, we will address only the methods most commonly used in the literature as discussed in Section 3.3.1. Notably, we will not be discussing metrics that explicitly account for competition or reliability (e.g. Merlin and L. Hu, 2017; Conway, Byrd, and Eggermond, 2018) as these do not yet seem to be in common use.

In general, given a travel time matrix \(T_{odt}\), momentary accessibility from one origin to a set of destinations is defined by Equation 3.1

\[
A_{o,t} = \sum_d f(T_{odt}) \cdot W_d
\]  

(3.1)

where \(f(T_{odt})\) is an impedance function that defines how accessible a destination is based on the travel time to that point; and \(W_d\) is a set of weights on the destination points. To enable comparisons between agencies, we set all weights to \(100/n\), standardizing accessibility scores to the range 0-100%, where 100% would indicate that all destinations are fully accessible. Accessibility can also be averaged over time \(t\) yielding \(\bar{A}_o\), a measure of average instantaneous accessibility from origin \(o\) to all other destinations \(d\) (Equation 3.2).

\[
\bar{A}_o = \frac{\sum_d \sum_t f(T_{odt}) \cdot W_d \cdot T^{-1}}{T}
\]  

(3.2)

Instead of averaging accessibility over all times in our large sample however, we break time into one hour windows, each of which is a more realistic representation of what an accessibility researcher might feasibly calculate from schedule data. We call this measure \(\bar{A}_{oh}\), letting \(h\) index hourly bins.

Accessibility researchers have defined the impedance function \(f(T_{odt})\) in many different ways. One of the most common, called cumulative opportunities, is just a binary measure of whether a destination is accessible within a given time threshold \(\theta\) (Equation 3.3).

\[
f(T_{odt}) = \begin{cases} 
1 & \text{if } T_{odt} \leq \theta \\
0 & \text{if } T_{odt} > \theta 
\end{cases}
\]  

(3.3)
Another alternative, the negative exponential function (Equation 3.4) diminishes smoothly as travel time increases and has been shown to more closely reflect the way peoples’ activity patterns actually change as a function of distance. Its parameter, $\beta$, controls how quickly the curve diminishes.

$$f(T_{odt}) = e^{-T_{odt}/\beta}$$

Equation 3.4

A visual comparison of the two functions with typical parameters is given in Figure 3.1.

![Figure 3.1: Negative exponential and cumulative opportunities impedance functions $f(T_{odt})$ with parameters $\theta = 45$ minutes (Equation 3.3) and $\beta = 30$ minutes (Equation 3.4).](image)

3.4.5 Measuring Accuracy

Absolute travel times and levels of accessibility among the four agencies in this study vary widely; in order to make comparison easier we look at the deviation of the retrospective dataset from the schedule in relative terms. We define error, $e$, for any measure of interest $m$ as the percent difference of the retrospective measure from the schedule-based measure (Equation 3.5).

$$e(m) = \left(\frac{m^{retro}}{m^{sched}} - 1\right) \cdot 100$$

Equation 3.5

To allow easier visual comparison of positive and negative percent changes, we use log scales in all plots making doubling and halving of the scheduled values visually equivalent.

For measures of error resulting in multiple values at different times ( $e(T_{odt})$ and $e(\bar{A}_{oh})$ ) we use medians and quantiles to describe the distribution of errors. In practice, such distributions tended to be skewed toward large values and our intent is to minimize the effect of outliers in the analysis as it is conceivable these could be due to missing or erroneous data in the retrospective GTFS package.

3.5 Results

We start the discussion of results in Section 3.5.1 with a look at error in estimated travel times between all points as these are the basis of the accessibility measures that follow. In Section 3.5.2 we explore average agency-level accessibility error and its sensitivity to the selection and parameterization of impedance functions. This allows us to select a reasonable impedance function parameterization for each agency and proceed to Section 3.5.3 where we observe spatial patterns in accessibility error. In Section 3.5.5 we take a close look at moment-to-moment variation in accessibility at a few selected origins.
3.5.1 Travel Times

In order to appreciate the accessibility measures that follow, it is necessary to understand something of the distribution of the travel times used to construct them. Figure 3.2 plots scheduled travel times $T_{sced}$ against $e(T_{odt})$. A range of absolute differences from the scheduled travel times are given in minutes on the right of each plot. Because the $T_{odt}$ matrix contains more than 100 million travel times, a density plot was necessary to avoid over-plotting points.

Figure 3.2: Scheduled travel times between all points ($T_{sced}$, on the X axis) are compared against $e(T_{odt})$ on the Y axis in the form of a density plot with labelled percentiles. Absolute differences between travel times ($T_{sced} - T_{odt}$) are also given in minutes on the right margin.

There is a lot of information in Figure 3.2, but perhaps the first thing to notice is that the magnitude of average travel times is quite different for each agency. The SF Muni is a geographically compact agency, with most cross-network trips taking less than an hour. TTC, the next largest by geographical
extent has most scheduled trips falling short of two hours. The JTA and MBTA by contrast are region-spanning transit agencies and have trips scheduled to take up to four hours during the rush hour periods. Such long trips are likely uncommon in reality, but are a natural result of calculating travel times between all points in such a large area. Reasonably parameterized accessibility functions will largely discount longer trips such as these.

The next thing to note is that the distributions for all four agencies have their densest parts right along the line indicating near-zero difference from the schedule; that is, a solid portion of all trips do appear to be behaving roughly as scheduled, with less than ±15% deviation from scheduled travel times. That being said, there is substantial variability around the expected times for all agencies but especially for the two with regional coverage. We can observe a definite tendency toward longer travel times in the retrospective dataset, though it may surprise some readers to note how symmetrical the distributions are, with almost as many trips appearing to take less time than scheduled. This is to be expected, partly as the result of conservative scheduling which takes account a priori of expected delays which end up not materializing much of the time (Wessel and Widener, 2016). Another possibility is that travel times are consistent on average, but arrival times are offset from the schedule, making precise momentary estimates quite inaccurate.

Differences in travel times between the two datasets for each agency, in percentage terms, are generally greatest for short to mid-length trips, tapering off for longer scheduled trips where greater relative differences equate to greater absolute differences. An interesting illustration of this can be seen in the plot for the JTA where there appear to be bands in the dataset corresponding to 30 and 60 minute absolute differences from on-schedule performance. These are likely related to the JTA’s practice of running many routes with clock-face scheduling, at 30 and 60 minute intervals throughout the day. A missed connection on a trip requiring a transfer would generally require a passenger to wait an additional half-hour or hour for the next vehicle. Note though that the pattern is visible both above and below the 0% line, indicating that the schedule may also incorrectly indicate that some connections are not possible which actually are feasible. A similar pattern is visible to a lesser degree in the data for the MBTA.

As we move on to using these travel time estimates to calculate accessibility, it is important to remember that many of the longer travel times shown in Figure 3.2 (e.g. beyond 2 hours) end up simply rendering certain destinations inaccessible, and that the impedance functions assign the highest values to shorter travel times. Most of the weight of the analysis that follows therefore is focused on the shortest travel times shown in these plots.

### 3.5.2 Parameter Selection and Agency-level Error

To calculate accessibility it is necessary to select parameters for the impedance functions described in Section 3.4.4. Normally, a parameter would be chosen based on theories about or observations of the phenomenon in question. However as our interest in this study is more general the selection is somewhat arbitrary. To account for this we start by exploring a range of parameters to discover the general magnitude of error in averaged place-based accessibility, $e(\bar{A}_{oh})$. Figure 3.3 shows the median and interquartile range of $e(\bar{A}_{oh})$ for both the cumulative and negative exponential functions as parameters $\theta$ and $\beta$ range over 5 minute intervals from 0 to 60 minutes.

The general picture here is that as parameters increase from zero, more and more of the city is deemed accessible. At very small values, weight is assigned mostly to very short trips where a large proportion
Chapter 3. The Accuracy of Schedule Data for Accessibility

Figure 3.3: The median and interquartile range are given for $\epsilon(\bar{A}_{ab})$ as parameters vary between 0 and 60 minutes for the cumulative opportunities and negative exponential functions.

of time is spent walking either directly to the destination or to and from stops. At intermediate values however weight is assigned to trips where time spent waiting for or travelling on transit is the largest proportion of most trips. This is where all of the difference between the two datasets exists, and where we see the largest errors. As parameter values get still higher however, error starts to taper off as a larger portion of destinations are deemed accessible via either dataset.

Figure 3.3 shows that at intermediate parameter values $\epsilon(\bar{A}_{ab})$ can range quite widely but is in general decidedly negative; at the median, we see accessibility scores roughly $0 - 16\%$ lower than scheduled across all agencies. Recall that in the last section we saw generally longer than scheduled travel times, so this difference is to be expected. There are differences between agencies here too, with the TTC and JTA showing lower average errors than either of the other two agencies. Potential reasons for this will be discussed later.

We also observe that across all four agencies the cumulative impedance measure shows more variance and greater average error than the negative exponential. The variability is to be expected as the cumulative function is discontinuous and will therefore generally exhibit sharp differences around threshold travel times. The difference in averages however is likely due to the fact that the cumulative function considers only potential trips up to its travel time threshold. It was shown in Figure 3.2 that such short trips tend to exhibit greater negative deviations from the schedule while longer trips, which are given some weight by the negative exponential impedance function are generally more concordant, at least in relative terms.
Based on these findings, the rest of this analysis will be based on accessibility measures with impedance functions as defined in Table 3.1. These parameters were selected as typical of transit accessibility studies generally, and typical of the range of errors observed for each agency. The SF Muni was given lower parameter values than the three other agencies as a result of its smaller geographic size and thus the quicker change in its error curve exhibited in Figure 3.3.

Table 3.1: Selected parameters for impedance functions defined in Section 3.4.4. Units are minutes.

<table>
<thead>
<tr>
<th>Agency</th>
<th>Cumulative $\theta$</th>
<th>Exponential $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTC</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>JTA</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>MBTA</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>Muni</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

### 3.5.3 Error in Place-based Accessibility

From an agency-level summary of error in the previous section we now move to measures of average error at particular origin locations. Recall that each origin has a measure of average instantaneous accessibility calculated for each hour in the study period. Figure 3.4 plots the median scheduled accessibility against the median $e(\bar{A}_{oh})$ by origin. Color is used to show the relative dispersion (interquartile range) of $e(\bar{A}_{oh})$ on a quartile scale ($\bullet$Q1,$\bullet$Q2,$\bullet$Q3,$\bullet$Q4). We can observe a clear relationship between the three variables: as scheduled accessibility increases, errors, either positive or negative, tend to become smaller and the variability in observed errors over time tends to decrease.

Origins with low average accessibility tend to be peripherally located and served by relatively few transit lines. Accessibility at such locations may even depend entirely on the timely arrival of vehicles on a single line which provides access to the rest of the network. Any disruptions to that line’s schedule adherence can have an outsized impact on error in momentary accessibility, and as performance will vary over time, may also tend to produce greater variability in error. By contrast, origins with high average accessibility tend to be centrally located and/or served by multiple transit lines. If one vehicle is running late another may be running early and total accessibility converges on an average. Note though that if that is the case, it would likely not be the same destinations which would end up being accessible. To illustrate some of these patterns we select two origins from each agency which will be subjected to a more detailed analysis in Section 3.5.5. These origins are highlighted in Figure 3.4 as well as in the maps in Section 3.5.4. Origins were selected for their ability to illustrate typical patterns in the dataset as a whole.\(^4\)

Figure 3.4 also shows that there is a very large range in average $e(\bar{A}_{oh})$ across origins. Again, the relation between average scheduled accessibility and variability seems to hold: the JTA with the lowest scheduled accessibility on average exhibits the highest range among errors, while higher-accessibility agencies like the SF Muni and the TTC have less variation among origins. These two agencies, serving only the centers of their respective regions have relatively frequent service and grid-like networks; most origins in our sample are served by multiple lines. The more radial agencies by contrast, without the redundancy of a grid structure, may tend to produce origins with access to the network depending on a smaller number of lines, thus more susceptible to interruption.

\(^4\)Many particular origins were explored in depth which are not shown here for the sake of space. Descriptive statistics such as the correlation between momentary accessibility scores, or the error in average accessibility were used to guide this exploration to diverse origins within each agency.
Figure 3.4: Median schedule-based $\bar{A}_{oh}$ (x axis) is plotted against Median $\epsilon(\bar{A}_{oh})$ (y axis) for each origin using the negative exponential impedance function. Color gives the relative dispersion of errors for each origin as a quartile classification (*Q1, Q2, Q3, Q4). Selected origins are highlighted in black and comparable results for the cumulative function are provided in Appendix A.

3.5.4 Spatial Patterns

To detect any systematic spatial patterns in $\epsilon(\bar{A}_{oh})$ we compute Moran’s $I$ for the median measure of error at each zone. As origin points were originally placed using a Voronoi tessellation, we construct an adjacency matrix using a Delauney triangulation. This is then thinned to a sphere-of-influence graph (See Michael and Quint, 1999) to eliminate any unreasonably long edges around the margins. As shown in Table 3.2 spatial autocorrelation is found to be strongly positive for all agencies and measures. The spatial distribution of median $\epsilon(\bar{A}_{oh})$ is mapped for all agencies at the end of this section (Figures 3.5, 3.6, 3.7, & 3.8) and visually confirms the findings of strong spatial autocorrelation. Selected points
discussed in Section 3.5.5 are highlighted and labelled.

Table 3.2: Moran’s $I$ of median $e(\bar{A}_{oh})$. All values are statistically significant ($P < 0.001$).

<table>
<thead>
<tr>
<th>Agency</th>
<th>Cumulative</th>
<th>Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTC</td>
<td>0.547</td>
<td>0.603</td>
</tr>
<tr>
<td>MBTA</td>
<td>0.815</td>
<td>0.855</td>
</tr>
<tr>
<td>Muni</td>
<td>0.679</td>
<td>0.681</td>
</tr>
<tr>
<td>JTA</td>
<td>0.692</td>
<td>0.805</td>
</tr>
</tbody>
</table>

As would be expected, it appears that the errors at particular origins are related to the lines that serve them. This is particularly visible in the map for the JTA, Figure 3.8, where definite linear patterns in errors are visible. The MBTA, as the other radially structured regional agency, Figure 3.6, exhibits similar linear patterns but also appears to have broader trends with the north-west portion of the region exhibiting greater average negative deviations in accessibility. Patterns in the TTC (Figure 3.5) and SF Muni (Figure 3.7), the two grid-structured agencies, are somewhat harder to discern. Both have obvious clusters of high and low values, but they do not seem as clearly related to particular lines. The TTC seems to show some consistency along the east-west subway line which could be attributable to the fact that subway departure times were copied directly from the schedule data. This consistency however does not appear as strong for the north-south running lines which were also duplicated from the schedule data.

Two previous studies of network-level GTFS reliability have presented maps of the TTC and MBTA using similar measures of error to that used in this research. Stewart (2017a) gives a map of $e(\bar{A}_o)$ for a central portion of the MBTA network. The spatial distribution of error, while measured to and from different locations, does look similar in places to the measures we produced for the MBTA. As his analysis was conducted using data for October 2016, this may indicate that the observed spatial pattern of error is persistent for the MBTA.

The work of Wessel, Allen, and Farber (2017) presents a brief 4-hour-average snapshot of error in transit access to jobs in Toronto. The map they present however, while covering the whole city of Toronto, does not appear to match the patterns observed in this study, though the magnitude of errors does seem to fall in the same range. It may be that the smaller time sample in that study produced less typical results, or it may be that patterns in spatial error are less stable for the TTC. Another possibility is that jobs are strongly clustered in downtown Toronto and this weighting skewed the accessibility results sufficiently that the maps are not comparable. Without further comparisons with other agencies and time periods, the utility of conjecture is limited.

While it is natural that there should be spatial patterns to this phenomenon the patterns do present a problem for any statistical application of schedule-based accessibility data. Models for example of mode share or social equity that use schedule-based accessibility as an input will have unobserved spatial patterns in their measurement error. Since this pattern would go unobserved by the typical researcher, it cannot be accounted for in the model and may produce inefficient or biased estimates.

---

5 Cf. Stewart, 2017a Figure 8.7 page 178 with our Figure 3.6.
6 Cf. Wessel, Allen, and Farber, 2017 Figure 4 page 95 with our Figure 3.5
Figure 3.5: TTC. 327 total O/D points. Note that the rapid transit lines were derived from schedule-based GTFS data only. Color indicates median $e(\bar{A}_{oh})$ for each zone using the negative exponential impedance function. Note that due to methodological differences, a direct comparison of this map to Chapter 2's Figure 2.4 is not possible. See Section 5.3.2 for further discussion of this methodological inconsistency.
Chapter 3. The Accuracy of Schedule Data for Accessibility

Figure 3.6: MBTA. 294 total O/D points. Note that the rapid transit lines were derived from schedule-based GTFS data only. Color indicates median $e(\bar{A}_{ab})$ for each zone using the negative exponential impedance function.
Figure 3.7: SF Muni. 294 total O/D points. Note that the Bay Area Rapid Transit System (BART) was not included in this analysis and is not shown. Color indicates median $c(A_{oh})$ for each zone using the negative exponential impedance function.
Chapter 3. The Accuracy of Schedule Data for Accessibility

Figure 3.8: JTA. 282 total O/D points. Color indicates median $e(\tilde{A}_{oh})$ for each zone using the negative exponential impedance function.
3.5.5 Momentary measures

To understand some of the phenomenon driving these errors we drill down into changing patterns of momentary accessibility at a selection of typical origins. Figure 3.9 shows the change in cumulative $A_{ot}$ for selected origins during the evening of Friday, November 10th. Scheduled accessibility is shown in blue and retrospective in red.

Figure 3.9: Change in $A_{ot}^{sched}$, $A_{ot}^{retro}$ from selected origins over the 4-7pm period on Friday October 10th 2017. $A_{ot}$ is given as a percentage of all origins accessible using the cumulative function. Origins are identified on the maps in Section 3.5.4 and in Figures 3.4 & A.1. Note the variable range of the Y axis.

Both A and G are typical of accessibility in zones with low-frequency service provided by a single transit line. A is an unusually low-accessibility zone for the TTC, but G is a relatively typical one for the JTA. Both exhibit a regular sawtooth-wave pattern where accessibility climbs as the arrival of a vehicle gets closer in time before falling again when travel times suddenly increase after the vehicle has just departed. For both origins, there is clear alignment between the schedule and retrospective scores, but also substantial misalignment in spots. Origin A appears to have a mostly on-time service, with arrivals occurring within a few minutes of the schedule, however the height and shape of peaks in accessibility appears to vary, perhaps as vehicles move more or less quickly to their destinations. Origin G by contrast shows substantial misalignment with the schedule, with some departures occurring as much as 25 minutes after they appear to have been scheduled. From minute to minute, this causes G to actually exhibit a slight inverse correlation between $A_{ot}^{sched}$ and $A_{ot}^{retro}$.

 Origins B and H by contrast are relatively high accessibility zones for the TTC and JTA, served by multiple lines going in several different directions. As a result, $A_{ot}$ moves up and down much more rapidly as arrivals occur more frequently. Both origins also have at least slightly over average $e(A_{ot})$, meaning that the schedule is less prone to overpredicting accessibility for these origins. Yet, it seems hard to argue based on Figure 3.9 that this is due to stricter schedule adherence. While there is a positive correlation between the two lines for both origins ($r \approx .2$), it rather seems to be the case that $A_{ot}^{retro}$ varies somewhat randomly from $A_{ot}^{sched}$, ie. with a pattern that is not easily discovered. For origin B, we can see that $A_{ot}^{retro}$ is decidedly higher across the three hours shown while H is close at most times...
Origins C, D, E, & F, from the MBTA and SF Muni are all scheduled for moderate relative levels of accessibility, however D and E achieve small to positive measures of median $e(\bar{A}_{oh})$ while C and F both show median levels of $e(\bar{A}_{oh})$ substantially lower than scheduled. Accessibility at origin D can be seen to track $A_{oh}^{sched}$ quite closely with a few deviations, mostly upward, pushing it toward generally higher than scheduled accessibility. Origin C by contrast, also from the MBTA, also appears to largely track the schedule data, though $A_{oh}^{retro}$ misses a few peaks that appear in the schedule.

Origins E and F, both from the SF Muni, seem to show accessibility scores that range much more widely, from moment to moment, likely due to the smaller geographic scale of the agency. Origin E appears to match scheduled levels of accessibility on average by swinging wildly both above and below the mark at any particular moment. Origin F by contrast seems to not quite match the scheduled swings in accessibility, showing generally lower peaks.

Generally speaking, where service is frequent, or provided by multiple lines, the correspondence between $A_{oh}^{retro}$ and $A_{oh}^{sched}$ appears quite weak. Accessibility scores for such origins are generally in the correct range, though the minutely correlation between the two may be close to zero or even negative. Where service is less frequent, there generally appears to be a stronger correspondence between momentary measures, however offsets in the waveforms due to off-schedule departures may produce substantial error even when error is averaged over a one-hour period, at least if wait times are included in the calculations. In such cases it may be more realistic to assume that a passenger has some ability to coordinate their arrival with the schedule or real-time data, minimizing the importance of long initial waits.

3.6 Discussion

This research set out to explore the differences between accessibility measurements commonly derived from schedule-based GTFS data, and accessibility as experienced by actual transit users in a world where late buses, traffic jams and missed transfers are not only possible, but a part of daily life. The big assumption is that our method of using GPS-based AVL data to produce a comparison dataset was able to produce accessibility metrics which better reflect the actual experience of accessibility than the schedule data. We know that there are some instances where this is not the case, where GPS data for example had irreconcilable spatial errors and had to be discarded. We also know that routing on the retrospective dataset introduces a theoretical problem in that it is not always possible for a transit user to know a priori which route choice will produce the fastest trip. If we assume that our data is accurate though, which seems appropriate, there are some interesting findings that transit accessibility researchers can take from this study.

First, common schedule-based measures of accessibility may overestimate net accessibility on average by about 5-15% or more. This finding has implications especially for any comparisons of accessibility between modes, either as input to a mode share model, or as a factor in policy decisions where accessibility outcomes are modelled for multiple alternative infrastructure investment options. Using only schedule-based accessibility measures in such situations may be akin to estimating automotive travel times without accounting for traffic congestion. It is understandable why these global differences in accessibility come about; for example, schedules generally show vehicles operating perfectly consistent headways, however

---

7The topic of passenger schedule coordination with low-frequency services has not been much discussed in the transit accessibility literature to date which typically determines simulated departure times without regard to the scheduled arrival. How to properly treat this consideration is an issue left for future accessibility researchers.
any stochastic variation in actual timing, either faster or slower, for the same number of vehicles will
tend to lengthen average headways and increase average waiting times even if speeds stay the same.
Or in some cases schedules may be optimized to minimize time spent at transfer points, however a late
running vehicle leads to a long wait until the next timed transfer. When we consider that both the
TTC and MBTA had their rapid transit systems copied from the schedule data rather than observed
with AVL systems, it may be reasonable to suspect that real accessibility would be even lower for those
agencies across the board.

Second, there are strong and consistent spatial patterns in origin-level accessibility, even when ac-
cessibility is averaged over one-hour periods. It appears that schedule data does not predict origin-level
accessibility with error that is spatially random, but under-predicts (and occasionally over-predicts) in
consistent ways at particular places. The period of this study is not long enough to see how stable
these patterns are in the long term but in the time we were able to observe there were indeed distinct
patterns in accessibility error. Origins with over- or under-predicted accessibility scores tend to clump
together, exhibiting strong spatial autocorrelation. This is to be expected as this is indeed a spatial
phenomenon - vehicles running early or late, slow or fast relative to the schedule will have an effect on
measured accessibility at every point they pass. Or, as Wessel (2015) suggests, this could be a problem
with the way recovery time is built into scheduled travel times rather than stationary route termini in
GTFS data. Regardless of the cause, this spatial error is of concern to any spatial use of GTFS-based
accessibility measures, whether the interest is trip-generation models or social equity. Especially where
accessibility is used as an input to spatial statistical models, unobserved spatial measurement error could
have a large and unknown impact on coefficients and probability estimates.

Third, we observe that the negative exponential impedance function seems to consistently produce
smaller average errors than the cumulative function. This is likely due to the fact that the negative
exponential function inherently includes more information, from more estimated travel times than the
cumulative function which effectively discards any information past its travel time threshold, \( \theta \). While
the cumulative function is easier to interpret\(^8\) the findings suggest that this ease be balanced against
its sensitivity to slightly varying travel times. Indeed, either of these impedance functions may perform
poorly in this context when compared to more advanced measures recently developed. Conway, Byrd,
and Eggermond (2018) propose measuring a median (or other quantile) travel time over some period
and using this as the basis of accessibility metrics. The results discussed in Section 3.5.5 would seem to
suggest that this might be a good strategy to reduce average error in schedule data, though that would
need to be tested. Other interesting approaches are suggested by Merlin and L. Hu (2017), who discuss
competitive measures of accessibility to scarce resources like jobs. It is not yet known what confidence
intervals might be placed on such measures derived from schedule data alone.

Finally, we have seen that as we go from particular travel time estimates (Section 3.5.1) to measures of
momentary accessibility (Section 3.5.5), to hourly averages (Section 3.5.3), the size of the average relative
difference between schedule and retrospective datasets decreases. This is essentially a reversion to the
mean as each step is an increase in the level of aggregation, averaging together more and more travel
times to create each measure. While we have not addressed measures of momentary accessibility in any
depth, it should be expected that such measures will have larger errors than measures with any degree of
temporal aggregation. This consideration raises serious questions about research which uses accessibility
measures based on a single departure time, as these would likely include a good deal of sampling error in

\(^8\)We use it ourselves in Figure 3.9 for this reason.
their observations. While the problem of sampling error in this context has been discussed elsewhere (e.g. Owen and Murphy, 2018b; Conway, Byrd, and Eggermond, 2018; Stepniak et al., 2019), those accounts have focused on the interactions between selected departure times and scheduled service dynamics and have not addressed the possible gap between scheduled and actual travel times.

In general, this research points to the need for more work toward understanding uncertainty and variability in transit travel. It may not be sufficient to use schedule data alone to see how accessibility across transit networks varies over time except perhaps in very simple cases, or where schedule adherence is known to be exceptional. There needs to be some understanding of how variability in transit travel times are actually experienced as a quasi-stochastic phenomenon and used to inform mode choice, route selection, and itinerary planning. We hope that this research helps move that conversation forward.
Chapter 4

The Effect of Route-choice Strategy on Transit Travel Time Estimates

4.1 Abstract

In studies of accessibility via transit, estimates of travel time typically rely on the calculation of a shortest-path between two points for a given departure time. Such shortest-paths are time-dependent and not always stable from one moment to the next. Given that actual transit passengers necessarily have imperfect information about the system, their route selection strategies are heuristic and cannot be expected to always achieve optimal travel times. Thus an algorithm that returns optimal travel times at all moments will tend to underestimate real travel times all else being equal. While several researchers have noted this as a theoretical issue none have yet measured the extent of the problem. This chapter observes and measures this effect by contrasting two alternative heuristic route-selection strategies to a standard shortest-path calculation made using standard trip-planning software and GTFS schedule data. The Toronto Transit Commission is used as a case study and we model actual transit operations for the agency over the course of a normal week with archived AVL data transformed into a retrospective GTFS dataset. Travel times are estimated using two alternative route-choice assumptions: 1) habitual selection of the itinerary with the best average travel time and 2) dynamic choice of the next-departing route in a predefined choice set. It is shown that most trips present passengers with a complex choice among competing itineraries and that the choice of itinerary at any given moment of departure may entail substantial travel time risk relative to the optimal outcome. In the context of accessibility modelling, where travel times are typically considered as a distribution, the optimal path method is observed in aggregate to underestimate travel time by about 3-4 minutes at the median and 6-7 minutes at the 90th percentile for a typical trip.

4.2 Introduction

Accessibility metrics are widely used in transport planning, where accessibility is defined as a measure of the relative ease of travelling to some set of destinations distributed around a region. Accessibility is concerned not only with where people do go, but also with where they might go and is often thought of as a measure of people’s freedom or ability to engage with opportunities, procure services, and generally
engage with the life of a city or region. Accessibility is typically estimated separately for different modes of transport and this study is concerned with accessibility by public transport, hereafter referred to as transit. Travel by transit generally has relatively low monetary costs, and transit accessibility studies have tended to emphasize travel time as the main or only component of accessibility measures (e.g. Owen and Murphy, 2018a). The estimation of accurate transit travel times is thus critical to an accurate measure of transit accessibility.

While accurate estimates of travel time can in general be derived from various techniques for observing a sample of actual travellers, such techniques are not able to be generalized well to the problem of estimating potential travel times. Rather than observe real passengers, transit accessibility researchers have relied instead on modelling transit systems using schedule data which has become widely available (See e.g. Zervaas, 2018) in a standard format known as the General Transit Feed Specification (GTFS). Such models, among other problems (See Chapter 3), do not realistically simulate passenger route choice. While some transit trips may have only one obvious way of getting from A to B, many others present passengers with several competing alternatives. Given the stochastic nature of transit operations, it is impossible to know a priori which will minimize travel time, and a real passenger is required to somehow make a decision with limited information which will impact their realized travel time. The problem of modelling route choice has been covered extensively in the literature on transit assignment and many models have been proposed (Gentile et al., 2016). Rather than take advantage of this body of work however current GTFS-based models have relied on algorithms which return the minimum possible travel time for any given moment of departure - even where the choice of an optimal itinerary would be extremely complex and require information which is not available to an actual passenger.

This chapter argues that more realistic assumptions about route-choice are needed in transit accessibility models and makes the case for this by quantifying the degree to which shortest-path algorithms produce unrealistically low estimated travel times compared to more realistic alternatives. We take the Toronto Transit Commission (TTC) as a case study and model transit operations with a retrospective GTFS dataset (Wessel, Allen, and Farber, 2017) describing the actual operations of the agency over the course of five weekdays. Using this dataset, we calculate travel times using both a standard shortest-path method and using two alternative heuristic strategies. One strategy is based on an assumption that people are creatures of habit, the other on an assumption that they may vary their choice based on real-time information at the time of departure. By observing the absolute differences in travel times between the shortest path and heuristic estimates we can then quantify the degree to which current models of accessibility may underestimate travel times and thus over-estimate accessibility. We also seek to provide context for the eventual development of better route-selection heuristics by quantifying the empirical choice scenario faced by passengers. To this end we propose an entropy-based measure of route-choice complexity and quantify the travel time risk associated with route choice at any given moment of departure.

Section 4.3 provides context for this chapter and further develops the conceptual problem of route-choice in stochastic transit networks. Section 4.4 lays out the methods and data used, providing a detailed illustration of the two tentative route-selection heuristics. Finally, Section 4.5 presents the results and Section 4.6 summarizes the findings.
4.3 Literature Review

4.3.1 Transit Accessibility

In transport research, accessibility is a relative measure of the ease with which some set of spatially distributed destinations can be reached by a potential traveller. In studies of transit specifically, this “ease” has typically been defined as a function of travel time only, though other cost functions have been developed. For example El-Geneidy, Levinson, et al. (2016) included the monetary fare price to account for the experiences of low income transit users - and Cui and Levinson (2018) propose a general (not specific to transit) measure of accessibility that incorporates external costs of travel such as insurance, pollution, and public subsidy of infrastructure. Nassir et al. (2016) propose a utility-based accessibility function, attempting to measure the subjective experience of using transit. They even consider route choice in this context, though only in as much as multiple route options are theorized to present travellers with more options on which to exercise their preferences.

Despite such efforts, the literature on transit accessibility is dominated by the use of estimated travel time as the only measure of accessibility (e.g. Farber and L. Fu, 2017; Boisjoly and El-Geneidy, 2016; Ma and Jan-Knaap, 2014; Owen and Levinson, 2016; Owen and Murphy, 2018b; J. Lee and H. J. Miller, 2018; Wessel and Farber, 2019a; Conway, Byrd, and Linden, 2017). As the internal monetary costs of using transit are relatively low, and the conversion between time and money inexact at best, this is not likely to change in the near future, nor do we suggest that it should. Travel time is likely to remain the largest impediment to travel that transit users encounter.

4.3.2 Estimating Travel Times

Given the dominance of travel time in the measure of accessibility, the accurate estimate of such times is critical to the study of transit accessibility.

Probably the most accurate way to estimate travel times on transit is to actually observe passengers making trips and see how long they take to reach a destination. In transit systems with smart-cards and distance-based fares, this can be possible if passengers are required to tap their card both when entering and leaving the system (Sun and Xu, 2012). Systems such as these can produce huge volumes of trip time measurements (Jang, 2010) and travel times between stations or stops can often be inferred quite easily by taking the difference between tap-in and tap-off times (H. Lee et al., 2017). When observed over time these systems can be expected to generate distributions of travel times that accurately characterize actual travel between stations. While fare systems like these can be extremely useful for service planning, they are relatively uncommon outside of Asia, except where they are applied only to rapid transit (e.g. WMATA) or commuter rail systems (e.g. GO Transit), but not the connecting bus/streetcar networks. In any case, such data are limited in their potential applicability to accessibility metrics by their reliance on observed travel behaviour. Pairs of stops that are not well frequented may have few or no observations connecting them, especially at particular times of day.

Similarly, several methods have been employed for observing transit passengers using connected mobile devices, either by recording their movements with GPS (Stenmeth et al., 2011) or detecting their presence with Bluetooth sensors (Dunlap et al., 2016). Such datasets would almost certainly produce a smaller sample of observed trips than the smart-card data, and would suffer from the same issue - that trips not commonly made cannot be observed.
By far the most common method of estimating travel times for the purpose of measuring accessibility however is to use an application which models the transit system from standardized schedule-based GTFS data, which has become very widely available in recent years (Zervaas, 2018). Some such applications were designed with passenger navigation assistance in mind, as with OpenTripPlanner (OpenTripPlanner Project, 2019), while others such as R5 (Conway, Byrd, and Linden, 2017) and UrbanAccess (Blanchard and Waddell, 2017) were developed explicitly for the purpose of transit accessibility analysis.

These applications take two different approaches to travel time estimation. OpenTripPlanner and R5, along with ArcMap’s GTFS Network Analyst (Morang, 2019) closely follow the data model defined by the GTFS and integrate it with a pedestrian network e.g. from OpenStreetMap. Each transit vehicle scheduled to pass from one stop to the next is modelled as a directed edge in a large space-time graph comprising transit trips, stops, and walking edges. This graph is then scanned by shortest-path algorithms such as RAPTOR (Delling, Pajor, and Werneck, 2014) or Dijkstra’s Algorithm. By contrast, UrbanAccess takes a faster, more heuristic approach, pre-processing the GTFS data into an essentially atemporal network graph, where travel times and waiting times are summarized as averages and associated with transit edges (Blanchard and Waddell, 2017). This approach, while extremely fast and useful in many circumstances, has substantial limitations such as the inability to consider timed arrivals, headway and speed variability, and variability over time (Conway, Byrd, and Linden, 2017). These two approaches correspond well with the dichotomy in transit assignment between schedule-based and frequency-based models (Gentile et al., 2016). The more detailed (i.e. schedule-based) approach is more common in the literature but has its own drawbacks. For example it produces false precision by assuming that schedule data is accurate (Wessel and Farber, 2019a) and as Conway, Byrd, and Linden (2017) note regarding R5:

The accessibility calculations assume that people have perfect information and choose the optimal path to reach their destination given their departure time.

This latter issue has been noted elsewhere in the transit accessibility literature as well (Morang, 2019; Wessel and Farber, 2019a; Kaplan et al., 2014) but to date has not been addressed directly.

4.3.3 Retrospective GTFS

Some of the most recent work in transit accessibility has addressed the assumption that GTFS schedule data is an accurate representation of transit operations. Wessel, Allen, and Farber (2017) outline a method for constructing a retrospective GTFS dataset based on an historic archive of automatic vehicle location (AVL) data. This retrospective GTFS archive then contains not expected future but rather observed past arrival times, and by virtue of its identical format can be used in the same types of network level analysis to which schedule-based GTFS is applied. A similar technique was used by Stewart (2017a) to produce more realistic travel-time accessibility metrics applied in the context of London, England and Boston, Massachusetts. Wessel and Farber (2019a) generalize the problem and assess the accuracy of schedule-based accessibility in the general case by studying four different North American transit agencies over the course of a week and comparing the differences between retrospective and schedule-based accessibility outcomes. They find that schedule-data over-estimate accessibility by about 5-15% across the board and that there is a great deal of spatial variation with some zones differing very substantially from the expected accessibility score.

However in their use of retrospective GTFS these applications make the same problematic assumption
of “perfect information” raised in the previous section all the more pressing. When shortest path routing is done on a retrospective GTFS package, one not only assumes that passengers have perfect knowledge of a schedule but that they have perfect knowledge of the actual future locations of all transit vehicles, including effects of traffic congestion, breakdowns, short-turns, etc. The shortcomings of this assumption have not yet been directly addressed.

### 4.3.4 Route Choice

Empirical studies of route choice in transit have made clear that actual passengers do not behave in quite the same way as shortest-path algorithms when selecting routes. While passengers have been observed in general to prefer routes that minimize travel time, walk distance, and number of transfers (Jánošíková, Slavík, and Kohání, 2014; Eluru, Chakour, and El-Geneidy, 2012), it has not been demonstrated that they are able to optimize these preferences very effectively from moment to moment. Rather than suppose that passengers have all the necessary information and are rational decision makers (See Gärling, 1998), we emphasize the role of habit formation and strategy development.

The role of habit in travel behaviour has been much discussed (See Gärling and Axhausen, 2003) and it seems reasonable to suppose that passengers making regular trips learn from their experiences and form habits which they perceive as generally optimal regarding mode choice, route choice, etc.\(^1\) While habit in the context of route choice specifically has not been well studied empirically, some evidence does point to its influence. Kim, Corcoran, and Papamanolis (2017) observe passengers in Brisbane, Australia and measure the apparent extent of habit in their route choice behaviour. They find that passengers seem more likely to stick to one of several possible routes when they are more frequent travellers (e.g. commuters), travelling between familiar O/D pairs, and when there are substantial potential time differences between the alternatives.

By contrast Spiess and Florian (1989) emphasize the role of strategy in route choice, suggesting that passengers engage in consistent strategies such as taking the first arriving vehicle from a set of available options at a given stop. While the routing strategy may be consistent the actual path used by the passenger can vary widely based on circumstances. The set of paths potentially derived from such a strategy is called the hyperpath\(^2\). Kurauchi, Schmöcker, Shimamoto, et al. (2014) find evidence of hyperpath-style routing among regular bus commuters in London, England. They speculate that such variability may be due to strategies such as taking the first bus to arrive (See Kurauchi, Schmöcker, Fonzone, et al., 2012) or selecting the fastest from multiple overlapping lines which would reach the destination.

### 4.4 Methods and Data

Our purpose in this research is to measure the size of the effect that the assumption of perfect information may have on travel time estimates. We do this by comparing standard “perfect information” travel time estimates to those based on two alternative assumptions:

1. Passengers select a route which is generally the best and stick with it as a matter of habit

\(^1\)For a more advanced model of learned route-choice behaviour see e.g. Wahba and Shalaby (2011)

\(^2\)See Gentile et al. (2016) for a contextualization of hyperpaths with a transit assignment modelling framework.
2. Passengers use a strategy based on available real-time information, taking the next departing vehicle toward their destination from a set of known alternatives.

In either case the resulting travel time estimates must be equal or higher and thus accessibility equal or lower than if the passenger were supposed to act with perfect information. We wish to find out how large this potential gap is and explore any consistent patterns in its distribution. We hypothesize that there is a substantial effect on travel time and that trips where the passenger would face a larger set of route choices will tend to be most overestimated in the perfect information scenario. We further expect that more route choices will be available on longer trips and trips that run diagonal to the street grid, thus requiring transfers.

4.4.1 Case Study - Data

We take as our case study the Toronto Transit Commission (TTC), which operates a large, grid-like network of high frequency transit lines spanning the City of Toronto, Ontario. It is the experience of the authors that this network often produces difficult and unsatisfying route choices, as many potential trips are served by multiple lines, few of which adhere closely to a published schedule, and many of which exhibit substantial variation in headway and in-vehicle travel time.

To accurately represent this variability we make use of a retrospective GTFS package describing the state of the transit network for the five weekdays of November 6th-10th, 2017. A retrospective GTFS, as outlined by Wessel, Allen, and Farber (2017) is a dataset in GTFS format that describes the exact operations of a transit agency as they were, based on observed transit vehicle locations, rather than as they were scheduled to be. The time period in question was a normal week in Toronto, marked neither by major holidays nor weather events and we assume it is representative of typical levels of service for the TTC. The dataset, and the code used to produce it are made available online.

A limitation of this dataset is that AVL data were not available for the subway system - the four subway lines (See Figure 4.1) are modelled in the retrospective GTFS according to the schedule. This will have the effect of artificially reducing service variability on these lines, though in reality these are likely already the least variable, as they are the only lines the TTC operates in a fully grade-separated right of way.

4.4.2 Measurement Strategy

To calculate an accessibility measure based on travel times, one must define first, a set of trips from origins to destinations (ODs) and second, an impedance function specifying the utility of a destination given the travel time between ODs.

Defining a set of ODs is necessarily application specific, though some researchers have adopted the convention of measuring residential access to jobs when discussing accessibility measures in abstracto (e.g. Allen and Farber, 2018a; Owen and Levinson, 2016). We take a different approach here, adopting a spatially uniform sampling strategy and giving roughly equal weight to all places served by transit in the city. All parts of Toronto are quite dense as the city is at the center of a larger region and this

---

3Unfortunately the TTC does not publish helpful statistics on system-wide reliability but only reports aggregate rates of on-time departures from terminal stations for individual lines. Even with perfect performance on such a metric there would be much opportunity for unreliable service along the length of each route. Without much more systematic analysis of the available AVL data (e.g. http://sausy.ca/2017/12/king-street-pilot-travel-time-variation/) a good quantitative measure of reliability is hard to come by.
strategy has the benefit of remaining agnostic to the distribution of activities one may wish to study. It also does not place undue emphasis on access to the extreme density of jobs in the downtown core.

Similarly, the choice of an impedance function may be guided by application, or for methodological studies by convention. At present however the literature on accessibility is quite active and no consensus has been established on impedance metrics that are appropriate in this context. Some measures like “cumulative opportunities” may be commonly used but are poorly justified by theory or observation. Many other methods are better justified by theory but not yet tested empirically or widely applied (e.g. B. Y. Chen et al., 2017; Zhang et al., 2018). To make the results of this study relevant to ongoing innovation in this area, we will not calculate accessibility as such but rather look directly at the travel times used to construct any such metric. Our hope is that researchers will be able to infer the implications of absolute changes in travel times for their own preferred impedance function. As all reasonable impedance functions share an interest in modelling, with more or less accuracy, reasonable human travel behaviour we can make our observations most relevant to the field by selecting for observation trips which have lengths characteristic of actual travel.

The goal then is to generate a set of trips which is spatially random yet representative of typical trip lengths for the region. This is accomplished by defining a large set of dispersed OD points, weighting the population of potential trips between them to be representative of typical travel behaviour in the city, and taking a weighted random sample from the population.

We define the set of possible ODs as the intersections of all major transit lines in the city. 440 points were manually placed at these locations with some exceptions where points would otherwise be very close together (Figure 4.1). Given that the density of the resulting point pattern exhibits some spatial variation, we further approximate a spatially uniform sample by assigning weights to each point based on a rough approximation of its catchment area. To measure this we constructed a Voronoi tessellation of the points and clipped it to a 1km buffer. The area of each catchment is applied as a weight to the points and weights are assigned to OD pairs as the product of the origin and destination weights.

The City of Toronto is quite large and a simple spatially random sample was skewed toward atypically long trips, as shown in Figure 4.2. To correct for this, we iteratively re-weighted all potential OD pairs in the sample according to the distance between them such that the probability density function (PDF) of weighted trip lengths matched a PDF of actual trips made on the TTC within the City of Toronto. Distance was estimated with Manhattan distance between ODs accounting for the \( \approx 17 \) degree rotation of most of the street network from true north. The empirical distribution of trip lengths was derived from the 2016 Toronto Transportation Tomorrow Survey (TTS, 2019). A weighted random sample of 1,000 OD pairs was taken from the population of all possible OD combinations. Figure 4.2 provides a comparison of the distribution of trip lengths between the sample and the empirical dataset.

---

4The quality of the retrospective GTFS data is poorest downtown due to 1) relatively poor GPS signal around tall buildings in the downtown core and 2) the predominance of the subway in this part of the city, which as we have noted is based only on schedule data.
Figure 4.1: 440 OD points were placed at major transit intersections and transfer points.

Figure 4.2: Distribution of trip lengths in the sample (red) versus the distribution of actual trips based on the Toronto Travel Survey (blue). The unweighted distribution of all possible OD pairs joining the sample points is given in grey.

4.4.3 Itinerary Choice Set Generation

We define an itinerary as a set of sequential walking and transit legs connecting an OD pair. Figure 4.3 provides a simple illustration with three possible itineraries, a, b and c. Itinerary a involves walking up to stop 1, taking routes 1 or 2 to stop 2 and walking to the destination. Itinerary b instead goes down to route 3, taking it between stops 3 and 4, while itinerary c has the user walk the entire route instead of waiting for transit. Walking segments are based on shortest-path network distance.

To generate a set of itineraries we use OpenTripPlanner to calculate shortest paths between ODs on the retrospective GTFS data at every minute of the day. For each departure, OpenTripPlanner returns a sequence of walking and transit legs, where transit legs have a start and end stop id as well as a route id. Many of the sequences thus generated exhibit minor variations in stop location or route selection. We cluster these various observations into distinct itineraries by treating as equivalent any
Figure 4.3: Three possible itineraries (a, b, c) on a simple OD pair. Routes serving the same stop pairs are treated indifferently.

path which has the same sequence of routes, but different stops or the same sequence of stops but different routes. The first condition can occur due to minor variations in the timing of arrivals at adjacent stops and the latter occurs often where two or more lines share a portion of their route. Itineraries are then defined by the specific details of the most frequent constituent observation. This clustering is done to prevent arbitrary distinctions between essentially identical options.

We break the day into four periods corresponding to general changes in weekday service levels for the TTC (Table 4.1). For each service period and OD pair, we define the set of potential itineraries as those which are the fastest alternative at any point in the service period across the five days for which we have data.

Table 4.1: Distinct weekday TTC service periods (Toronto Transit Commission, 2019).

<table>
<thead>
<tr>
<th>Service Periods</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning peak</td>
<td>6am - 9am</td>
</tr>
<tr>
<td>Midday</td>
<td>9am - 3pm</td>
</tr>
<tr>
<td>Afternoon peak</td>
<td>3pm - 7pm</td>
</tr>
<tr>
<td>Early Evening</td>
<td>7pm - 10pm</td>
</tr>
</tbody>
</table>

4.4.4 Measuring Route-choice Complexity

OD pairs which yield larger, more complex choice-sets are expected to produce situations in which it is more difficult for passengers to achieve an optimal travel time. To measure this, we borrow the concept of entropy from the field of information theory. For background on the concept of entropy in information theory see Gray (2011). By treating the optimal itinerary at any given moment as the discrete outcome of a stochastic process, we can see that the problem is similar to that of measuring the information content of a stream of characters. We define entropy, $S$, in Equation 4.1 where $P_i$ is the probability of
Chapter 4. Imperfect Route Choice

4.4.5 Route Selection Strategies

Individuals engage in a wide variety of strategies for selecting transit routes, some of which have been modelled and studied empirically for decades (Gentile et al., 2016; Y. Liu, Bunker, and Ferreira, 2010). It is not our purpose here to model this variety accurately or to innovate on existing methods - only to suggest some easily implemented alternatives to the problematic assumption that passengers can optimize their travel time based on perfect information of the system. We define two alternatives based on an assumption of imperfect information.

The first, habitual route choice, we define as the selection of a single itinerary which minimizes mean travel time over a given time window. While our dataset covers only a week, we make the assumption that the service was typical and that a passenger could have learned by experience from past weeks that were essentially identical in terms of average travel time performance. As travel time distributions tend to be skewed toward long trips (Xue et al., 2011) the strategy of minimizing a mean also tends to give preference to more reliable itineraries.

For the real-time strategy we assume a passenger has perfect information only about the arrival time of vehicles in the first leg of their itinerary and seeks to minimize their initial wait and walking time where waiting and walking are treated equivalently. This could be approximately the case for example if a passenger checked accurate real-time departure estimates before starting their trip. A passenger is assumed to take the itinerary which would let them board a vehicle toward their destination as soon as possible, from the set of all possible itineraries. Where itineraries share a common first leg, and thus a common departure time, the one with the better mean travel time is pursued to the destination. A fuller implementation of a hyperpath strategy involving multiple decision points would be desirable but is beyond the scope of this research.

Finally, we calculate travel times for the perfect information scenario. While these were initially generated with OpenTripPlanner as part of the process of identifying itineraries, we chose to recalculate these for methodological consistency with the two alternatives. We define the perfect information scenario as using only itineraries in the choice set defined for the other alternatives. Unlike the habitual alternative, route choice can vary from moment to moment and unlike the real-time strategy, no itinerary is ever used which would result in a sub-optimal arrival time.

4.4.6 An Applied Example

In this section we illustrate our method by applying it to the first author’s morning commute of about 4.6 kilometers. OpenTripPlanner generated five distinct itineraries for this OD pair during the 6-9am morning peak period over the five days for which we have data. Table 4.2 provides details of each

\[ S = - \sum_i P_i \log_2 P_i \] (4.1)

The use of a base-2 logarithm means that entropy is measured in bits. A value of 0 bits corresponds to 1 possible itinerary only, 1 bit to two equally likely itineraries, 2 bits to four and so on.

---

5Travel times were estimated from the retrospective GTFS data using software which has been made available online at https://github.com/Nate-Wessel/routing. We did not estimate walking times independently but relied on walking distances between transit legs returned by OpenTripPlanner.
itinerary along with mean travel times and values of $P_i$. These correspond to Figure 4.4, which shows minute-by-minute variation in travel times for each itinerary along with the optimal itinerary at any given moment of departure.

Table 4.2: Stats for each itinerary illustrated in Figure 4.4. Entropy for this OD pair is $S = 1.67$ for the AM-peak period. Figure 4.3 indicates how itineraries are to be interpreted.

<table>
<thead>
<tr>
<th>Label</th>
<th>Itinerary</th>
<th>time</th>
<th>$P_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>{w137, s3177, r47, s3184, w57, s14477, r2, s14482, w828}</td>
<td>31.5</td>
<td>0.494</td>
</tr>
<tr>
<td>b</td>
<td>{w137, s3177, r47, s3184, w57, s14477, r2, s14482, w2, s9557, r510, s9560, w344}</td>
<td>33.3</td>
<td>0.338</td>
</tr>
<tr>
<td>c</td>
<td>{w137, s3177, r47, s3180, w54, s9600, r506, s9614, w583}</td>
<td>36.3</td>
<td>0.047</td>
</tr>
<tr>
<td>d</td>
<td>{w137, s3177, r47, s3184, w57, s14477, r2, s14479, w28, s5525, r94, s5533, w217}</td>
<td>37.7</td>
<td>0.107</td>
</tr>
<tr>
<td>e</td>
<td>{w838, s9600, r506, s9614, w583}</td>
<td>38.8</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Figure 4.4: Travel time (including initial wait) for an OD pair varies across five possible itineraries between 6-9AM. The two colored bars across the bottom of each plot indicate the optimal itinerary (top) and the real-time strategy (bottom). The habitual strategy is constant across time and is not shown.

Itineraries $a$, $b$, $c$, & $d$ share a common first leg, namely walking to a nearby stop and catching the
Chapter 4. Imperfect Route Choice

#47 bus northbound. These four options diverge from there, with c transferring to the next line, the #506 streetcar, after only two stops, and a, b, & d proceeding further up to the subway (Line #2). a, b, & d then differ on whether they take the subway all the way to the closest station (a) or reduce the necessary walk by transferring to another line (b & d) from either of two stations. It is interesting to note that itinerary d, which involves a relatively long diversion from the subway to the #94 bus is only optimal in the earliest part of the period, reflecting increasing congestion on the street that line operates on. Itinerary b is not effected by this, as the #510 streetcar operates in a partially separated right of way. Itinerary e by contrast to the previous alternatives disregards the #47 bus, instead walking a few hundred meters of the route up to the #506 streetcar and taking it the rest of the way.

Itinerary a produces the lowest mean travel time of the alternatives at 31.5 minutes, making it the consistent choice of the habitual strategy. As four of the five itineraries share the same first leg, the real-time strategy is only able to choose between a & e in this case; when there is a big gap in the #47’s service, it is able to choose to walk up to the streetcar rather than wait for a bus that is not forthcoming. Once it has opted to wait for the #47 however, the strategy as we have defined it is locked into taking the itinerary with the best mean travel time which in this case is a.

Other OD pairs likely present itineraries where the real-time strategy is able to make more distinct choices, though in this case it plays a limited role. Though not modelled here, it may be the case that many actual passengers make use of real-time information throughout their trip, enabling them to select subsequent legs based on real-time information. For example it is quite easy in practice to choose between itineraries a and b. The only difference between the two is in whether to walk or use transit for the final leg and the station where the choice must be made has a real-time arrival display and other visual cues.

4.5 Results

The main goal of this research is to demonstrate that more realistic route-choice assumptions generate travel time estimates which are substantially higher than those produced by standard shortest-path algorithms. To quantify this, Section 4.5.2 measures the shift toward longer travel time distributions associated with these strategies. To appreciate those results however it is helpful to first understand something of the complexity of the paths which would be required to achieve an optimal travel time for all departures in the first place. Thus Section 4.5.1 first provides measures of the path complexity associated with the sample of OD pairs. Finally, Section 4.5.3 further illustrates the choice situation, giving a passenger’s perspective of the matter by measuring the potential upside travel time risk associated with route choice at any given moment of departure.

4.5.1 Optimal Path Complexity

Only a small portion of OD pairs in the sample produce only one itinerary across an entire service period. A hypothetical passenger on the TTC, intending to minimize their travel time, would thus face at least a couple viable options for most potential trips, whether or not they were strictly aware of these. Relative distributions of itinerary counts are given in Figure 4.5 separately for each of the four daytime service periods.

The distribution of itinerary counts is fairly similar throughout the day, though the midday and PM-peak service have somewhat longer tails. This corresponds to generally longer travel times and
increased congestion relative to the AM-peak and evening periods. A typical (median) trip produces 5 distinct itineraries in a given service period which roughly matches our expectations and corresponds to the example in Section 4.4.6. The long tails of the distributions do come as somewhat of a surprise, with 10% of OD pairs producing 12 or more itineraries.

Such large counts however may perhaps only reflect the inclusion of itineraries which are only the optimal path for as little as a single minute in the departure window. The measure of entropy described in Section 4.4.4 is designed to account for this, factoring in consideration of how much time in the service period each of the set of itineraries is optimal. Figure 4.6 plots entropy per OD pair as a cumulative density function for each service period. Given that an entropy of 2 bits corresponds to a scenario with four equally likely outcomes and an entropy of 3 bits to 8 equally likely outcomes, we can see from the chart that optimal route-choice on the TTC can be extremely complex for some trips. Similar to Figure 4.5 we see generally similar patterns across the service periods, though with somewhat higher values in the midday and pm-peak periods. Note that the median value of $\approx 1.6$ again corresponds well with the example given in Section 4.4.6. It may be important to recall here what was shown in that example: that distinct itineraries can share a large portion of their route and that relatively minor differences in travel time can flip the optimal itinerary rapidly between a few different values. The last leg of a trip for example may be walked or rely on transit if a vehicle happens to be coming.

This high level of path complexity is concordant with the idea of the hyperpath presented by Schmöcker et al. (2009), though of course we question whether a typical passenger is capable of or interested in allowing such complexity to manifest in their actual travel behaviour. While not shown here, preliminary results indicate that a comparable entropy measure derived from a schedule-based GTFS package for the same period produced a similar distribution overall, though with values shifted somewhat toward lower values of $S$. This is to be expected as the schedule data typically contains repeating patterns with little or no random variability in service.

$S$ is fairly stable for OD pairs across the four service periods with Pearson correlation values ranging from $r = 0.87$ (AM-peak vs. midday) to $r = 0.82$ (AM-peak vs. evening). Exploratory regression
analysis was able to discover significant patterns in the distribution of $S$, showing that OD pairs with higher values of $S$ tend to be associated with longer average travel times and/or greater distances between origin and destination. They also tend to be oriented at a diagonal to the orthogonal street/transit grid of the city, thereby being more likely to require transfers as few direct diagonal services exist. OD pairs which would likely involve travel crossing a subway line also produced higher $S$ values, perhaps due to the use of subway stations as major transfer points in some places. Finally, OD pairs that happened to be aligned with well travelled corridors generally had somewhat lower values of $S$, which we speculate may be due to a streamlining or upgrading of service in busy corridors. To avoid burdening the reader with unnecessary detail here, a fuller description of the variables and regression results is offered in Appendix B.

While the \textit{a priori} choice of a travel-time optimizing path on a particular trip may be daunting, it is entirely possible that the difference in travel times between the optimal and heuristic alternatives is actually quite negligible. This is especially true since the alternatives are based on the same set of itineraries and trips and thus different strategies may result in travel times identical to those of the optimal-choice scenario. Sections 4.5.2 and 4.5.3 explore the implications of the alternative route-selection heuristics for estimated travel times, first in terms of the shift in travel time distributions and then in terms of travel time risk associated with route choice at a moment of departure.

### 4.5.2 A Shift Toward Longer Travel Times

Studies of transit accessibility are often based on distributions of travel times, summarizing accessibility over a particular time period such as the morning commute (e.g. Owen and Murphy, 2018a; Allen and Farber, 2018a). To measure the potential overestimate of accessibility attributable to the use of shortest-path based estimates, we compare the distributions of travel times derived from the standard optimal-path model with those derived with the two alternative strategies defined in Section 4.4.5.

Figure 4.7 plots the global distribution of travel times across all sampled OD pairs for the three route-selection strategies. Our sample includes trips taking anywhere from several minutes to two hours or more and travel times are generally slightly longer in the midday and PM-peak periods. As expected
the distributions for both heuristic strategies are seen to have shifted somewhat toward longer travel times relative to the standard measure. While this shift does not look like much visually, it is important to appreciate that Figure 4.7 covers a two hour period in order to show the whole range of travel times, making small relative differences difficult to notice. Across all four periods, median travel times for the habitual and realtime strategies are about 3 or 4 minutes longer respectively while at the 90th percentile they are shifted 6 to 7 minutes or about 8-10% relative to the optimal-path estimates. Basic descriptive statistics of each distribution are given in Table 4.3.

Table 4.3: Descriptive statistics of the travel time distributions shown in Figure 4.7. Values are in minutes.

<table>
<thead>
<tr>
<th>Period</th>
<th>Strategy</th>
<th>Mean</th>
<th>Median</th>
<th>90th %ile</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM-peak</td>
<td>Optimal</td>
<td>37.2</td>
<td>33.5</td>
<td>65.5</td>
</tr>
<tr>
<td></td>
<td>Habit</td>
<td>40.1</td>
<td>36.1</td>
<td>70.8</td>
</tr>
<tr>
<td></td>
<td>Real-time</td>
<td>40.8</td>
<td>36.5</td>
<td>72.7</td>
</tr>
<tr>
<td>Midday</td>
<td>Optimal</td>
<td>38.5</td>
<td>34.7</td>
<td>67.2</td>
</tr>
<tr>
<td></td>
<td>Habit</td>
<td>41.8</td>
<td>37.8</td>
<td>73.0</td>
</tr>
<tr>
<td></td>
<td>Real-time</td>
<td>42.2</td>
<td>38.1</td>
<td>74.3</td>
</tr>
<tr>
<td>PM-peak</td>
<td>Optimal</td>
<td>40.7</td>
<td>37.0</td>
<td>71.1</td>
</tr>
<tr>
<td></td>
<td>Habit</td>
<td>44.3</td>
<td>40.5</td>
<td>77.7</td>
</tr>
<tr>
<td></td>
<td>Real-time</td>
<td>44.9</td>
<td>40.7</td>
<td>79.3</td>
</tr>
<tr>
<td>Evening</td>
<td>Optimal</td>
<td>36.7</td>
<td>33.5</td>
<td>63.3</td>
</tr>
<tr>
<td></td>
<td>Habit</td>
<td>39.6</td>
<td>36.2</td>
<td>68.6</td>
</tr>
<tr>
<td></td>
<td>Real-time</td>
<td>39.8</td>
<td>36.3</td>
<td>69.5</td>
</tr>
</tbody>
</table>

Just as $S$ varied substantially between OD pairs, we should expect to see a good deal of variability in travel time differences when we disaggregate these global statistics into summaries per OD pair and service period. Figure 4.8 shows distributions of relative difference in the median travel time per OD pair.
and time period. From this perspective we can see that some ODs have distributions shifted as much as 12 minutes toward longer travel times, even in the middle of their respective distributions, though more than half of OD pairs have median shifts of less than three minutes relative to the standard optimal-path distributions.

Figure 4.8: Distributions of difference in median travel times for each strategy relative to the standard optimal method. Observations are per OD pair and time period.

The realtime strategy seems to produce slightly longer travel times than the habitual strategy on average, and as in the previous plot it is interesting to note how similar the two heuristic methods are in their degree of difference from the optimal method. One possible explanation is that the real-time strategy as defined here can only distinguish between the first legs of potentially involved itineraries, defaulting after that to the safest itinerary using that first leg. It may be the case that this often just results in the same choice as the habitual strategy. While the relative distributions may look similar however, the two measures are only correlated at $r = 0.56$.

As with $S$ in Section 4.5.1, we find that the heuristic travel time distributions both tend to be shifted further away from the optimal path scenario, in absolute terms, when trips are longer, oriented at an angle to the grid, crossing a subway line and in corridors less frequently travelled. The interaction with these variables is not terribly strong though, only explaining 18-19% of the variation in median travel time differences. Precise details of the regression models used are provided in Appendix B.

### 4.5.3 Route Choice & Travel Time Risk

Another way to think about the difference in travel times attributable to route selection strategy is in terms of risk associated with route choice at a particular moment of departure. That is, when adopting a reasonable route-selection heuristic, what are the odds that there was actually a non-selected alternative that would have been much faster? This relates to the concept of “regret”, which has been applied before in the context of route choice behaviour (Caspar, 2012; Ben-Elia, Ishaq, and Shiftan, 2013). Briefly, the idea is that people seek to make choices such that a non-selected alternative does not perform significantly better than the one chosen. In a sense, this is the inverse of risk aversion, pushing people to make risky choices to avoid losing out on potentially large gains. It also maps loosely onto the concept of “upside risk” in finance. Our interest in this metric in the context of this chapter is that we would
expect OD pairs with large upside risk to encourage passengers to engage in more dynamic strategies in an effort to optimize their travel times. With the data we have we can measure the extent of this phenomenon, though of course we do not attempt to associate our observations with actual passenger behaviour in this chapter.

We find that in aggregate, across all ODs and times, the habitual strategy results in an optimal itinerary choice 59% of the time while the real-time strategy is able to do only slightly better at 60%. Only a small percentage of OD pairs were able to achieve perfect travel time optimality at all times though more than half achieved optimal travel times for more than half of the time. To illustrate some of this variation, Figure 4.9 plots a distribution of the difference between the optimal and heuristic travel times for a random sub-sample of 100 OD pairs. Service periods did not show substantial differences and are aggregated together for this plot.

![Figure 4.9: Distributions of difference in travel time for a given moment of departure. Each line represents one OD pair. The red line indicates the 90th percentile observation.](image)

It seems clear that while a majority of departure times give equal values for the optimal and heuristic strategies, there is a decidedly long tail for many OD pairs where either heuristic strategy results in travel times 10 or even 20 or more minutes longer than was strictly necessary. In reality, some such trips, especially for shorter expected travel times, might induce passengers to change their strategy if possible. This can sometimes be seen in Toronto where passengers, after waiting longer than they would like will hail a cab, start walking to another stop, rent a bike, or otherwise leave their ordinary itinerary and make alternate arrangements. The effect is sometimes contagious, as one person talking with others and departing can trigger nearby passengers to question their willingness to continue waiting. It is important to consider however that the theoretical optimal time at any given moment is not something that passengers are actually privy to. What they can experience is a delay relative to an expected travel or wait time along with some knowledge that alternate paths might have been taken; if they have tried the alternatives, they may also have some knowledge of the likely travel times on these to compare their experience to.

One consistent pattern in Figure 4.9 is that the real-time strategy has slightly more extreme values...
at the tails; this may be the result of impatience regarding the first leg leading to bad choices in regard to total travel times. Recall also that the habitual strategy tends to prefer routes with few large outlying travel times. For both strategies almost all OD pairs in all service periods show increased travel times by the 90th percentile.

### 4.6 Conclusions and Discussion

It has been noted in the literature that transit accessibility models may falsely assume that passengers have perfect knowledge of the transit system and are able to manifest this knowledge in their moment-to-moment route choice. This chapter has spelled out the implications of this assumption and put it to an empirical test for one major transit agency see whether making more realistic assumptions about passenger route choice ends up producing reasonably similar results. The TTC was chosen because its large grid of frequent transit lines and questionable schedule adherence were expected to maximize the difference between the standard model and more realistic heuristic strategies by presenting passengers with complex and dynamic itineraries among which to choose. Section 4.5.1 showed that this expected route-choice complexity was quite real with passengers having multiple distinct itineraries to choose from for almost all OD pairs. Given the median optimal route-choice entropy of about 1.7, it is not surprising that almost half of departures modelled with the alternative route-selection strategies chose an itinerary that would end up being sub-optimal in terms of travel time, relative to the standard shortest-path model.

At a moment-to-moment level, we found that route-choice has serious implications for expected travel times, with a small portion of trips taking much longer than a path-indifferent travel-time estimate would predict. We describe this as the “travel time risk” associated with route choice for an OD pair and find it to be strongly related (0.41 ≤ r² ≤ 0.58) to the entropy of the time-optimal path. Despite this potentially large difference in travel times given a particular moment of departure, we find that when travel times are considered as distributions over a time window, the shift toward longer travel times seems relatively small with the average (median) trip gaining less than 5 minutes at the 90th percentile. Again though, there appears to be a long tail with some OD pairs seeing large shifts in travel time distributions.

We attempted to identify characteristics of OD pairs that were related to greater route-choice entropy and greater shifts in travel time distributions, but were unable to explain more than about 20-25% of the variation with simple spatial measures. In general, longer trips and those likely to require transfers produced greater differences from the standard perfect-information route-choice model. To offer some explanation for the greater part of the variance would require further study.

To our knowledge, this study is the first to look at this phenomenon directly and much work remains to be done. It is important for example to consider that our definition of more realistic route-choice algorithms (Section 4.4.5) is far from a fully developed model of human behaviour. While denying transit passengers omniscience about their ultimate travel time, we still attribute to them too much capacity in some areas and too little in others. For example, the realtime strategy we defined has no ability to change course midway through a trip though it is well known that actual passengers often engage in this behaviour (Spiess and Florian, 1989; Schmöcker et al., 2009). Conversely, the habitual strategy as we defined it seems optimistic, assuming that passengers have an accurate view of the mean travel times on all of their alternatives when it seems unlikely they would even know what all the alternatives
are. Indeed, both strategies seem fanciful when we consider how much information actual passengers are likely have about trips to places they have never been before; this is especially relevant in the context of accessibility studies and needs much more consideration.

How much our findings can be extended to other agencies or to schedule-based data is somewhat unclear. The TTC is a large agency with most routes offering frequent service. Other agencies with different service patterns, for example radial routes and infrequent service, likely have much more constrained route choice for most trips, though perhaps the implications of choosing a sub-optimal route would be more dramatic. It may be useful to look at the problem in a slightly different way by noting that an abundance of alternative routes implies that the network is resilient to sub-optimal route choice by passengers. By choosing to study a more “resilient” agency we may have failed to observe places where route choice has more substantial consequences.

As for the use of schedule-based GTFS data in transit accessibility models, we can only suggest that route choice complexity will appear diminished and that travel times will tend to exhibit less variability overall. Previous research has suggested that schedule-based GTFS data has serious limitations in its ability to accurately model transit services as people actually experience them, producing potentially large overestimates of accessibility and systematic spatial errors (Wessel and Farber, 2019a; Stewart, 2017a). While schedule-based analysis is likely to continue into the near future due to the availability of data, we would suggest that more realistic models are needed both of route-choice behaviour and of on-the-ground transit fleet operations.

This study while not itself conclusive is, we hope, suggestive of directions which could be taken in future research on accessibility. As transit operations are modelled with a higher degree of detail and thus typically a stronger presumption of precision, it becomes necessary to consider how precise it is possible to be while building a model of reality in which human agency and human limitations play an active part.
Chapter 5

Conclusions

This dissertation argues that measures of transit accessibility must rely on models which better approximate how real transit systems operate and how they are used by passengers. Transit cannot be treated as a fully known and predictable phenomenon, running precisely according to schedule except perhaps in a few exceptional cases. Urban transit rarely works like this and people do not typically behave as though they expect it to. I have developed two major implications of this line of thought, first in Chapter 2 putting forth a method for replacing fixed schedule data with detailed descriptions of actual past performance. Chapter 3 showed that this substitution produced fairly dramatic changes in measured accessibility across four diverse agencies. Next, given that perfectly optimal route-choice in this context is impossible, alternative route selection strategies were modelled in Chapter 4 and this too was shown to have a meaningful effect on estimated travel times and thus accessibility.

This chapter outlines some broad conclusions that can be drawn from these findings and makes the case that they have important implications for ongoing work on transit accessibility. Specifically, I discuss the relationship between reliability and accessibility, relating this to applications in transit planning, measures of social equity and recent ways of looking at temporal variation in accessibility. I also address some of the limitations of this work and suggest avenues for the remediation of these deficits in future studies.

5.1 Accessibility as a Function of Reliability

Accessibility and reliability are deeply interrelated concepts for public transit, though both have been defined in many different ways. Perhaps the broadest definition of accessibility is to say that it is the very purpose of any transport system, a measure of the utility of that system for reaching some set of valued destinations from some place in space and time. The way this has actually been instantiated as a quantitative measure is multifarious and is the subject of ongoing development. Reliability can be just as slippery a concept, defined broadly perhaps as the degree to which one can depend on something to behave in an expected way - though for transit it has worn the guise of any number of measures of variation: percent on-time arrivals, number of missed runs, 90th percentile wait times, or any metric of travel time distributions, and so on. The two concepts are related in that one cannot use something to reach a destination if it cannot be trusted to behave predictably. This has been shown empirically in the context of mode choice (Sweet and M. Chen, 2011; Bhat and Sardesai, 2006), ridership (Chakrabarti
and Giuliano, 2015), and route choice (Tilahun and Levinson, 2010) with passengers generally choosing to avoid less reliable alternatives.

Despite a large body of research on measuring, assessing, and modelling transit reliability, there has been astonishingly little interaction between this and research on transit accessibility as such 1. To my knowledge the topic has been addressed directly and in detail only by Stewart (2017a), in a dissertation published the same year as Wessel, Allen, and Farber (2017). In this dissertation Stewart develops a selection of “constrained accessibility metrics” including a constraint of reliability measured in terms of travel time. He uses methods very similar to those described in Chapter 2 to estimate travel times in order to observe reliability-constrained accessibility in two case studies. Hendren et al. (2015) argue convincingly that transit reliability metrics should follow the lead of automotive reliability measures and take a passenger-oriented perspective; rather than using statistics at the level of stops or routes (a management perspective e.g. El-Geneidy, Horning, and Krizek (2011)) we should primarily look at distributions of cross-network travel times as these are the main outcome from a passenger’s perspective. They suggest the use of tap-on tap-off smart card systems as a way of measuring passenger-centric travel times, though of course this leaves many agencies that lack such systems without accurate means for systematically estimating OD-level travel times. Gittens and Shalaby (2015) maintain this focus on passengers’ perspectives and suggest the partitioning of total travel time into waiting and travelling components which are experienced differently by passengers.

This move toward travel time variability as a central measure of reliability meshes nicely with the conceptualization of accessibility via transit as largely based on (variable) travel times. In this context, the time-sensitive, schedule-based model of transit accessibility advanced by Lei and Church (2010) is a huge advance on previous work, allowing researchers, by combining GTFS schedule data with trip-planning applications, to produce detailed depictions of the way travel times are expected to vary depending on the time of departure. Clearly such time-dependant variation in estimated travel time has been deemed important for the study of transit accessibility as a number of studies have been promulgated looking at various ways of summarizing or sampling from such travel time distributions (Stepniak et al., 2019; Conway, Byrd, and Eggermond, 2018; Owen and Murphy, 2018b; Boisjoly and El-Geneidy, 2016). Others have delved directly into the variation, using this as a kind of accessibility metric in its own right (Farber, Morang, and Widener, 2014; Farber, Ritter, and L. Fu, 2016).

This way of assessing variability is a good start, but variation in transit travel times clearly depends on more than just the moment of departure. It depends on what route you decide to take, how lucky or skilled you are in timing your arrival at stops, and how easy transit itself makes these things by behaving more or less predictably. As discussed in Section 4.3.2, the ideal way of measuring this is to observe real passengers making trips with as much detail as possible, and derive from these actual trips a distribution of empirical travel times. The problem with this approach is that it does not extend to the set of all possible trips and that the tools to accomplish this are available to only the few transit agencies having tap-on, tap-off smart card systems implemented system-wide 2.

Whether we think of things in terms of accessibility or reliability, the key outcomes are travel times and and travel time variability. As such, accurate measures of this are critically important. Schedule based measures of accessibility cannot account for unreliable service - a retrospective accessibility measure

---

1It is interesting to note for example that Cats (2011) develops an elaborate transit simulation model including explicitly simulated unreliability and uncertain passenger route choice while only mentioning accessibility once in passing in the whole 290 page dissertation.

2Or involve questionable use of private data collected by other means.
however builds reliability into the model implicitly, in whatever form it may take. This may include for example poorly timed transfers, delayed vehicles, missed runs, erratic headways, and so on. The resulting estimated travel time distributions can be neatly incorporated into the growing literature addressing “reliable accessibility metrics”, though these to date have primarily focused on variability attributable to schedules only (Stewart, 2017a; Conway, Byrd, and Eggermond, 2018).

5.2 Implications for Ongoing Research

5.2.1 Reliability in Transit/Land-use Planning

This research has some potential to change the way accessibility measures are used as an input to transit planning and land-use decisions. Several studies have looked at how the addition of new scheduled services impacts accessibility (Ma and Jan-Knaap, 2014) and have used the accessibility outcomes of alternate scenarios to weigh alternative proposals (Farber and Grandez, 2016). This approach to transit planning has been worked out most extensively by researchers associated with Conveyal (Conway, Byrd, and Linden, 2017; Conway, Byrd, and Eggermond, 2018; Stewart, 2017b; Stewart and Zegras, 2016), a company which offers accessibility-based consulting services to transit agencies. The problem with these approaches however is that at present they are only able to compare services which are presumed to adhere perfectly to a schedule. This research however raises a different kind of question: in an effort to increase accessibility, how might one balance a plan to improve reliability against a plan to operate new service? Given that schedule non-adherence has a major impact on measured accessibility (Chapter 3), it must be assumed that improvements to reliability could have potentially large, measurable impacts as well.

This means that changes to routing or infrastructure could, in theory at least, be more fairly balanced against changes to the reliability of a system. Imagine for example that there are two transit plans competing for the same limited budget. One involves adding service to a line and increasing its frequency while the other would reduce automotive traffic on congested downtown streets without adding additional service. Planners are interested in knowing which alternative increases accessibility more for a given expenditure. In order to model this scenario we would need to explicitly account for reliability in the initial state as well as expected reliability and service levels in the two alternative conditions. We might start by using a retrospective GTFS dataset as described here to measure accessibility in the initial condition. The planned future states then could be compared in a few different ways. We could make iterative changes to the retrospective GTFS data, for example increasing the spacing between vehicles or increasing their speed slightly until an expected reliability outcome was measured. Or we might insert extra runs at particular departure times, slotting these between observed trips and jiggling the departure times a bit until things fit. Alternatively, some parameters of service and reliability could be measured from the initial state and a dataset could be produced artificially by some means which would mimic these real characteristics, such as percent on-time arrivals, mean travel times, etc. By changing the parameters of the model used to produce this dataset, we could model additional service, faster travel, etc.

Such changes would of course need to be more rigorously specified and all the details of such a modelling strategy worked out, but the important point is that without an accurate unreliable starting condition, modelling such a change scenario would seem to be much more difficult. It seems clear in any
case that once reliability enters the equation, it is no longer adequate to consider modelling proposed changes with schedule data only.

5.2.2 Accessibility and Spatial Equity

One of the most common themes in the transit accessibility literature is the uneven distribution of transit services among various population groups (e.g. Pereira et al., 2018; Stewart, 2017a; Farber, Ritter, and L. Fu, 2016; Kahrobaei, 2015). In the United States the analysis of such spatial disparities is even required by law any time substantial changes are proposed for a transport system (Karner, 2018; Kramer and Goldstein, 2015). Analyses such as these are concerned with the fine-grained spatial distribution of the benefits of transit, measured in terms of accessibility. The empirical findings in Chapters 3 and 4 indicate however that measures of accessibility based on schedules and trip-planning route estimates produce substantial spatial errors. This was addressed most directly in Chapter 3 where spatial errors (Equation 3.5) were mapped for each agency and found in every case to be strongly autocorrelated (Table 3.2). While that chapter largely declined to speculate about the direct causes or correlations of these spatial patterns since this was not measured explicitly, I had the opportunity to speak with several people more familiar with the MBTA and SF Muni while presenting the research in Chapter 3 at the Transportation Research Board 2019 conference. They indicated that the mapped patterns of error (Figures 3.6 & 3.7) corresponded to general impressions they had of service reliability in different parts of their cities with less reliable areas showing lower than scheduled accessibility.

Most research on reliability has come out of a civil engineering context and has not had much of the spatial or equity focus of research in planning or geography. It is difficult therefore to substantiate any potential association between reliability concerns and the spatial distributions of low income or minority groups if such associations exist. One way of speculating about the direction of this potential relationship however is to look at some well studied differences between passengers of rail and bus lines in general. It has long been held that bus riders are poorer on average than rail passengers (Lu, 2018; Garrett and Taylor, 1999) due to rail’s ability to attract “choice” riders. A big part of this attractive power is that rail services tend to be faster and more reliable, operating at higher frequencies than nearby bus services and using designated right-of-ways (Ben-Akiva and Morikawa, 2002). There is even a substantial literature looking at the ability of rail transit to gentrify the places it serves possibly displacing low income residents (e.g. Grube-Cavers and Patterson, 2015). If such associations are true, then it is likely that areas with lower reliability will also tend to be lower income neighborhoods. If so, this could indicate that spatial inequity in the provision of transit service may have been generally underestimated by conventional models. More research would be needed to confirm or reject such associations. Verbich and El-Geneidy (2016) for example take a spatial equity approach and emphasize the importance of standardized measures of reliability for inter-agency comparisons, but are unable to actually generate any such measures for their study due to the incompatible metrics reported by different agencies.

5.2.3 Inter-modal and Inter-agency Comparisons

One clear finding of this research is the strong suggestion that previous estimates of accessibility via transit have generally overestimated what was practically achievable in reality. This was an expected outcome and conforms with the theoretical justifications outlined in Chapter 1. Chapter 3 found that
net accessibility was 5-15% lower on average than schedule-based measures would indicate, though this varied quite a bit over time and space. Chapter 4 extended this for the TTC by estimating the degree to which even this measure was a net over-estimate due to overly optimistic route-choice assumptions. By applying the same accessibility functions used for the TTC in Chapter 3 to the distributions of travel times produced in Chapter 4, I estimate that this consideration accounts for an additional \( \approx 5\% \) reduction in overall accessibility. Other considerations not addressed in this research (see Section 5.3) would likely indicate even larger overestimates, especially when less formal transit operations are considered.

If this is so, this could indicate that most studies of accessibility via transit in the last decade have probably over-estimated net accessibility by 10-20% or more on average. This finding is relevant to any high-level comparison of transit accessibility measures to other modes (e.g. Widener, 2017). For instance, if planners wished to measure the relative utility of a transit project to a change in pedestrian infrastructure, they might justify one or the other based on estimated net accessibility outcomes. If their transit measure is based on current standard methods, they should assume that it overestimates accessibility for transit. Of course, this says nothing about the accuracy of methods for assessing accessibility changes for other modes of transport. It is likely the case that current methods also overestimate accessibility by other modes as well.

Further, given the variation in the degree of overestimation between just the four agencies studied in Chapter 3, any inter-agency comparisons are thrown into question as well. For example, Owen and Murphy (2018a) use standard schedule and trip-planner based accessibility methods to produce a report of detailed accessibility statistics for the 49 largest US metropolitan areas. They rank cities according to net transit accessibility\(^3\) though many of the reported differences between cities are easily within a \( \pm 10\% \) range. They also report year-to-year change in accessibility for all metropolitan areas\(^4\), giving these to two decimal places and finding all values to be within a range of +17% to -6%. However given that their study systematically fails to account for reliability and changes in reliability, the absolute values of their results and even the relative rankings are cast in serious doubt. My purpose in giving this example is not to throw shade on the work of Owen and Murphy, but to illustrate that the use of present estimates of accessibility without any attempt to define confidence intervals or quantify measurement error can lead to overconfidence in the accuracy of results. Such studies, if they continue with present methods, must acknowledge their inability to account for reliability and their tendency to overestimate accessibility.

### 5.2.4 Time Geography and Temporal Variation in Accessibility

Numerous studies have looked at temporal dynamics in accessibility, though to date these have been based on travel time variability attributable to schedules only rather than also considering reliability issues (e.g. Widener, Minaker, et al., 2017; Farber, Morang, and Widener, 2014; Farber and L. Fu, 2017; Farber, 2016; Goliszek, 2017). The issues raised in this dissertation significantly problematize this approach, especially to the degree that it deals with small-scale moment-to-moment changes in accessibility. This was illustrated clearly by Figure 3.9, reproduced in part below as Figure 5.1. Minute-to-minute changes in accessibility according to schedule-based measures were shown in Section 3.5.5 to sometimes correspond quite poorly with measures based on retrospective GTFS data. At times, consistently delayed vehicles could even induce negative correlations between schedule-based and retrospective accessibility measures.

\(^3\)Table 3 of Owen and Murphy (2018a).
\(^4\)Table 5 of Owen and Murphy (2018a).
However temporal studies of changes in transit accessibility, while generally concerned with times of departure are not typically addressed toward particular historical moments, such as the evening peak period of Friday, November 10th, 2017 shown in Figure 5.1. Rather, they are concerned with general temporal patterns of variability in accessibility. Given that real accessibility does vary between moments and days however, a measure of general variation in accessibility must also take this into account with a measure of accessibility which is robust to variation. For example, we might want to observe how access to grocery stores changes over the course of a day due to an interest in access to healthy food for people working irregular schedules. It would be insufficient to observe the schedule only, or to observe one real day only. Rather, we would have to measure at least several days worth of observed accessibility and produce an aggregate measure to summarize the distribution of travel times from general moments of departure. Conway, Byrd, and Eggermond (2018) for example use the 90\textsuperscript{th} percentile travel time as a robust measure, providing some assurance to a departing passenger that they will likely have time to complete the trip.

Reliability oriented methods should be extended as well to explicitly time-geographical studies of person-based accessibility (e.g. Hägerstrand, 1970). Rather than averaging over travel time variation attributable to a schedule (e.g. Allen and Farber, 2018b) one should look at actual variations in travel time attributable to both schedule and reliability issues. This could allow us to model for example passenger preferences for on-time arrival at the destination by approximating a buffer time added to their departure.

### 5.2.5 Health Geography and Path Exposure Modelling

Neutens (2015) provides a review of recent work on issues of accessibility in health geography and argues that more temporally aware and time-geographical methods are needed from transport geographers. He also makes the case for more person-based measures of accessibility and exposure, rather than simply measures associated with peoples’ home locations (see Kwan, 2009). Such measures are beginning to be used in models of exposure during travel and while using transit, specifically whether in the context of exposure to elemental (Fraser, 2016; Fraser and Chester, 2017), social (Burdick-Will, Stein, and Grigg, 2019), or other health related factors (Widener, 2017).

Such models have typically been estimated by deriving a shortest path from schedule-based GTFS data and trip-planning applications. From that path, measures of geographic proximity are used to estimate exposure to nearby phenomenon during the trip. As argued in Chapter 4 the shortest path at any given moment may well not be the one likely to be taken. Buliung et al. (2013) provide empirical evidence...
evidence of this in the context of walking and driving shortest paths but to date the issue of generating more realistic paths for transit users has not been addressed directly in this context. I would suggest that the existence of a large literature on route-choice among transit passengers (See Section 4.3.4) tends to indicate that other factors beyond travel time minimization are at play. Assuming that people are at least as much influenced by actual transit services as by schedules, we should assume that a retrospective GTFS would be a better starting point and that heuristic route-selection algorithms may produce more accurate results in some contexts.

5.3 Limitations

5.3.1 Scope, Data, & Computation

This research addresses methodological issues in the use of a standardized dataset which is available for hundreds if not thousands of transit agencies over the course of the last decade. Chapter 3 analyzed four agencies while Chapters 2 & 4 used only the Toronto Transit Commission. All chapters used less than a week of observations. While efforts were made to pick a typical week for observation and diverse agencies as case studies, this remains a very limited sample from a very large population.

The selection of only a few cities was necessitated by the substantial time and attention that was required for the production of reasonably accurate retrospective GTFS data. Quality of the source data varied between cities and many idiosyncrasies needed special attention. In particular, the OpenStreetMap data used for map-matching AVL data to the street/rail networks was often not inclusive of major transit facilities not accessible to the general public. I spent a great deal of time looking through bad matches to find the source of the problem before mapping bus garages and amending turn restrictions which did not apply to buses. In San Francisco, to make the matching work, I had to spoof a substantial underground streetcar tunnel so that it aligned with the misaligned AVL position data, which meant researching and debating the precise position of a tunnel on the other side of the continent. Sampling frequencies in the AVL data also varied, introducing inconsistencies in the standards I tried to establish for data quality and necessitating a large amount of manual sampling and verification of data for each agency.

Despite these efforts not all data could be cleaned and processed to my satisfaction; a small portion of the AVL data for each agency had to be removed from the analysis as discussed in Section 2.4.4. While every effort was made to ensure that such removed trips were not spatially or temporally concentrated\(^5\), it is not known what effect this missing data had on the analysis. It might be possible in future research to conduct a sensitivity analysis by artificially dropping some number of trips and measuring the impact on any particular measured outcome. This could address the problem of data which is known to be missing to some measurable degree by placing confidence intervals on any estimates. It would not however be able to address the “unknown unknown” of data missing from the real-time source data itself. Such a concern raises broader issues of the need to validate data against some kind of standard. This is difficult however as this research itself is based on the assumption that the real-time data is a viable standard against which GTFS schedule data can be validated.

The time needed for calculations was also a major limitation. OpenTripPlanner in particular was poorly optimized for large scale analysis\(^6\), let alone analysis of irregular data not based on a schedule.

---

\(^5\)Such clusters were sought out as indicators of problems which could be resolved in some other way prior to the analysis.

\(^6\)Conveyal has developed a newer tool better suited to this purpose called R5.
Computers, including resources on a super-computing cluster, spent months churning through large datasets with tools not built for parallel computing. Purpose-built software would be extremely helpful for any future analysis of retrospective GTFS data.

This analysis was also limited by the availability of data. Transit agencies which did not offer high-quality, open, realtime data could not be considered for this analysis. The tools developed in Chapter 2 were built around the TTC which supplies realtime data through an API provided by NextBus. As I continued working with that tool, the agencies I had to choose from for a more extended analysis were necessarily clients of the same company. Many of the roughly forty “agencies” in that cohort were tiny (single streetcar lines or university shuttle systems) while others had low sampling frequencies for their vehicle locations feeds or even big gaps with missing location data, as was the case for Los Angeles. The retro-gtfs software is currently being extended by other researchers to work with the GTFS-realtime standard, which though more common can also be more difficult to work with due to the way it is encoded. It seems very likely that many of the same data quality issues will be present in that data source as well.

### 5.3.2 Methodological Consistency & Validity

Each of the three main chapters implements a different spatial sampling strategy and different ways of calculating accessibility, making the results of the three parts difficult to compare directly. This research began with Chapter 2 using as a demonstration a very simplistic and conventional measure of cumulative access-to-jobs from a set of population weighted census zones. On reflection however this metric seemed to be quite sensitive to a zone’s distance from the dense cluster of jobs in downtown Toronto. It also had the drawback of excluding jobs just outside of the city boundary like the area around Pearson International Airport which is the second largest employment cluster in Canada. Chapter 3 sought to correct these problems by using a set of unweighted OD points which was spatially dispersed along each agency’s lines; it also added consideration of the negative exponential impedance function to supplement the more sensitive cumulative opportunities measure. The rationale for this choice is discussed in Section 3.4.3. Finally, Chapter 4 did away with the impedance functions entirely, dealing instead with the underlying distributions of travel times on which these are typically based, as explained in Section 4.4.2. It also used a somewhat different set of OD points and sampled from these rather than calculating travel time exhaustively as before.

Thus an effort toward methodological progress resulted in inconsistency between the chapters. This is unfortunate given that each uses the TTC as a case study; it would otherwise be natural to compare, contrast, and combine the results of the three chapters as each builds on the one before. In particular, Chapters 2 and 3 both present maps of error (See Equation 3.5) in averaged place-based accessibility for the TTC (Figures 2.4 & 3.5 respectively), though distinctly different spatial patterns are apparent in the two maps. Without consistent methods, it may be impossible to know whether such differences are attributable to the different distribution of destinations, the different time period, the different means of aggregating momentary access scores, or to some combination of these and other factors. It would also be desirable to be able to use the findings of Chapter 4 to refine Chapter 3’s estimate of the degree to which net accessibility has been overestimated in standard models. Again however the inconsistency makes this difficult; the most expedient means of addressing this question is likely to conduct a new study with this particular issue in mind.

Reviewers throughout this research have objected that any particular accessibility measure may not
be ideal for some purpose or generalizable to another measure which is closer to their own interests
due to issues of MAUP, scale, boundary effects, parameter selection and so on. This applies to the
selection of an impedance function and parameters, to the choice of origins and destinations and their
respective weights, and to the use or neglect of e.g. competitive (Merlin and L. Hu, 2017) or subjective
utility (Nassir et al., 2016) paradigms. Rather than narrowing the focus of this research to some very
particular and defensibly valid accessibility application, I hoped to maintain generalizability byretreating
toward increasing methodological abstraction. Some other researchers have taken a different approach,
seeking to maintain generalizability by offering a variety of applied measures (Stepniak et al., 2019;
Pereira et al., 2018) or the use of a range of parameters (Owen and Murphy, 2018a). While this may be
useful to a reader with some particular metric in mind, it also tends to complicate the work overall by
introducing too many different methods, datasets, and variables.

I believe this research took the right approach given the options available, however it must be noted
that the neglect of more traditional applications of accessibility metrics introduces its own limitations.
In particular, Chapter 3 disregarded explicit landuse and the actual spatial distribution of travel demand
in each city, instead using transit service itself as a proxy. This resulted in generally more uniform dis-
tributions of origins and destinations than would be the case if one were looking at common accessibility
measures such as total residential access to jobs or retail outlets. This relative dispersion may have un-
duly weighted observations on parts of the transit networks which serve relatively few real travellers. It
is possible that these less frequented routes are more reliable because they are less congested. However it
is also possible that they are less reliable because their improvement is a lower priority to operators. The
degree and direction to which this observation bias has skewed the results is difficult to guess without
further study. Similarly, the choice of simple cumulative and exponential impedance functions limits
the extension of this research to more advanced and theoretically valid measures such as those based on
perceived utility. Such a measure might for example account for the different perceptions of time spent
waiting versus time spent travelling and be much more sensitive to long or irregular wait times which
can result from unreliable service. However as trip travel time components were not analyzed in this
research, it is difficult to speculate how this work would apply to the use of such measures.

5.3.3 Perspective and Selection Bias

This work was written in Toronto by a regular user of the TTC - a very large agency with very frequent
yet often surprisingly unreliable services and a uniquely rectilinear network structure (See Mees, 2010).
My focus in this dissertation on reliability and route choice clearly comes out of my own interests in this
context and there are serious issues which I have not addressed here - issues that tend to more directly
affect other kinds of passengers or other kinds of transit agencies.

For example, I would encourage future researchers to look into the ways accessibility is affected by
issues of onboard crowding. It is often the case in Toronto that vehicles are simply too full for passengers
to board, especially during the rush hours. This does not effect all services evenly and likely has a unique
spatial distribution which would tend to lower access to some parts of the city more than others. Many
major transit agencies around the world struggle with crowding to some degree. One could also choose to
incorporate fares into an accessibility measure (as Cui and Levinson, 2018; Nassir et al., 2016) and look

---

7 Conceptually, accessibility can be thought of as measuring the utility of the transport-landuse system as an integrated
whole (See Geurs and Van Wee, 2004) so it makes little sense *prima facie* to disregard half of the equation. However the
separate consideration of modes in much accessibility analysis substantially complicates the picture especially when the
use of different modes (e.g. car versus transit) exhibits distinct spatial distributions.
at the way fare zone boundaries create discontinuities in accessibility surfaces for low income users. One could surely spend a lifetime studying the way various mobility, vision, or cognitive impairments affect accessibility via transit. Grisé et al. (2018) address wheelchair accessibility in a very preliminary way using GTFS data, though in general much better data is needed than is presently available; this applies both to the way GTFS models wheelchair accessibility as well as the way street/pedestrian networks are often represented at a scale not appropriate to routing pedestrians in detail. Such specialized issues have been given exceedingly little attention in the recent profusion of transit accessibility literature.

5.3.4 Extrapolation of Findings to Other Types of Agencies

Given the limits addressed above, this work necessarily speaks most directly to Toronto and places like it — dense urban areas with large and congested transit systems. The findings presented here may have limited relevance to agencies or services which do manage to maintain a relatively high degree of schedule adherence, or which provide few viable alternative routes for most trips. For example I would expect this work to have little to say to an accessibility analysis that focused solely on regional commuter rail systems which tend to be both both radial in structure and fully separated from competing traffic.

By contrast, this work may be especially relevant in places where schedules are even less important or reliable than they seem to be in Toronto. Other countries with more informal transit such as jitneys and competing private contractors may have even more to gain from an analysis of actual operations rather than schedule data. Williams et al. (2015) for example discuss their project to construct a GTFS dataset for Nairobi’s “semi-formal” bus network, though it seems clear from their methods that the resulting schedule would be only a rough approximation of actual service. It is not known how viable it would be to actually monitor vehicles in such a case with GPS devices as the technique outlined in Chapter 2 would require, but it seems clear that lacking such data our ability to measure accessibility via such services may be very limited at present. Similarly, Pereira et al. (2018) make a study of schedule based accessibility via transit in Rio de Janiero, despite doubts that all scheduled services were actually being operated as specified by private contractors responsible for providing the services. In this case an attempt was made to use retrospective GTFS data, however a low sampling frequency combined with missing AVL data prevented any analysis which would have produced a sufficient degree of accuracy.\footnote{Substantial effort went into making this dataset work and I would like to thank Dr. Pereira for giving me the opportunity to test my methods against such a tricky adversary.}

5.4 Extensions Beyond Accessibility

5.4.1 Historically Informed Trip-planning

The approach of looking to past transit performance to inform forward-looking measures (see Section 1.3.1) can be applied to more than just travel time estimates and accessibility metrics. Retrospective GTFS data could be used to inform transit trip-planning applications rather than or in addition to schedule and real-time data. This is already being done in automotive trip-planning applications where estimates of edge-level traffic speeds which vary by time of day are now often based on large GPS probe datasets gathered from drivers offering up their locations via onboard navigation systems.

This could be similar to the approach used in Chapter 4, where historical travel time outcomes were generated for a variety of alternate itineraries. Section 4.4.6 of that chapter used my morning commute
to illustrate that method and the exercise has actually changed the way I to get to the office, indicating that I may have been using a sub-optimal itinerary for more than a year without knowing it. I did the same analysis for the return commute in the evening and found that the optimal strategy suggested a different return itinerary and that I was not optimizing my time well there either. I do not generally use real-time data in trip planning except where it is available at stations. My own heuristic is more like Chapter 4’s habitual strategy - though my experience was that on-the-ground trial and error was not enough to learn an optimal habitual route even for someone studying the issue.

At present, real-time trip-planning applications typically base their main network graph on schedule data, updating this with real-time predictions where possible at any given moment. Such models are only more useful than schedule data alone when queried en-route or at the moment of departure. This means that the useful information contained in real-time (AVL) data is only truly accessible to someone who is willing to allow the trip-planner to determine rather than inform their route-choice. More conservative travellers such as myself need more understanding of and control over the alternatives than current trip-planning applications tend to allow.

What I envision is that retrospective GTFS or similarly structured archived AVL data could undergird an historically informed transit trip-planning tool, which could give travellers general rather than specific route suggestions. For example, say I want to plan a trip to an unfamiliar airport for an early morning flight. To feel comfortable I need to know a route and estimated travel time before setting an alarm and going to bed. Present models would only be able to use schedule-based data for a planned route so far in the future, however there should be benefit to observing how various alternative routes to the airport have fared in the past for similar early morning trips, perhaps even for similar times of year, weather conditions, etc. Such problems may perhaps be amenable to machine learning algorithms such as have long been used in real-time arrival predictions (e.g. Chung and Shalaby, 2007). Instead of using historical data for a single route however, these could learn network level characteristics for trips involving transfers, etc. Another alternative is just to define a set of possible itineraries for a trip and provide the passenger aggregate descriptive statistics on each alternative’s past performance, allowing them to sort these by any number of criteria such as fastest, most reliable, etc.

5.5 Open Data Standards and the Future of Accessibility

This research was made possible thanks to AVL data I was able to scrape from public-facing real-time API’s and store on my own computer for post hoc analysis. I have no access to AVL data from before I started my observation period and no knowledge of what has happened since I stopped. Some transit agencies archive their own AVL data and retain this for various internal purposes, though I have not known an agency to release such data to the general public. There is however currently work underway at the Center for Urban Transportation Research (CUTR) at the University of South Florida to continuously archive large amounts of publicly available real-time transit data from a number of agencies and make it available retrospectively for a variety of applications (Nieves et al., In Review). This would include any data provided by these agencies through their implementation of the GTFS-realtime standard such as vehicle locations and estimated stop-level arrival predictions. The possibilities for a standardized archive of historical data like this are at least as great as the possibilities presented...
by the standardization of schedule data in the General Transit Feed Specification. This is especially true if, like GTFS, such an archive eventually provides reliable access to data spanning several years or more. The reason GTFS has taken off as a tool for accessibility analysis is precisely that it is available for so many transit agencies all around the world in the same format. When combined with other standardized data sources like OpenStreetMap, accessibility applications based on open standards have great potential to attract a large pool of users and developers, extending their utility beyond the confines of any particular municipality.

While detailed, realistic models of particular transit systems have been developed in the past (E.g. E. J. Miller et al., 2015; Cats, 2011) the real potential of the techniques described in this research lies in the use of open, standard data formats now widely available. Transportation modelling typically takes as its purview a single administrative region’s transport system. Operating as a part of the government that has commissioned their work, modellers have access to internal data on transport operations not available to anyone else.\(^{10}\) They also have the time and resources to develop an application around all kinds of local or administrative idiosyncrasies and prerogatives. Such models can provide a rational foundation for internal planning and decision processes however they provide no means of testing alternative perspectives or enabling cross-border comparisons. The literature on public transit accessibility however has asked precisely the kinds of questions that such administrative modelling has not been able to answer. It has for example helped to develop perspectives on social justice which are fundamentally critical of administrative priorities. It has developed methods to allow non-experts to test their ideas in the realm of transit network planning (Stewart, 2017b). It has opened the possibility of massive inter-agency and inter-regional comparisons (Owen and Levinson, 2016). It has opened transit route and accessibility modelling to public health researchers (Neutens, 2015). And it has allowed for a standardization of software enabling accessibility analysis by relative laypeople for any imaginable purpose, whether critical, commercial, academic, or driven by simple curiosity (McGurrin and Greczner, 2011). Open data also enhances the reproducibility of any such efforts by enabling people to document their sources clearly and share data.

While the wide availability of GTFS data has opened many doors (Antrim, Barbeau, et al., 2013), I believe the development of standard, open historical AVL or real-time datasets when combined with GTFS and other open data\(^{11}\) could eventually lead to much more complete and accurate models of transit travel and accessibility than are presently available. Such models will be better able to fulfill the promise of ongoing work on the many frontiers of transit accessibility research.

---

\(^{10}\) There do exist standardized and even open-source transport modelling frameworks such as MATsim (Kucirek, 2012), though it seems that these often still require substantial efforts toward customization and localization.

\(^{11}\) See for example the proposed GTFS-ride standard which combines GTFS with data from an agency’s automated passenger counting system.
Bibliography


Antrim, Aaron, Sean J Barbeau, et al. (2013). “The many uses of GTFS data–opening the door to transit and multimodal applications”. In: Location-Aware Information Systems Laboratory at the University of South Florida, p. 4.


Ben-Elia, Eran, Robert Ishaq, and Yoram Shifman (2013). “If only I had taken the other road...”: Regret, risk and reinforced learning in informed route-choice”. In: Transportation 40.2, pp. 269–293.


Dunlap, Matthew et al. (2016). “Estimation of origin and destination information from Bluetooth and Wi-Fi sensing for transit”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2595, pp. 11–17.


Farber, Steven (2016). “Connecting People to Places: Spatiotemporal Analysis of Transit Supply Using Travel-Time Cubes”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2595, pp. 11–17.


Fraser, Andrew and Mikhail V Chester (2017). “Transit system design and vulnerability of riders to heat”. In: Journal of Transport & Health 4, pp. 216–225.


Hendren, Patricia et al. (2015). “Transit travel time reliability: shifting the focus from vehicles to customers”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2535, pp. 35–44.

Hu, Wen Xun and Amer Shalaby (2016). “Route and Segment Level Analyses of Bus Route Reliability and Speed Using Automated Vehicle Location Data”. In: *Transportation Research Record*.

Jang, Wonjae (2010). “Travel time and transfer analysis using transit smart card data”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2144, pp. 142–149.


Kramer, Anna and Alexandra Goldstein (2015). “Meeting the Public’s Need for Transit Options: Characteristics of Socially Equitable Transit Networks”. In: *Institute of Transportation Engineers. ITE Journal* 85.9, p. 23.


Lee, Young-Jae et al. (2001). “Effect of automatic vehicle location on schedule adherence for mass transit administration bus system”. In: Transportation Research Record: Journal of the Transportation Research Board 1760, pp. 81–90.


Lu, Ryland (2018). “Does Investing in Rail Transit Benefit the Poor? A Comparative Study of Rail and Bus Travel by Low-Income Households in the California Household Travel Survey”. In: Transportation Research Record 2672.6, pp. 11–20.


Mandelzys, Michael and Bruce Hellinga (2010). “Identifying causes of performance issues in bus schedule adherence with automatic vehicle location and passenger count data”. In: Transportation Research Record: Journal of the Transportation Research Board 2143, pp. 9–15.


OpenTripPlanner Project (2019). OpenTripPlanner on Github. URL: https://github.com/opentripplanner/OpenTripPlanner.


Thompson, Gregory L (2001). “New insights into the value of transit: modeling inferences from Dade County”. In: Transportation research record 1753.1, pp. 52–58.


Appendix A

Extraneous Figures

This section contains figures referenced in the text but not included there in an effort to maintain clarity.
Figure A.1: Median schedule-based $\bar{A}_{oh}$ (x axis) is plotted against Median $e(\bar{A}_{oh})$ (y axis) for each origin using the cumulative impedance function. Color gives the relative dispersion of errors for each origin as a quartile classification ($Q1$ $Q2$ $Q3$ $Q4$). Selected origins are highlighted in black and comparable results for the negative exponential function are given in Figure 3.4.
Appendix B

Characterizing OD Pairs

This section discusses in detail the regression models covered only briefly in Sections 4.5.2 & 4.5.3.

Given the variability which exists between OD pairs on measures of entropy and travel time difference, one may rightly ask what qualities of an OD pair are associated with higher or lower values. Are there spatial patterns? My approach to answering these questions was to conduct some exploratory regression analysis to see how well the variables of interest can be predicted and what attributes are positively or negatively associated. I produced a number of simple OD-level variables, listed below, based on the origin and destination locations. A few other variables were tried but found to be unrelated and are not listed here.

Table B.1: Description of regression variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>grid_dist</td>
<td>Manhattan distance in kilometers between origin and destination accounting for the orientation of Toronto's street grid</td>
</tr>
<tr>
<td>from_grid</td>
<td>Minimum angle between the line formed by the OD pair from the street grid of Toronto, which is defined as an orthogonal axis rotated 17 degrees counterclockwise from true north.</td>
</tr>
<tr>
<td>crosses_sub</td>
<td>Boolean variable indicating whether a straight line connecting the origin and destination crosses one of the subway lines</td>
</tr>
<tr>
<td>real_flow</td>
<td>Approximation of actual daily transit travel demand between origin and destination based on data from the 2016 Toronto Travel Survey.</td>
</tr>
<tr>
<td>max_from_sub</td>
<td>Maximum distance in kilometers of either origin or destination from the nearest subway line</td>
</tr>
</tbody>
</table>

My initial assumption was that trips would be more complex and harder to optimize with a heuristic strategy when they are longer and at greater angles relative to the street grid. The reason for this is that longer trips will tend to present more opportunities for transfers to other lines and that trips at an angle to the grid will require transfers due to the generally orthogonal structure of the transit network in the city. Trips on the diagonal can generally either go “up-and-over” or “over-and-up”, presenting at least two alternatives. I also suspected that OD pairs relying heavily on the subway would tend to produce fewer options and more similar travel times due to 1) the fact that this part of the system was modelled with schedule data and was thus artificially reliable and 2) the extremely short headways ($\approx 3$ minutes) and reasonably fast travel on these lines were expected to dominate nearby alternatives. Finally, I expected OD pairs with fewer actual passengers would tend to have less reliable service on
Appendix B. Characterizing OD Pairs

the assumption that the TTC would prioritize heavily travelled routes with e.g. express services or designated rights of way.

The analysis presented here is exploratory in spirit and meant primarily to suggest relationships with variables to help the reader understand some potential causes of route-choice complexity (measured with S) and differences in estimated travel time outcomes which can be measured in a variety of ways. The intention is not to offer generalizable evidence of any kind as many of these variables will have little relevance in the context of other transit agencies. I focus rather on the degree to which the variables of interest can be explained by the initial hypotheses. Rather than conducting a separate analysis for each time period, these are combined together in the same models. To the extent that some of the variation in S, etc. exists between time periods, some of that variation will not be explainable by these models.

The results of regressing S (See Equation 4.1) on the variables in Table B.1 are given in Table B.2 below. The expected signs are seen for variables except for those to do with the subway: crosses_sub which is positive and max_from_sub which is insignificant. That crosses_sub indicates a positive relationship may mean that crossing the subway is positively associated with the complexity of the optimal path, S. This contradicts expectations but may reflect the fact that subway stations are often transfer points serviced by multiple lines. Roughly 26% of the variance is explained by the model.

Table B.2: Regression predicting S per OD pair and time period. N = 4,000, r² = 0.26

<table>
<thead>
<tr>
<th>variable</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.715</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>grid_dist</td>
<td>0.052</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>from_grid</td>
<td>0.016</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>crosses_sub</td>
<td>0.248</td>
<td>2.19e-13</td>
</tr>
<tr>
<td>real_flow</td>
<td>-0.004</td>
<td>1.70e-06</td>
</tr>
<tr>
<td>max_from_sub</td>
<td>-0.007</td>
<td>0.245</td>
</tr>
</tbody>
</table>

The same independent variables are now applied to a model of the difference in median travel times between the heuristic strategies and the optimal-path method; the distribution of this variable was shown in Figure 4.8. Table B.3 gives results for the habitual strategy and Table B.4 for the realtime strategy.

Table B.3: Regression predicting difference in median travel times for the habitual strategy per OD pair and time period. N = 4,000, r² = 0.19

<table>
<thead>
<tr>
<th>variable</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>1.023</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>grid_dist</td>
<td>0.114</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>from_grid</td>
<td>0.022</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>crosses_sub</td>
<td>0.254</td>
<td>0.00241</td>
</tr>
<tr>
<td>real_flow</td>
<td>-0.013</td>
<td>2.53e-10</td>
</tr>
<tr>
<td>max_from_sub</td>
<td>0.048</td>
<td>0.00100</td>
</tr>
</tbody>
</table>

Regressions on other travel-time-related variables produced very similar results, including models for differences in mean travel times and travel times at the 90th percentile. Except for the changing influence of max_from_sub, all three models show similar relationships, reflecting the fact that S is strongly correlated with any measure of difference in travel times. For the models given here, S has a Pearson correlation with the difference in median travel time of 0.78 for the habitual strategy and 0.61 for the realtime. In part, this relationship is baked into the definition of the heuristic alternative strategies, which will almost certainly choose the sub-optimal itinerary more often when the information
Table B.4: Regression predicting difference in median travel times for the real-time strategy per OD pair and time period. $N = 4,000, r^2 = 0.20$

<table>
<thead>
<tr>
<th>variable</th>
<th>$\beta$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.688</td>
<td>1.59e-06</td>
</tr>
<tr>
<td>grid_dist</td>
<td>0.156</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>from_grid</td>
<td>0.027</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>crosses_sub</td>
<td>0.443</td>
<td>5.59e-05</td>
</tr>
<tr>
<td>real_flow</td>
<td>-0.013</td>
<td>1.44e-06</td>
</tr>
<tr>
<td>max_from_sub</td>
<td>0.085</td>
<td>7.79e-06</td>
</tr>
</tbody>
</table>

content of the optimal path is higher. This however may be a necessary feature of any possible heuristic route-selection strategy. If that is the case then an estimate of $S$ associated with a given trip might reasonably be used as a way of inflating travel time estimates derived from a standard shortest-path routing algorithm, at least until better algorithms are developed. For the TTC, a 1 bit increase in $S$ was associated with a roughly 1.8 minute increase in estimated travel time at the median for the habitual strategy and 1.9 minutes for the realtime strategy. This coefficient would however likely be very different for different agencies.