

On the accuracy of schedule-based GTFS for measuring accessibility

Nate Wessel

University of Toronto

Steven Farber

University of Toronto Scarborough

Abstract: In this paper 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 access tend to produce less reliable estimates with more variability from hour to hour in real accessibility, while higher access zones seem to converge on an estimate 5-15 percent 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 access. 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.

Article History:

Received: December 12, 2018

Accepted: May 10, 2019

Available online: June 17, 2019

1 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 *et al.* 2013).

Copyright 2019 Nate Wessel and Steven Farber.

doi: [10.5198/jtl.u.2019.1502](https://doi.org/10.5198/jtl.u.2019.1502)

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – NonCommercial License 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

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 country (Owen and Levinson 2016). 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 access (e.g., Pereira *et al.* 2018; Widener *et al.* 2015) to accessibility as an input to mode-share models (e.g., Boisjoly and El-Geneidy 2016; Owen and Levinson 2015) to accessibility as a metric for use in transport planning decisions (e.g., Farber and Fu 2017; Farber and Grandez 2016; Stewart 2017b).

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., Bertini and El-Geneidy 2003; El-Geneidy *et al.* 2011), 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 paper 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.

2 Background

2.1 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 time¹, 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 *et al.* (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 access. And El-Geneidy *et al.* (2016a) compare the spatial distribution of transit access 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 Lee and Miller (2018) compare accessibility outcomes before and after a proposed bus rapid transit project in Columbus, Ohio. Conway *et al.* (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 access over time within a single city (Farber and 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 *et al.* 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., 8 a.m. for the morning commute) and estimated travel times from that moment only (e.g., Ma and Jan-Knaap 2014; Widener *et al.* 2015). 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 *et al.* 2016a), 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 access over time (e.g., Anderson *et al.* 2012; Farber and Fu 2017). 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 access is to compute momentary accessibility measures which are averaged together into a single score. This is the approach

¹ For examples of other metrics applied to accessibility, see Cui and Levinson (2018); El-Geneidy *et al.* (2016b).

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 *et al.* (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., Allen and Farber 2018b; Widener *et al.* 2017) or two (Allen and Farber 2018a; Farber and 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 *et al.* (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 paper 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.

2.2 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., Lee *et al.* 2001; Mandelzys and Hellinga 2010). 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 way that the results of the kind of studies described above are essentially unbiased. There is however substantial reason to doubt this.

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 2018).

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 2018), but it can be observed that agencies using timepoints often fail to include this data in their public GTFS data².

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 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 differences; without some empirical measurement, we are left to speculate.

2.3 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 access 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 access 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 access 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).

² The 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.

3 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.

All data used in this paper is available online at <https://osf.io/s5k3b/files/>.

3.1 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 GTFS package and AVL data for most of the fleet available through the NextBus API. A map of each agency is provided in Section 4.4.

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 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 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 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 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.

3.2 Retrospective GTFS

We implement the method described by [Wessel *et al.* \(2017\)](#) for constructing a routable retrospective GTFS package based on historical vehicle locations; we will briefly outline the method here, but refer readers to that paper 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 provides access to 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 6 through Friday the 10 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 5 & 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.

3.3 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-access problem with notions of competition ([Merlin and Hu 2017](#)) or skill-appropriateness ([Lee and 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 *et al.* 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 tessellation³ 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 4.4. Roughly 300 points per agency was chosen as a balance between coverage and computational feasibility. Some points that failed to properly snap to the street network were removed.

The choice of one particular distribution for origins and destinations will necessarily influence the outcome of the study and researchers are cautioned to think carefully about how results might have been different if accessibility were calculated to or from more concentrated or dispersed locations (e.g., [Stępnik et al. 2019](#)). Alternative spatial distributions were not tested in this study.

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 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 $((4 + 3) * 5)$ 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 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 access 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 2.1. Notably, we will not be discussing metrics that explicitly account for competition or reliability (e.g., [Conway et al. 2018](#); [Merlin and Hu 2017](#)) 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 1

$$A_{o,t} = \sum_d^D f(T_{odt}) \cdot W_d \quad (1)$$

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 percent, where 100 percent 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

³ For background on centroidal Voronoi tessellations, see [Du et al. \(1999\)](#).

o to all other destinations d (Equation 2).

$$\bar{A}_o = \sum_d^D \sum_t^T f(T_{odt}) \cdot W_d \cdot T^{-1} \quad (2)$$

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}_{ob} , letting b 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 θ (Equation 3).

$$f(T_{odt}) = \begin{cases} 1 & \text{if } T_{odt} \leq \theta \\ 0 & \text{if } T_{odt} > \theta \end{cases} \quad (3)$$

Another alternative, the negative exponential function (Equation 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, β , controls how quickly the curve diminishes.

$$f(T_{odt}) = e^{-T_{odt}/\beta} \quad (4)$$

A visual comparison of the two functions with typical parameters is given in Figure 1.

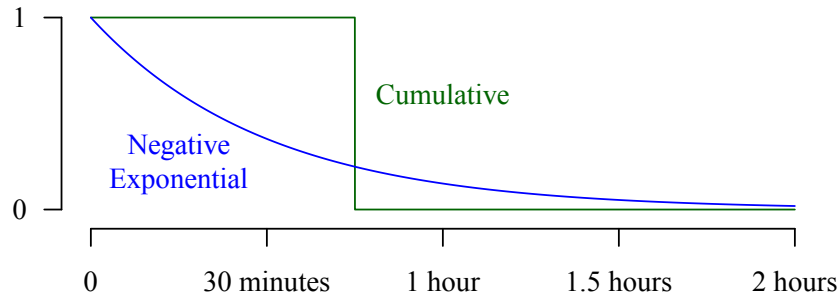


Figure 1: Negative exponential and cumulative opportunities impedance functions $f(T_{odt})$ with parameters $\theta = 45$ minutes (Equation 3) and $\beta = 30$ minutes (Equation 4).

3.5 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 5).

$$e(m) = (m^{retro} / m^{sched} - 1) \cdot 100 \quad (5)$$

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 $e(\bar{A}_{ob})$) 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 for example to missing or erroneous data in the retrospective GTFS package.

4 Results

We start the discussion of results in Section 4.1 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 4.2 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 4.3 where we observe spatial patterns in accessibility error. In Section 4.5 we take a close look at moment-to-moment variation in accessibility at a few selected origins.

4.1 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 2 plots scheduled travel times T_{odt}^{sched} 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.

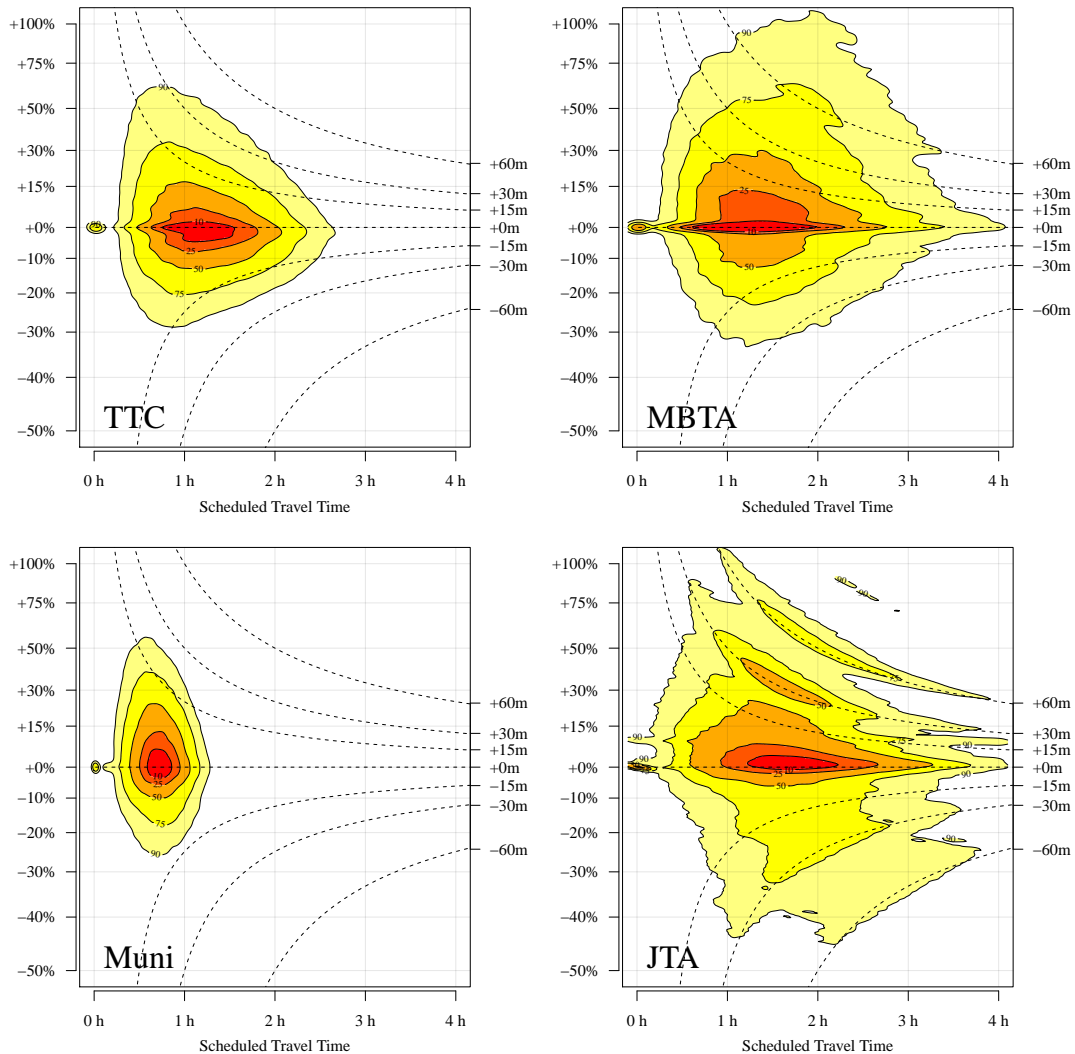


Figure 2: Scheduled travel times between all points (T_{odt}^{sched} , on the X axis) are compared against $e(T_{odt})$ on the Y axis in the form of a density plot with labelled percentiles. Absolute differences between travel times ($T_{odt}^{retro} - T_{odt}^{sched}$) are also given in minutes on the right margin.

There is a lot of information in Figure 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 smallest 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 $\pm 15\%$ 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 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.

4.2 Parameter Selection and Agency-level Error

To calculate accessibility it is necessary to select parameters for the impedance functions described in Section 3.4. 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, $e(\bar{A}_{ob})$. Figure 3 shows the median and interquartile range of $e(\bar{A}_{ob})$ for both the cumulative and negative exponential functions as parameters θ and β 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 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

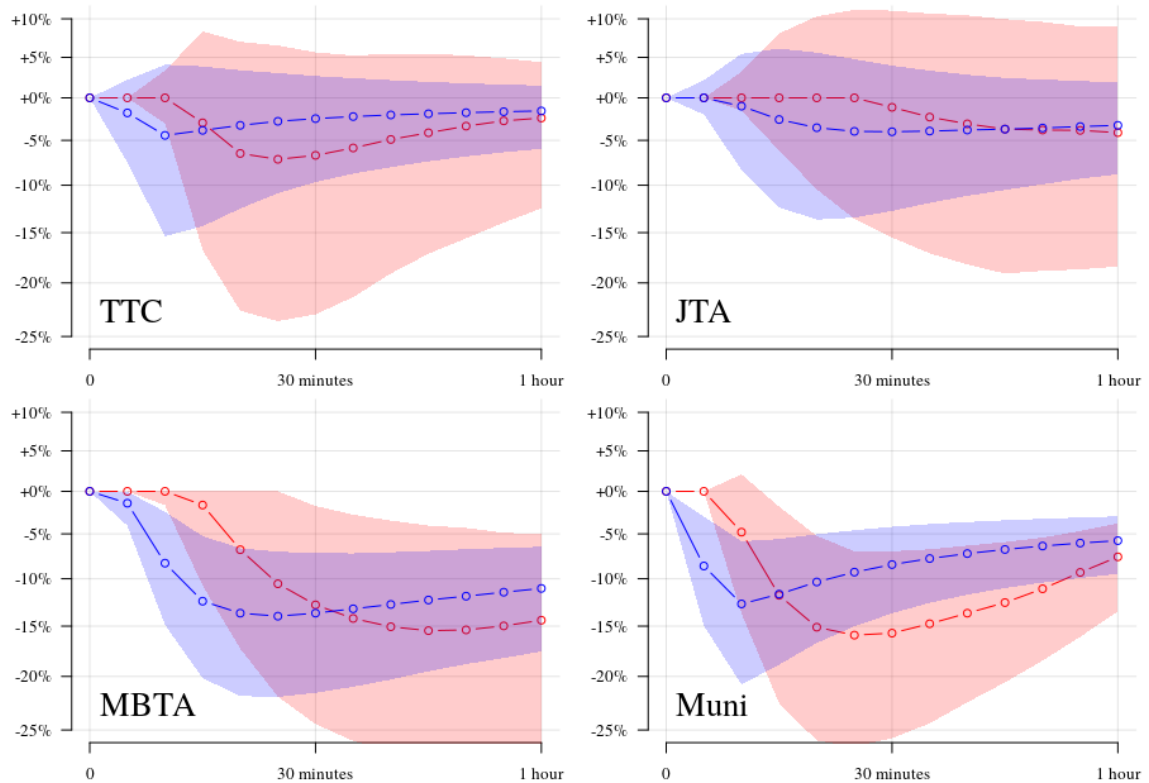


Figure 3: The median and interquartile range are given for $e(\bar{A}_{ob})$ as parameters vary between 0 and 60 minutes for the **cumulative opportunities** and **negative exponential** functions.

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 shows that at intermediate parameter values $e(\bar{A}_{ob})$ 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 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 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.

Table 1: Selected parameters for impedance functions defined in Section 3.4. Units are minutes.

Agency	Cumulative θ	Exponential β
TTC	45	30
JTA	45	30
MBTA	45	30
Muni	30	20

4.3 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 4 plots the median scheduled accessibility against the median $e(\bar{A}_{ob})$ by origin. Color is used to show the relative dispersion (interquartile range) of $e(\bar{A}_{ob})$ on a quartile scale (• Q1 • Q2 • Q3 • Q4) relative to each agency. 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 outsize 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 these patterns we select two representative origins from each agency which will be subjected to a more detailed analysis in Section 4.5. These origins are highlighted in Figure 4 as well as in the maps in Section 4.4.

Figure 4 also shows that there is a very large range in average $e(\bar{A}_{ob})$ 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-access 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.

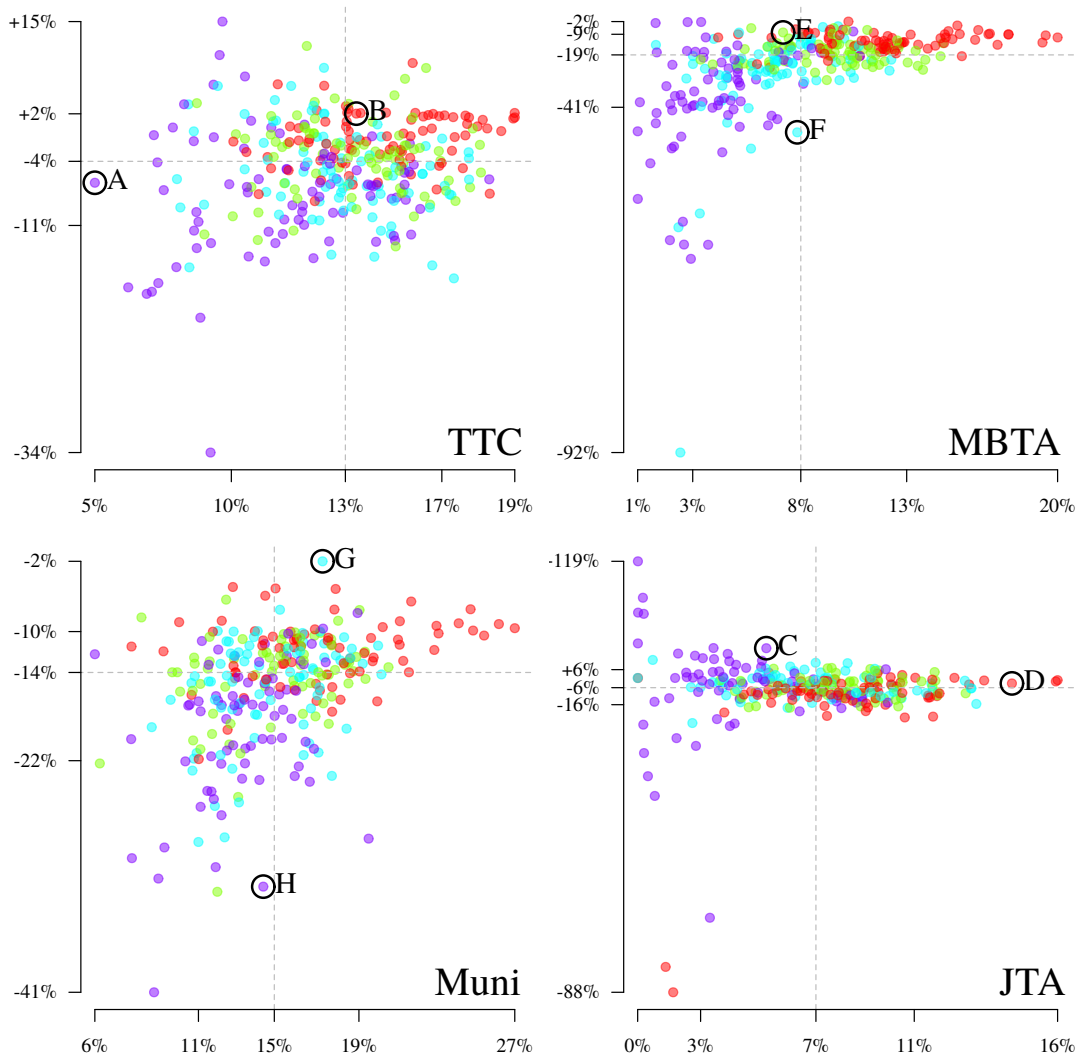


Figure 4: Median schedule-based \bar{A}_{ob} (x axis) is plotted against Median $e(\bar{A}_{ob})$ (y axis) for each origin using the negative exponential impedance function. Color gives the relative dispersion of $e(\bar{A}_{ob})$ for each origin as a quartile classification (● Q1 ● Q2 ● Q3 ● Q4) relative to each agency. Selected origins are highlighted in black and comparable results for the cumulative function are provided in Appendix A.

4.4 Spatial Patterns

To detect any systematic spatial patterns in $e(\bar{A}_{ob})$ 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 Delaunay 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 2 spatial autocorrelation is found to be strongly positive for all agencies and measures. The spatial distribution of median $e(\bar{A}_{ob})$ is mapped for all agencies at the end of this section (Figures 5, 6, 7, & 8) and visually confirms the findings of strong spatial autocorrelation. Selected points discussed in Section 4.5 are highlighted and labelled.

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 8, where definite linear patterns in errors are visible. The MBTA, as the other radially structured regional agency, Figure 6, exhibits

Table 2: Moran's I of median $e(\bar{A}_{ob})$. All values are statistically significant ($P < 0.001$).

Agency	Cumulative	Exponential
TTC	0.547	0.603
MBTA	0.815	0.855
Muni	0.679	0.681
JTA	0.692	0.805

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 5) and SF Muni (Figure 7), the two grid-structured agencies, are somewhat harder to discern. Both have obvious clusters of high and low values, but they don't 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 MBTA⁴. 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 *et al.* (2017) presents a brief 4-hour-average snapshot of error in transit access to jobs in Toronto⁵. 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's 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.

⁴ Cf. Stewart (2017a) Figure 8.7 page 178 with our Figure 6.

⁵ Cf. Wessel *et al.* (2017) Figure 4 page 95 with our Figure 5

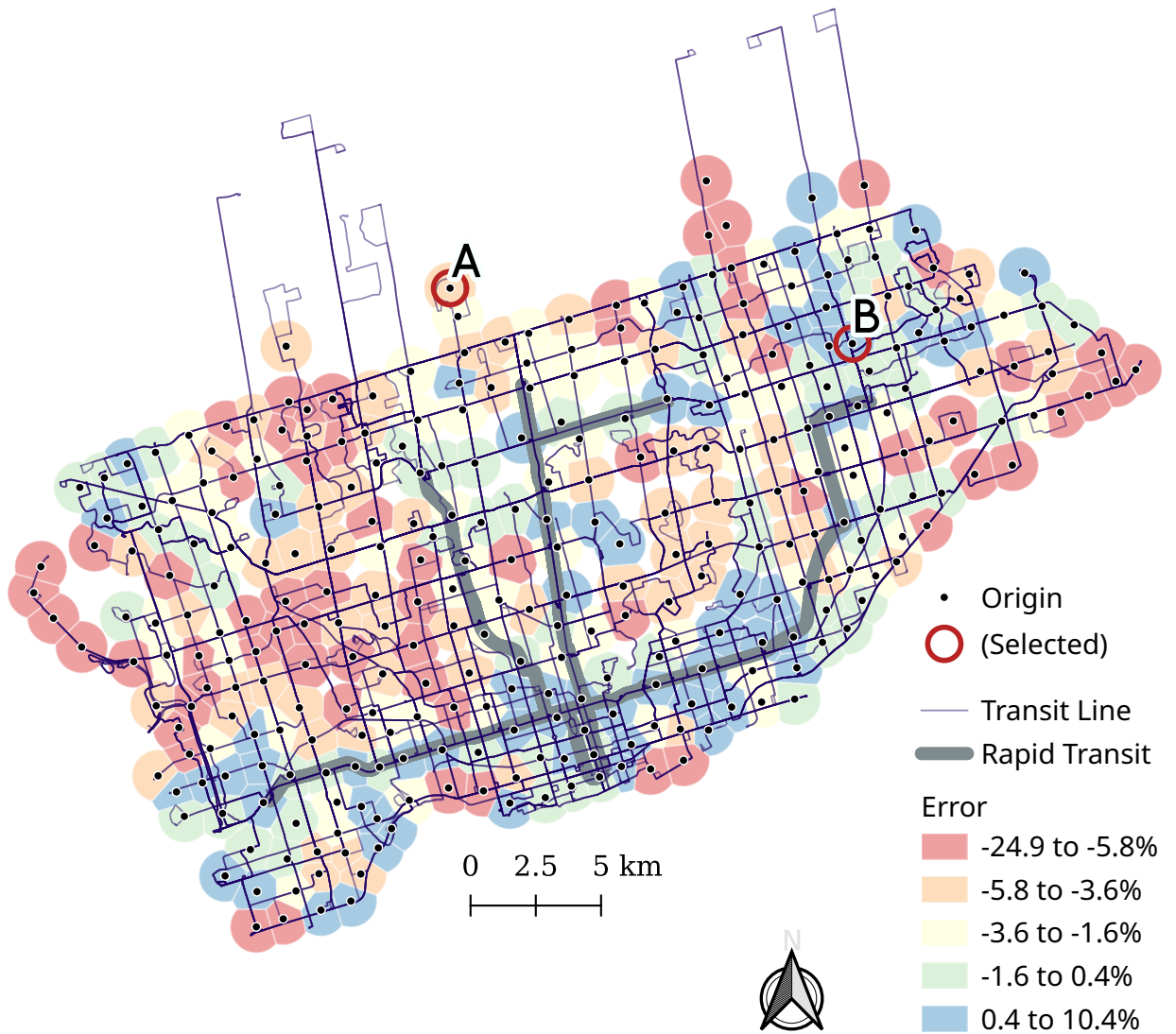


Figure 5: TTC. 327 total O/D points. Note that the rapid transit lines were derived from schedule-based GTFS data only. Color indicates median $e(A_{ob})$ for each zone using the negative exponential impedance function.

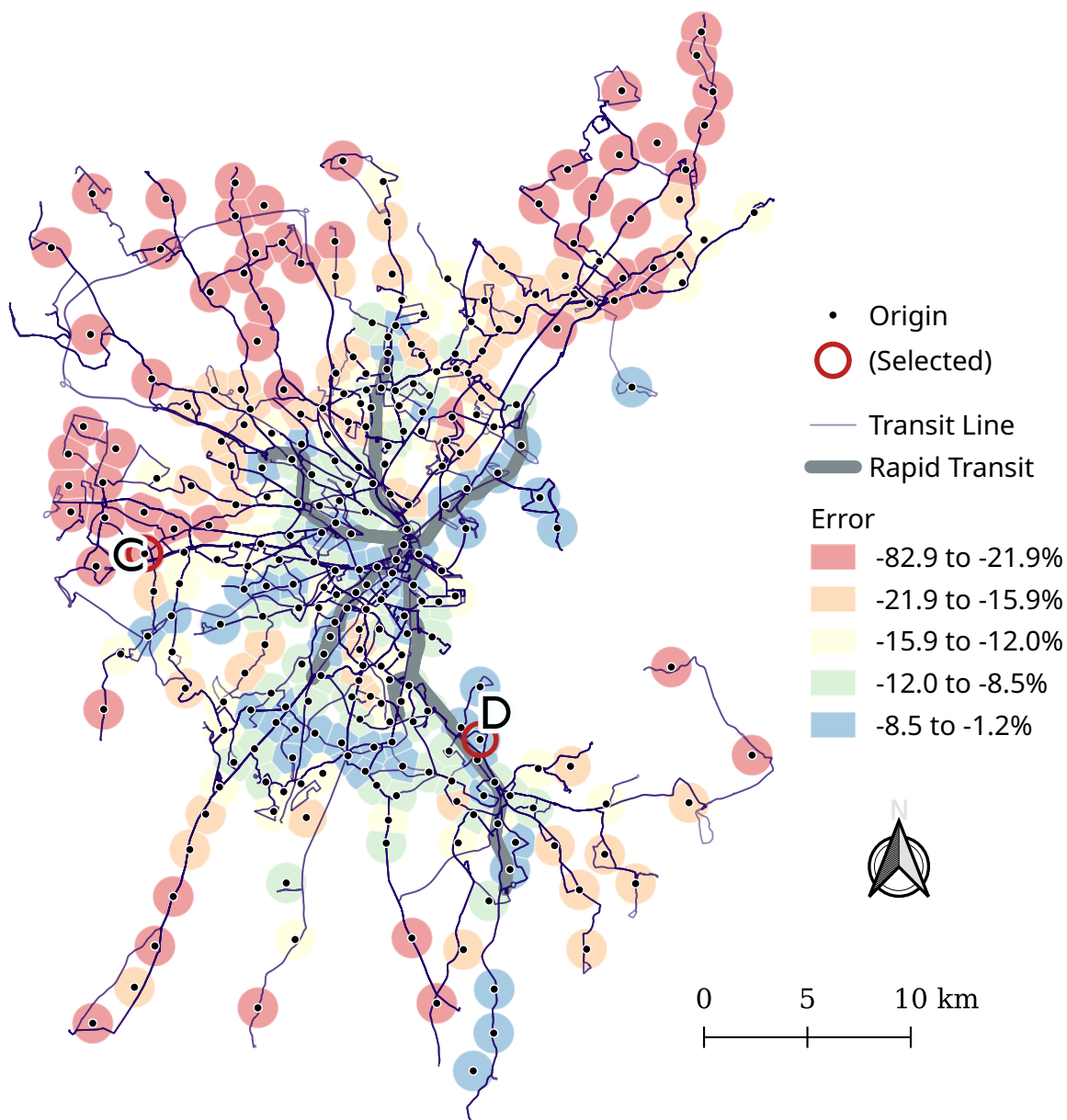


Figure 6: MBTA. 294 total O/D points. Note that the rapid transit lines were derived from schedule-based GTFS data only. Color indicates median $e(\bar{A}_{ob})$ for each zone using the negative exponential impedance function.

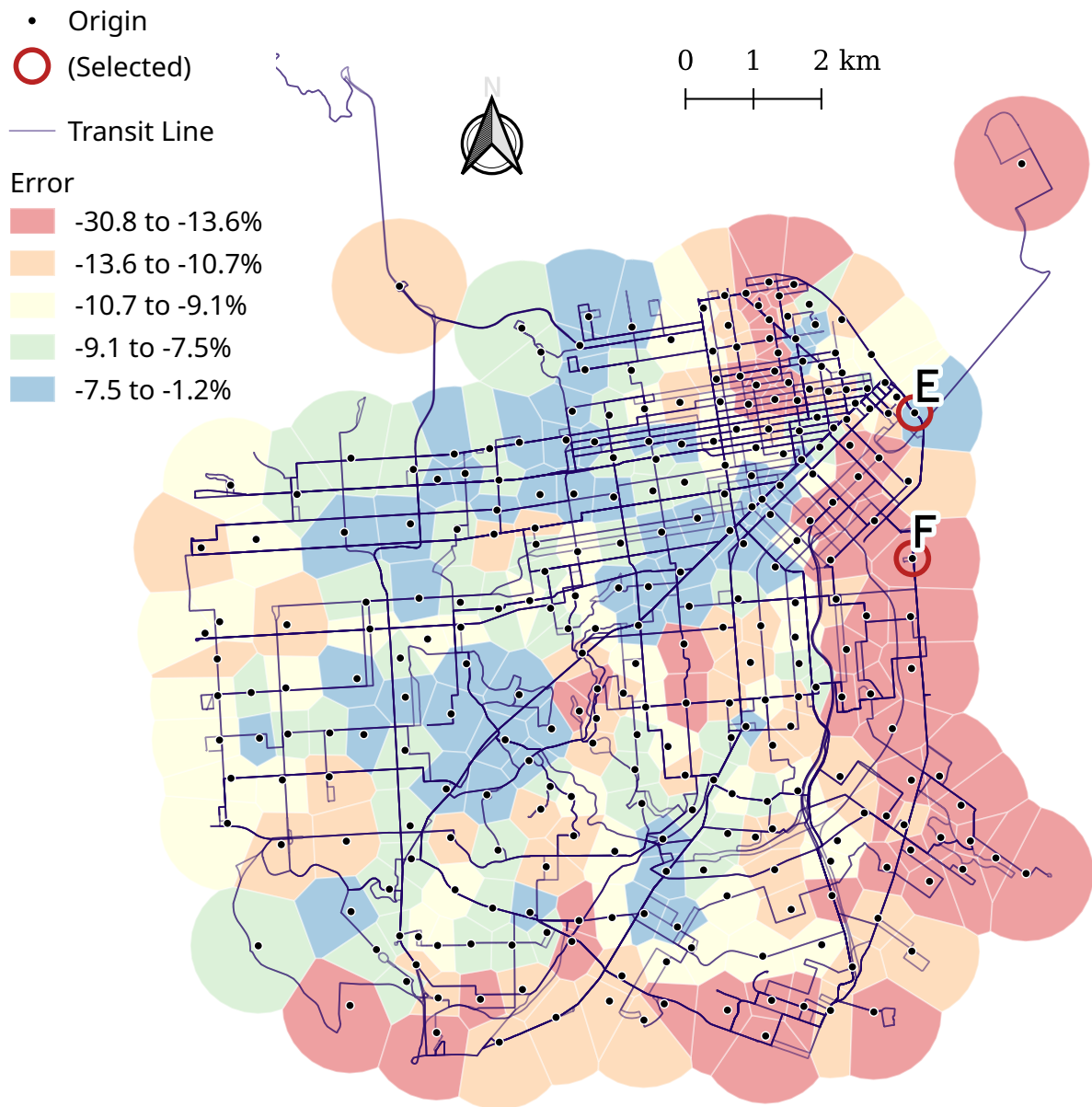


Figure 7: SF Muni. 294 total O/D points. Note that the Bay Area Rapid Transit System (BART) was not included in this analysis and is not shown. Color indicates median $e(\bar{A}_{ob})$ for each zone using the negative exponential impedance function.

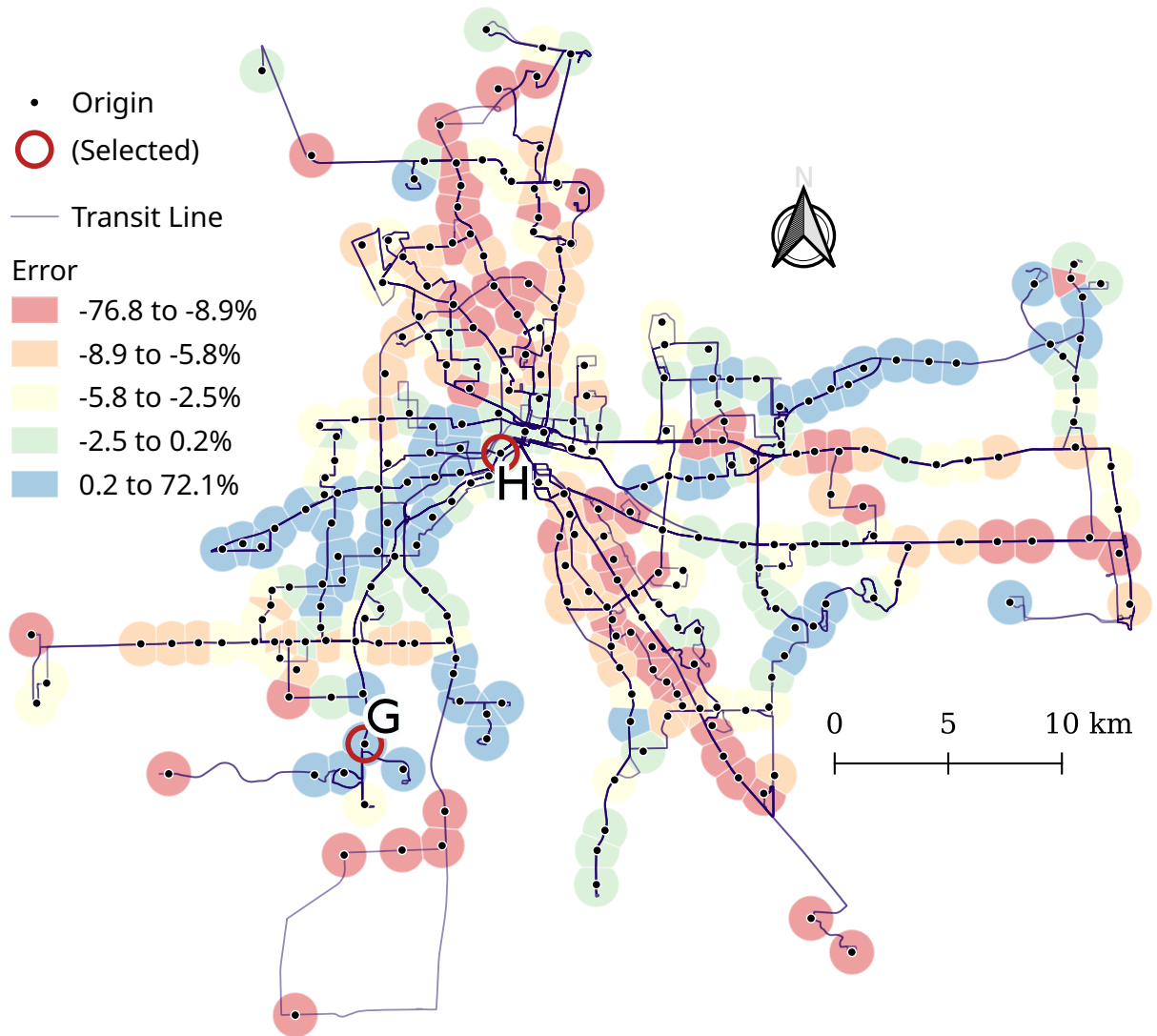


Figure 8: JTA. 282 total O/D points. Color indicates median $e(\bar{A}_{ob})$ for each zone using the negative exponential impedance function.

4.5 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 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.

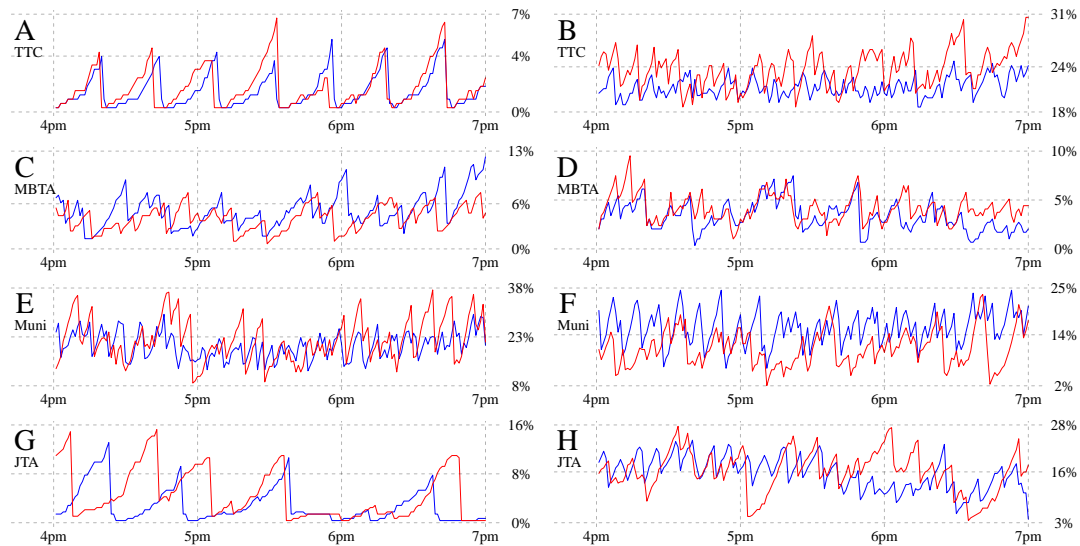


Figure 9: Change in A_{ot}^{sched} , A_{ot}^{retro} from selected origins over the 4-7pm period on Friday October 10th 2017. A_{ot} is given as a percentage of all origins accessible using the cumulative function. Origins are identified on the maps in Section 4.4 and in Figures 4 & 10. Note the variable range of the Y axis.

Both A and G are typical of access in zones with low-frequency service provided by a single transit line. A is an unusually low-access 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 of 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 (less negative) $e(\bar{A}_{ob})$, meaning that the schedule is less prone to overpredicting access for these origins. Yet, it seems hard to argue based on Figure 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} , i.e. 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 and higher or lower only sporadically.

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 $e(\bar{A}_{ob})$ while C and F both show definite negative median $e(\bar{A}_{ob})$. Access at origin D can be seen to track A_{ot}^{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_{ot}^{retro} misses a few peaks that appear in the schedule.

Origins E and F, both from the SF Muni, seem to show access 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_{ot}^{retro} and A_{ot}^{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.

5 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 percent 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's understandable why these global differences in access come about; for example, schedules generally show vehicles operating perfectly consistent headways, however 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 accessibility 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, θ . While the cumulative function is easier to interpret⁶ 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 et al. \(2018\)](#) 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 4.5 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 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 4.1) to measures of momentary accessibility (Section 4.5), to hourly averages (Section 4.3), 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 their observations. While the problem of sampling error in this context has been discussed elsewhere (e.g., [Conway et al. 2018](#); [Owen and Murphy 2018b](#); [Stępnia 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 access 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.

⁶ We use it ourselves in Figure 9 for this reason.

References

- Allen, J. and S. Farber. 2018a. Generating measures of access to employment for Canada's eight largest urban regions. Technical report, University of Toronto. URL <https://osf.io/preprints/socarxiv/pvrd9/>.
- Allen, J. and S. Farber. 2018b. How time-use and transportation barriers limit on-campus participation of university students. *Travel Behaviour and Society*, 13:174–182.
- Anderson, P. A., A. Owen, and D. M. Levinson. 2012. The time between: Continuously-defined accessibility functions for schedule-based transportation systems. Technical report, University of Minnesota.
- Anselin, L. and D. A. Griffith. 1988. Do spatial effects really matter in regression analysis? *Papers in Regional Science*, 65(1):11–34.
- Antrim, A., S. J. Barbeau, et al. 2013. The many uses of GTFS data—opening the door to transit and multimodal applications. *Location-Aware Information Systems Laboratory at the University of South Florida*, p. 4.
- Bertini, R. and A. El-Geneidy. 2003. Generating transit performance measures with archived data. *Transportation Research Record: Journal of the Transportation Research Board*, 1841(1):109–119.
- Boisjoly, G. and A. El-Geneidy. 2016. Daily fluctuations in transit and job availability: A comparative assessment of time-sensitive accessibility measures. *Journal of Transport Geography*, 52:73–81.
- Conway, M. W., A. Byrd, and M. van Eggermond. 2018. Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning. *Journal of Transport and Land Use*, 11(1).
- Cui, M. and D. Levinson. 2018. Full cost analysis of accessibility. *Journal of Transport and Land Use*, 11(1).
- Du, Q., V. Faber, and M. Gunzburger. 1999. Centroidal voronoi tessellations: Applications and algorithms. *SIAM review*, 41(4):637–676.
- El-Geneidy, A., R. Buliung, E. Diab, D. van Lierop, M. Langlois, and A. Legrain. 2016a. Non-stop equity: Assessing daily intersections between transit accessibility and social disparity across the Greater Toronto and Hamilton Area (GTHA). *Environment and Planning B: Planning and Design*, 43(3):540–560.
- El-Geneidy, A., D. Levinson, E. Diab, G. Boisjoly, D. Verbich, and C. Loong. 2016b. The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transportation Research Part A: Policy and Practice*, 91:302–316.
- El-Geneidy, A. M., J. Horning, and K. J. Krizek. 2011. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation*, 45(1):66–79.
- Farber, S. and L. Fu. 2017. Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis)investments over time. *Computers, Environment and Urban Systems*, 62:30–40.
- Farber, S. and M. Grandez. 2016. Transit accessibility, land development and socioeconomic priority: A typology of planned station catchment areas in the Greater Toronto and Hamilton Area. *Journal of Transport and Land Use*.
- Farber, S., M. Z. Morang, and M. Widener. 2014. Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, 53:149–159.
- Fransen, K., T. Neutens, S. Farber, P. De Maeyer, G. Deruyter, and F. Witlox. 2015. Identifying public transport gaps using time-dependent accessibility levels. *Journal of Transport Geography*, 48:176–187.
- Google. 2018. *General Transit Feed Specification Reference*. URL <https://developers.google.com/transit/gtfs/reference/>.
- Lee, J. and H. J. Miller. 2018. Measuring the impacts of new public transit services on space-time accessibility: An analysis of transit system redesign and new bus rapid transit in Columbus, Ohio,

- USA. *Applied Geography*, 93:47–63.
- Lee, Y.-J., K. Chon, D. Hill, and N. Desai. 2001. Effect of automatic vehicle location on schedule adherence for mass transit administration bus system. *Transportation Research Record: Journal of the Transportation Research Board*, 1760(1760):81–90.
- Ma, T. and G. Jan-Knaap. 2014. Analyzing employment accessibility in a multimodal network using GTFS: A demonstration of the Purple Line, Maryland. Technical report, University of Maryland, National Center for Smart Growth.
- Mandelzys, M. and B. Hellinga. 2010. Identifying causes of performance issues in bus schedule adherence with automatic vehicle location and passenger count data. *Transportation Research Record: Journal of the Transportation Research Board*, 2143(2143):9–15. URL <https://doi.org/10.3141/2143-02>.
- Merlin, L. A. and L. Hu. 2017. Does competition matter in measures of job accessibility? Explaining employment in Los Angeles. *Journal of Transport Geography*, 64:77–88.
- Michael, T. and T. Quint. 1999. Sphere of influence graphs in general metric spaces. *Mathematical and Computer Modelling*, 29(7):45–53.
- Owen, A. and D. Levinson. 2015. Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, 74:110–122.
- Owen, A. and D. Levinson. 2016. Developing a comprehensive US transit accessibility database. In *Seeing Cities Through Big Data*, pp. 279–290. Springer.
- Owen, A. and B. Murphy. 2018a. Access across America: Transit 2017. Technical report, University of Minnesota. URL <http://hdl.handle.net/11299/199920>.
- Owen, A. and B. Murphy. 2018b. Temporal sampling and service frequency harmonics in transit accessibility evaluation. *Transportation Research Record*. URL <https://trid.trb.org/view/1497217>.
- Pereira, R. H., D. Banister, T. Schwanen, and N. Wessel. 2018. Distributional effects of transport policies on inequalities in access to opportunities in Rio De Janeiro. *SSRN Electronic Journal*. doi: [10.2139/ssrn.3040844](https://doi.org/10.2139/ssrn.3040844).
- Stępnia, M., J. P. Pritchard, K. T. Geurs, and S. Goliszek. 2019. The impact of temporal resolution on public transport accessibility measurement: Review and case study in Poland. *Journal of transport geography*, 75:8–24.
- Stewart, A. F. 2017a. *Advancing accessibility: Public transport and urban space*. Ph.D. thesis, Massachusetts Institute of Technology. URL <http://hdl.handle.net/1721.1/111444>.
- Stewart, A. F. 2017b. Mapping transit accessibility: Possibilities for public participation. *Transportation Research Part A: Policy and Practice*.
- Wessel, N. 2015. *Discovering the space-time dimensions of schedule padding and delay from GTFS and real-time transit data*. Master's thesis, University of Cincinnati. URL http://rave.ohiolink.edu/etdc/view?acc_num=ucin1445342602.
- Wessel, N., J. Allen, and S. Farber. 2017. Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS. *Journal of Transport Geography*, 62:92–97. URL <http://saussy.ca/wp-content/uploads/2017/11/retro-GTFS-paper.pdf>.
- Wessel, N. and M. Widener. 2016. Discovering the space–time dimensions of schedule padding and delay from GTFS and real-time transit data. *Journal of Geographical Systems*, pp. 1–15. ISSN 1435-5949. doi: [10.1007/s10109-016-0244-8](https://doi.org/10.1007/s10109-016-0244-8). URL <http://dx.doi.org/10.1007/s10109-016-0244-8>.
- Widener, M., S. Farber, T. Neutens, and M. Horner. 2015. Spatiotemporal accessibility to supermarkets using public transit: An interaction potential approach in Cincinnati, Ohio. *Journal of Transport Geography*, 42:72–83.
- Widener, M. J., L. Minaker, S. Farber, J. Allen, B. Vitali, P. C. Coleman, and B. Cook. 2017. How do changes in the daily food and transportation environments affect grocery store accessibility? *Applied Geography*, 83:46–62.
- Zervaas, Q. 2018. Transitfeeds.com. URL <https://transitfeeds.com/>.

Zygo, A. 2017. *Measuring and Improving Seniors' Access to Medical Facilities*. Master's thesis, University of Connecticut. URL https://opencommons.uconn.edu/gs_theses/1042.

A Appendix: Additional figures

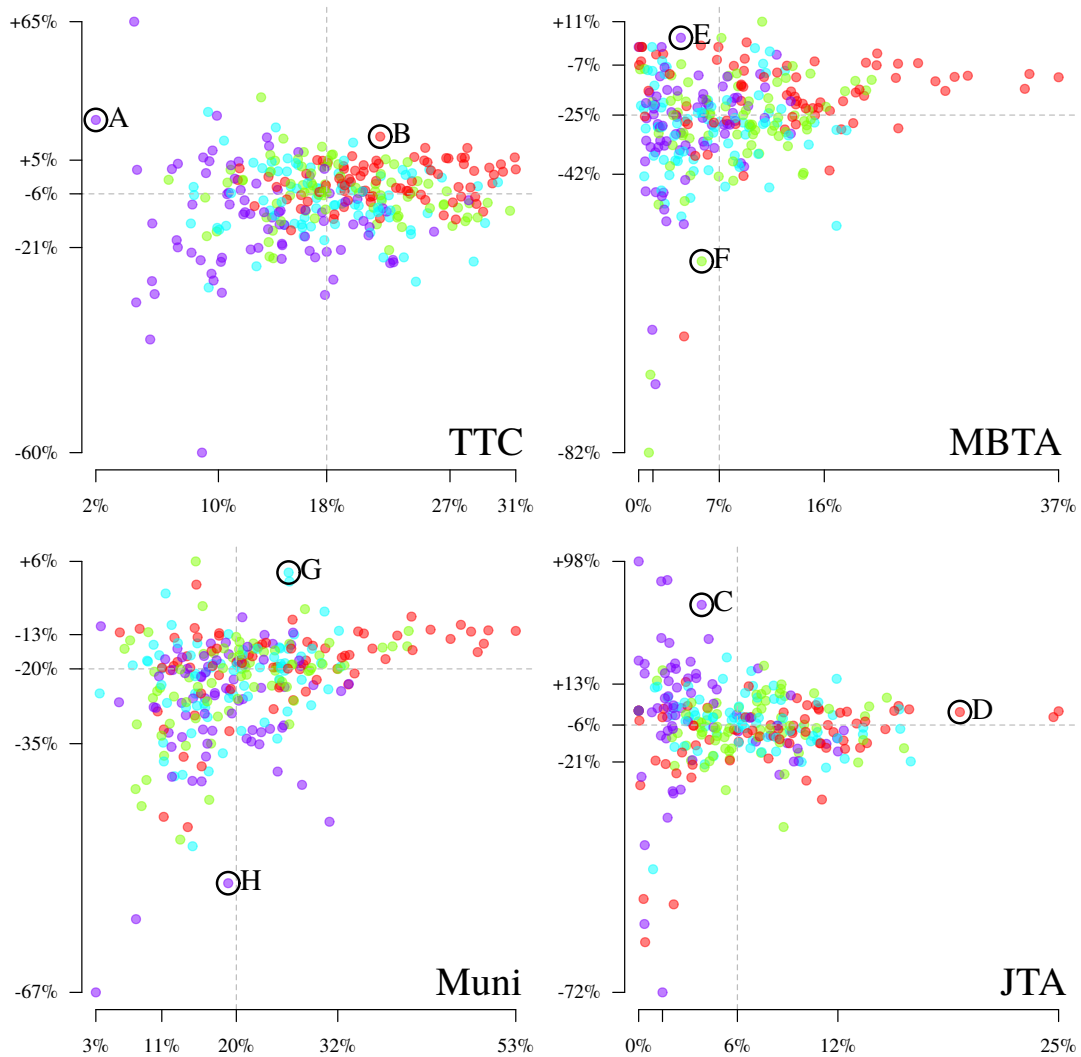


Figure 10: Median schedule-based \bar{A}_{ob} (x axis) is plotted against Median $e(\bar{A}_{ob})$ (y axis) for each origin using the cumulative impedance function. Color gives the relative dispersion of errors for each origin as a quartile classification (● Q1 ● Q2 ● Q3 ● Q4). Selected origins are highlighted in black and comparable results for the negative exponential function are given in Figure 4.