Understanding Bus Operations Using High-Resolution Vehicle Location Data

by

Yuzhu Huang

B.S., University of Washington (2018)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Yuzhu Huang
Department of Urban Studies and Planning
May 12, 2023

Certified by: John P. Attanucci
Research Associate, Center for Transportation and Logistics
Thesis Supervisor

Certified by: Anson F. Stewart
Research Scientist, Department of Urban Studies and Planning
Thesis Supervisor

Accepted by: Jinhua Zhao
Associate Professor of City and Transportation Planning
Chair, Transportation Education Committee

Understanding Bus Operations Using High-Resolution Vehicle Location Data

by

Yuzhu Huang

Submitted to the Department of Urban Studies and Planning
on May 12, 2023, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

High-resolution location data (heartbeat data) of transit fleet vehicles is a newfound data source for many transit agencies. On its surface, the heartbeat data can provide a wealth of information about all operational details of a recorded transit vehicle trip, from its location trajectory to its speed and acceleration profiles. In reality, the heartbeat data is often noisy and recorded at inconsistent frequencies, making it a challenging task for analysts to interpret the data as is. This thesis delves into the task of extracting useful operational information about bus vehicles from heartbeat data. In particular, the thesis focuses on three aspects of how heartbeat data can be used to enable operational analysis of transit routes.

First, a methodology is proposed to convert the raw, timestamped coordinate data into a continuous and smooth vehicle trajectory function of each bus trip. A case study using historical heartbeat data collected from a real-world bus trip is presented to showcase how a complete trajectory combined with the vehicle speed profile could allow for qualitative assessment of bus operations. Then, details are provided on how one can analyze the trajectories of multiple bus trips in aggregate to quantify the different types of delay encountered by bus vehicles, including stop dwell time, signal delay, crossing delay, and congestion delay. Case studies are presented to demonstrate how one can quantify each type of delay for a specific bus route or corridor served by multiple routes. Lastly, a thorough discussion is carried out about how one can conduct observational before-after studies using heartbeat data to draw conclusions about the effectiveness of transit improvement projects. A case study is provided to illustrate how one can evaluate the effectiveness of a stretch of bus-only lane by calculating the travel time savings due to the project.

The technical discussions presented in this thesis provide a solid foundation for conducting in-depth analysis of bus operations using heartbeat data. The methodologies will allow transit analysts to gain better insight into the performance of transit routes and corridors, thus allowing transit agencies to develop more targeted strategies for continuously improving transit services.

Thesis Supervisor: John P. Attanucci
Title: Research Associate, Center for Transportation and Logistics

Thesis Supervisor: Anson F. Stewart
Title: Research Scientist, Department of Urban Studies and Planning

Acknowledgments

First and foremost, I would like to thank the MBTA for providing me the opportunity to conduct research on bus vehicle heartbeat data. My interest in this data was piqued four years ago when I was working on a transit speed and reliability project as a consultant. But at the time, I was intimidated by the size of the data and did not have the technical capability to carry out an in-depth analysis of it. I really want to thank Dr. Anson Stewart for suggesting the research idea of analyzing heartbeat data as the topic for my thesis. If it wasn't for your suggestion, I would have not thought about revisiting this old problem with new solutions.

I also want to thank John Attanucci for your continuous support and guidance throughout the research process. You are always able to offer great insights into how data should be interpreted and presented, and I had learned a great deal from you. I am grateful for Prof. Jinhua Zhao, Prof. Haris Koutsopoulos and Dr. Awad Abdelhalim for your advice on the policy implications and the algorithmic improvement of my methodologies, the outcome of the research has been greatly improved thanks to your suggestions. I want to thank Prof. Cathy Wu for the guidance and inspiration in our exploration of the Braess's Paradox and the strong support for my next adventure.

I would like to thank everyone at the JTL-Transit Lab for the friendship and unwavering encouragement you have given me. I want to especially thank Dan, Devin, Dingyi, and Emma for the camaraderie we have developed over the past two years of graduate studies. I can't imagine having a better cohort of peers than you all.

I want to thank everyone who have supported me and been rooting for me since Day 1 of my journey in transportation. I started working in the transportation industry as a 19-year-old sophomore in college. I will never forget the many Seattle mornings driving to my 5:30 am TMC shift on northbound I-5. As an international student fighting for a place all on my own, I have faced many obstacles and struggled to find opportunities that allow me to learn and grow. Transportation is the one place that offered me one opportunity after another, and gave me a strong sense of

belonging and achievement. I am forever grateful for Nirni, Maan, Sayuri, Tony, Craig, Mark, Terry, Erica, Matt, Kate, Liz, Wendy, Pete, Glenn, Kevin, and so many others who are both extraordinary scientists or engineers and, above all, the kindest people. I have been very fortunate to have met all of you in my time as a civil engineering student and in my young career as a traffic engineer. Thank you for the opportunities and mentorship you have given me and for the immense trust that you had in my success.

Lastly, I want to thank my parents and my family, as well as many close friends, especially Jingyi and Suning, who have supported me unconditionally and encouraged me to pursue my dreams. I would not have been where I am today without you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

As more sensors and devices are installed on transit vehicles for monitoring various aspects of transit fleet operations, data with increasing quality and granularity becomes available to transit agencies. One recent data source addition is the high-resolution GPS data of vehicle locations, often called "second-by-second data" (referred to as "heartbeat data" hereinafter), that records the historical location and other metadata of each transit vehicle at almost every second.

The heartbeat data by nature resembles the Automated Vehicle Location (AVL) data that is widely used by transit analysts in the evaluative studies of the operational performance of transit routes. The nature of the AVL data is such that it records the times a transit vehicle enters and departs from bus stops. As a result, the types of assessment a transit analyst can conduct using AVL data are mainly on the exact behaviors of vehicles while dwelling at bus stops, such as bus dwell time, door-open time and door-closed dwell time, as well as the average behaviors of vehicles while traveling between bus stops, such as average running time and speed between stops.

While analysts can use these metrics obtained from AVL data to identify areas with reduced performance, it is difficult to determine what exactly contributes to the reduced performance in each area. As an example, an analyst may find that the average speed of buses is slower within one particular stop-to-stop segment than in

others, thus deciding that the segment is an area of concern and requires improvement. However, the low average speed between two bus stops could be caused by slow-moving traffic due to congestion, or by stopping delays incurred by traffic signals. The average speed information alone offered by the AVL data is not enough to determine which one of the two is the primary source of delay, thus making it challenging to pinpoint the most effective transit improvement strategy.

In comparison, the heartbeat data offers timestamps of vehicles not only at bus stops, but also at locations in between them. Therefore, the heartbeat data contains a richer set of data than the traditional stop-level AVL data and allows analysts to understand the exact behaviors of vehicles at any location along its route. Such detailed information offers the unique opportunity to uncover the interactions between bus vehicles and other road infrastructure and points of conflict besides bus stops that could impede bus movement such as traffic signals, pedestrian crossings, etc.

This thesis provides an in-depth look into how one can leverage the wealth of information contained in the heartbeat data to conduct detailed operational analysis of transit vehicles, buses in particular, that is not possible using AVL data. The raw heartbeat data often comes with noise and is presented in the form of coordinates and timestamps, thus a method is proposed to convert the timestamped coordinates to data points on a easy-to-understand time-space diagram. To account for the noise and inconsistent time gaps inherent in the heartbeat data, data smoothing algorithms are explored to convert the discrete time-space data into continuous trajectories, which then allow for the extraction of the position and speed of the bus vehicle at any point in time into its trip. Then, using the reconstructed bus trajectories, a method is proposed to decompose the observed bus travel time into free-flow travel time and different categories of delays. Finally, the quantified delays of bus vehicles are used to conduct before-after studies of transit improvement projects in order to evaluate the effectiveness of the strategies.

Throughout the thesis, case studies are provided following the discussions of methodologies using real-world heartbeat data to demonstrate the practicality and applicability of this research. A thorough literature review is conducted to illustrate

the contribution of each component of this thesis.

1.2 Research Objectives

This research aims to accomplish the following goals:

1. Define the characteristics of an ideal bus trajectory, and identify methods to convert discrete timestamped bus coordinates to continuous bus trajectories.

2. Propose a systematic framework to decompose the total bus travel time into different components, including free-flow travel time, dwell time, signal delay, congestion delay and loss time.

3. Illustrate how different sources of delays can be used to evaluate the effectiveness of transit improvement strategies, such as transit signal priority and bus-only lane projects.

1.3 Research Approach

To meet the above research objectives, various data processing, smoothing, probabilistic and statistical techniques are used. A separate chapter is dedicated to the discussion of each research objective, where the proposed methodologies are discussed in detail, followed by case studies carried out using real-world data. The focus of the discussion around methodologies is to build the technical foundation and provide a blueprint for working with heartbeat data and the products derived from it, with the goal in mind that the methodologies are reproducible by any analyst and able to be used in working with similar data from any transit agency. The purpose of the case studies, on the other hand, serve as demonstrations for how the methodologies can be applied to real-world data, and offer insight into the type of output that can be expected.

The analysis of heartbeat data described in this thesis focuses on bus vehicle rather than rail vehicles. Although heartbeat data offers an incredible amount of information

for rail operations just as it does for bus operations, the significant differences in the nature of rail and bus operations means that many assumptions and methodologies made in this thesis about bus operations cannot be directly applied to rail. However, with some adjustments to the basic assumptions to fit in the context of rail vehicles, the concepts presented in this thesis can be modified for rail operations as well.

1.4 Data Sources

1.4.1 Heartbeat Data

Data Content

For the task of reconstructing vehicle trajectories, the most relevant fields in the heartbeat data are the ones that describe the time and location variations of the bus vehicles. One standard data protocol that serves such purpose is GTFS-RT, which offers specifications for tables including `VehiclePositions` that is designed for reporting vehicle location information and the timestamp for when the location is measured [1]. Heartbeat data, such as archived GTFS-RT data[1], often provides historical records of the location of every bus vehicle as frequent as every a few seconds.

For the purpose of reconstructing a complete and continuous vehicle trajectory from heartbeat data, the `latitude, longitude` and `timestamp` columns will suffice, as long as the data is recorded at roughly every 6 seconds, which is the frequency that have been empirically tested by the author. In addition to storing vehicle positions data, GTFS-RT also provides helpful metadata of each vehicle trip such as `trip_id`, which is useful if one hopes to join the vehicle trajectory with the location of bus stops to observe vehicle behavior around such infrastructure, or to analyze the performance

[1]More specifically, the heartbeat data used by the author are archived snapshots of MBTA's GTFS-RT data taken every second from the public-facing API (`https://cdn.mbta.com/realtime/VehiclePositions.json`). To the author's knowledge, the public-facing GTFS-RT data of the MBTA is created by combining vehicle location data recorded by the Samsara tracking device (`https://www.samsara.com/pdf/docs/MBTA_Case_Study.pdf`) and trip information recorded in the CAD/AVL system.

of the trip against schedule. In these cases, the data field `trip_id` can be used to query information about the bus trip, such as its route number, stop IDs, stop locations, etc. In other extended applications of heartbeat data where the `bearing` data is proven valid and of high quality, the bearing information can also provide insight into the operational impact of vehicle maneuvers such as moving in and out of bus lanes, although the use of `bearing` data is not the focus of the discussion in this thesis.

A snippet of an example object stored in the GTFS-RT file published by the Massachusetts Bay Transportation Authority (MBTA) is shown in Figure 1-1a.

Note that it is not necessary for the heartbeat data to be stored in the GTFS-RT format for the methodologies proposed in this thesis to work, as any data source that can provide high-resolution (roughly 6 seconds per record) timestamped location data would work in reconstructing vehicle trajectories discussed in Chapter 2, and any data that additionally contains the `trip_id` information, matching with those in the static GTFS feed, would work in the delay analysis of transit routes or transit corridors presented in Chapter 3.

Data Frequency

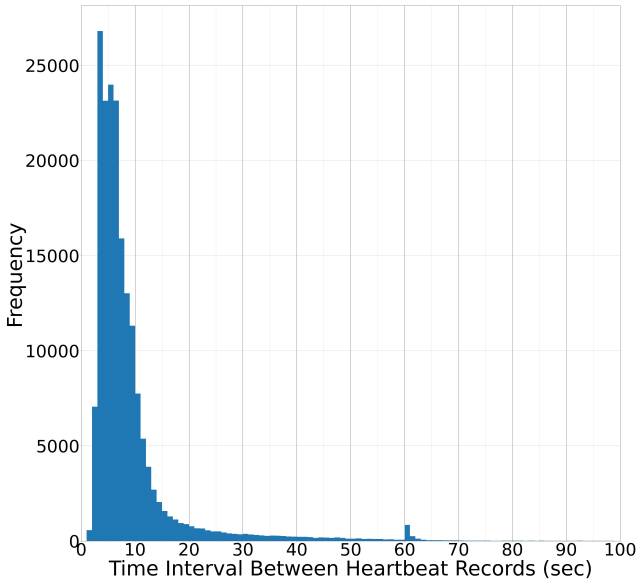
The heartbeat data recorded for each trip provides a snapshot in time of the location of the transit vehicle. However, such snapshots are not recorded at a constant frequency. An examination of nearly 900 outbound trips by MBTA's Route 1, which amount to around 185,000 heartbeat data records, shows that majority of data is recorded in intervals shorter than 10 seconds with a median frequency of 6 seconds, mode of 3 seconds and mean of 9 seconds. A histogram of the frequency of these GTFS-RT records is presented in Figure 1-1b,

1.4.2 Static GTFS Data

If the heartbeat data provided by the public agency contains information about each vehicle trip, e.g. the `trip_id` column, then one can query additional information such as the location of bus stops or the scheduled arrival time at each bus stop to

```
{
    "id": "1651130687_1915",
    "is_deleted": false,
    "vehicle": {
        "block_id": "C01-1",
        "capacity": 36,
        "current_stop_sequence": null,
        "load": null,
        "location_source": "samsara",
        "occupancy_percentage": null,
        "occupancy_status": null,
        "operator": {
            "first_name":
            "id":
            "last_name":
            "logon_time":
            "name":
        },
        "position": {
            "bearing": 10,
            "latitude": 42.280583999,
            "longitude": -71.09266201,
            "speed": 13.6294
        },
        "revenue": true,
        "run_id": "123-1001",
        "stop_id": null,
        "timestamp": 1651130687,
        "trip": {
            "overload_id": null,
            "overload_offset": null,
            "route_id": "28",
            "schedule_relationship": "SCHEDULED",
            "start_date": null,
            "trip_id": "51090957"
        },
        "vehicle": {
            "id": "y1915",
            "label": "1915"
        }
    }
},
```


(a) Sample entity object from the GTFS-RT Vehicle Positions JSON file provided by the MBTA.



(b) Histogram of data frequency of nearly 900 outbound Route 1 trips recorded in MBTA's GTFS-RT data.

Figure 1-1: Sample heartbeat data object and data frequency.

conduct more in-depth operational analysis.

Data Content

Provided that the given heartbeat data contains the `trip_id` column, then a few join operations are needed in order to query the corresponding `stop_id`s and stop coordinates corresponding to each trip. As illustrated in Figure 1-2, the `stop_times` table in the static GTFS feed contains the stop-level trip information that can serve as the basis in connecting the `trip_id` field with `stop_id`. The result of the join operation is one table that contains all columns listed in Figure 1-2. Among these columns, `stop_lat` and `stop_lon` provide the coordinates of bus stops along the pattern of the bus route that the trip is serving. These bus stops, along with traffic signals and pedestrian crossings can be matched to the same time-space reference frame as the vehicle trajectory, allowing analysts to dive into analyzing the behavior of

buses near various infrastructure. The `route_id` and `direction_id` columns allow for easy labeling of heartbeat data by route and by direction. The `arrival_time` column could be used to compare with the actual arrival time observed from heartbeat data to understand on-time performance of bus trips.

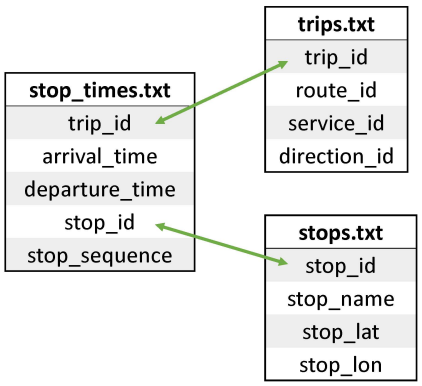


Figure 1-2: Static GTFS table and fields used to supplement heartbeat data.

1.4.3 Data Flow Diagram

The data flow diagram is illustrated in Figure 1-3. In summary, the heartbeat data alone can be used to construct vehicle trajectories. The vehicle trajectories combined with static GTFS data can then be used to conduct delay analysis to quantify the various types of delays and conduct before-after studies to understand the effectiveness of transit improvement strategies.

1.5 Literature Review

1.5.1 Trajectory Smoothing Techniques

Researchers have looked into analyzing transit performance from the trajectory of vehicles. Hall and Vyas analyzed bus vehicle trajectories obtained from location tracking devices and evaluated bus speed by dividing the length of road segments

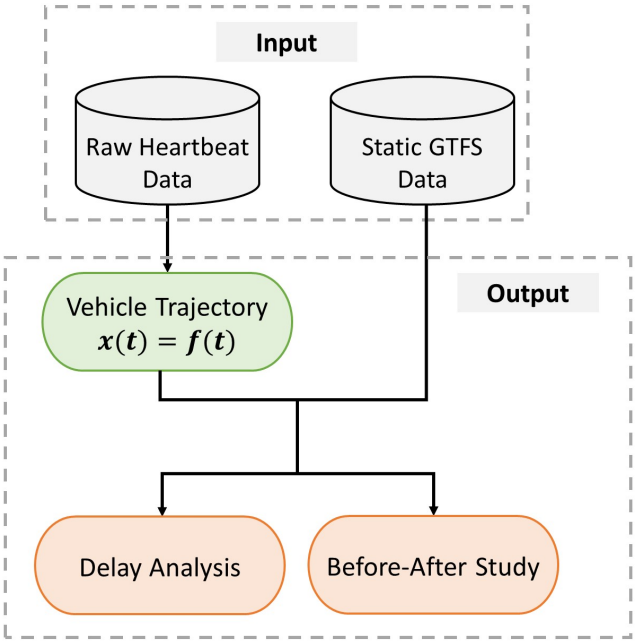


Figure 1-3: Data flow diagram.

between stops by the travel time between stops without dwell time, and compared the speed with automobile speed [2]. Cathey and Deiley used the Kalman Filter to construct trajectories of transit vehicles traveling on freeways and compared the speeds obtained from the trajectory to those calculated from loop detector data [3], and concluded that granular AVL data gathered from bus vehicles can serve as an additional data source for freeway performance monitoring. However, the transit location data used is collected every 20 seconds, and the researchers did not mention the monotonicity of the trajectory.

To the author's best knowledge, there has not been previous research that attempted to extract operational measures of transit vehicles by directly reconstructing a continuous, smooth, and monotonic transit vehicle trajectory from high-resolution location data. This study proposes a trajectory-building algorithm for bus vehicles to address this gap in the literature.

1.5.2 Transit Delay Analysis

From a review of the literature, the author finds that many researchers who have explored using heartbeat data (or similar but lower frequency data) have focused on decomposing the vehicle travel time into different categories by bus state.

Colghlan et al. categorized the total bus travel time by bus state into signal delay, queueing delay, boarding/alighting time, and free-flow travel time [4]. They assumed that signal delay is any delay from within 80 ft of a traffic signal, did not account for vehicle acceleration and deceleration, and used the decomposed travel time of each bus state as input to a regression model that predicts stop-to-stop travel time. Lind and Reid analyzed heartbeat data by decomposing the passenger in-vehicle travel time into time in-motion and time stopped, and used the result to calculate the total trip time budget [5]. Aemmer et al. used the GTFS-RT data and proposed a method to decompose the delay of transit vehicles into systematic and stochastic delays derived from comparing the linear speed to the 95th percentile speed in each segment, and used the results to evaluate transit reliability and efficiency [6].

Although these previous studies looked at using heartbeat data to analyze bus operations, researchers had mainly focused on extracting particular types of bus delay without examining all possible components of bus delays holistically. Furthermore, there is no research involving heartbeat data that aggregates the performance of multiple transit routes on the same corridor.

1.5.3 Effectiveness of Transit Improvement Strategies

In practice, analysts usually evaluate the effectiveness of transit improvement strategies by taking the difference between the analyzed metric observed before the installation of the treatment and that after the treatment, such as the study commissioned by the City of Cambridge in evaluating the effectiveness of a bus-lane project [7]. In research literature, such methodology still prevails as researchers often examine the effectiveness of a transit treatment by comparing specific operational metrics before and after the treatment. Sakamoto et al. [8] evaluated the effectiveness of a bus

priority lane in Shizuoka City, Japan by comparing the absolute difference in queue length, jam length, travel time and dwell time of buses in the study corridor before and after the treatment. Kim [9] drew conclusions about the performance of bus lanes in Seoul, South Korea by calculating the change in bus and car speed measured along the treated road segment before and after the project. The method of directly comparing the observed metric before and after the treatment, referred to in the remainder of the thesis as the naive method, naively assume that the metrics observed in the "before" period are perfectly representative of what the metrics would have been in the "after" period had there been no treatment. This assumption, however, can be challenged because factors other than the treatment could have impacted the unobserved, counterfactual travel time in the "after" period as well.

A field of research that had encountered the same challenge but had overcome it with more statistically-sound methodologies is the field of road safety. In particular, Hauer [10] presented several frameworks for evaluating the effectiveness of road safety projects, and the methodologies have since been widely adopted by transportation agencies. Therefore, the author attempts to adapt the naive method and the comparison group method explored by Hauer to fit in the context of transit improvement project, in the hope of overcoming the simplistic assumption inherent in the naive method.

1.6 Thesis Organization

This thesis is organized such that each chapter focuses on the discussion of one research objective as listed in Section 1.2. Chapter 1 introduces the background and motivation underlying the thesis, provides an overview of the thesis structure, and presents a literature review concerning each research topic covered in the thesis. Chapter 2 presents the methodologies concerning reconstructing a complete and continuous vehicle trajectory from raw heartbeat data. Chapter 3 dives into the task of categorizing and quantifying the magnitude of different types of delays. Chapter 4 provides an in-depth discussion about how one can carry out observational before-after stud-

ies to evaluate the effectiveness of transit improvement strategies. In each chapter, thorough discussions about the proposed methodologies are provided to serve as the technical foundation for practical applications, which are demonstrated through various case studies carried out using real-world data collected from public agencies. Chapter 5 summarizes the contributions and implications of the research and highlights opportunities to expand the study of heartbeat data beyond this thesis.

Chapter 2

Reconstructing Transit Vehicle Trajectory

2.1 Introduction and Motivation

At first glance, the heartbeat data seems to provide information about the exact location of each vehicle in such detail that it should be able to tell everything an analyst would wish to know about a vehicle's trip. However, a closer examination of the data, an example of which is visualized in Figure 2-1a, illustrates that the coordinate data can be noisy and vehicles may appear to straddle the road network, and the data recorded can have inconsistent frequency and result in incomplete location information within time intervals, making it difficult to directly read the exact location and speed of the vehicle from the heartbeat data.

The task of extracting operational information from the heartbeat data could be easier if one could reconstruct a complete transit vehicle trajectory. Note that the notion "trajectory" refers to a continuous, not discrete, timestamped location data of the transit vehicle consisting of two fields - time into trip, T, and distance into trip, D. This chapter discusses the data smoothing methodology an analyst can use to convert the noisy, discrete, and inconsistent raw heartbeat data to a continuous, smooth, and monotonic vehicle trajectory. Such a trajectory would provide reliable information about the exact location, speed, and acceleration of the transit vehicle at

any time and distance into the trip, which would be valuable to performance analysts and decision makers when trying to understand how transit vehicles interact with the built environment at a very granular level.

To accomplish the goal of reconstructing bus vehicle trajectories from raw heartbeat data, the definition of what an "ideal" trajectory should be is provided. Then, a map-matching process is introduced that can be used to snap each recorded coordinate to the road segments that the bus traverses. With each coordinate matched to a specific position on a specific road segment, the series of coordinate data can then be turned into the more analytically convenient "distance into trip" data, each of which has a corresponding "time into trip" information, together forming the discrete time-distance data series. Finally, a smoothing algorithm is proposed to convert the discrete time-distance data into a continuous data series that satisfies the characteristics of an ideal vehicle trajectory.

2.2 Properties of An Ideal Bus Trajectory

Since the heartbeat data records vehicle location at a high frequency, it provides a large volume of information that could be used to infer measures other than the location of the vehicle, such as vehicle speed and acceleration. To reconstruct a continuous transit vehicle trajectory from discrete timestamped location data and to be able to retrieve the speed and acceleration of the vehicle at any point in time, a series of data smoothing processing steps are needed. It is worth noting that the objective of data smoothing is not to obtain the "smoothest" curve possible, but rather a trajectory that resembles the real-world behavior of the bus vehicle as realistically as possible. Therefore, a few properties must be considered for the result to be representative of an actual transit vehicle trip. These properties include:

- The trajectory should be non-decreasing, i.e. the distance into the trip at a later time should not be smaller than the distance into the trip at a previous time.

- The trajectory is made up of composite cubic polynomials. The position of the vehicle is a result of control inputs and follows vehicle kinematics, and as Nagy et al. studied, cubic polynomials are the lowest order curves that can be used to generate the trajectory of car-like robots [11].
- The trajectory function should be continuous and at least once differentiable so that the speed profiles are also continuous and can be easily retrieved from the function.

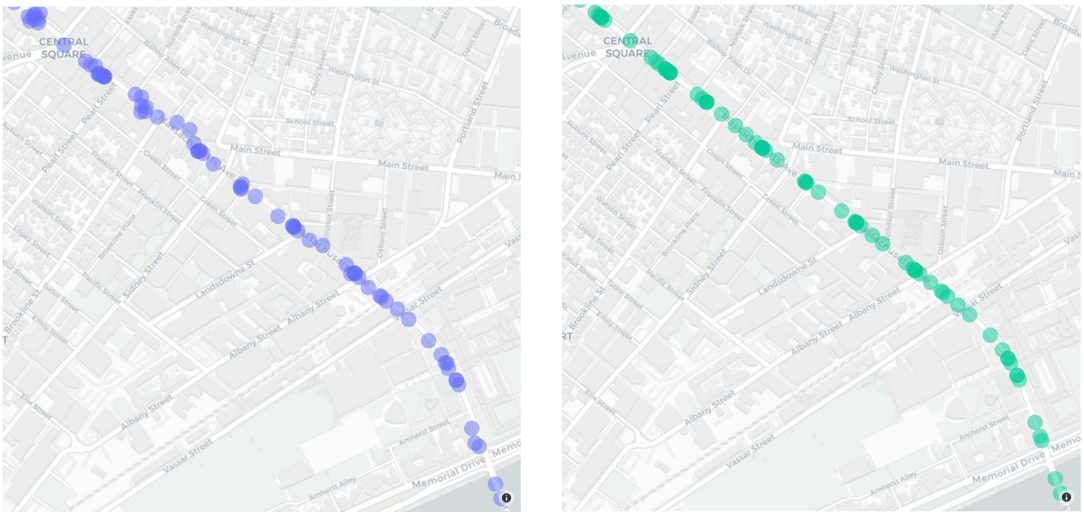
2.3 Processing of Raw Coordinates

2.3.1 Map Matching

The bus vehicle locations recorded in the raw heartbeat data are subject to noise and inaccuracy, which can possibly be attributed to the limited ability of GPS technology in providing accurate position of moving objects in dense urban environments [12]. This results in coordinates that are not accurately located within the actual road segment. To mitigate the issue with noisy coordinates, a map-matching process is needed to "snap" each coordinate point to the closest road segment to restore the location information as much as possible. One map-matching tool that is easy to use and freely available is the Valhalla map engine, which offers the "trace attributes" service that allows the client to post a request with a series of raw coordinate data and retrieve a corresponding series of coordinates snapped to the OpenStreetMap road network while guaranteeing the geographic order of the data points, i.e. data points are kept in an upstream-to-downstream order [13, 14]. The result of processing raw heartbeat data through the map-matching service is shown in Figure 2-1b, where it can be seen that the processed coordinates are realigned with the road network.

2.3.2 Obtaining Time-Distance Data

The map-matching process infers the most probable location along the road network for each raw coordinate, and returns the ID of the road segment and the position



(a) Raw heartbeat data points. (b) Map-matched heartbeat data points.

Figure 2-1: Raw and map-matched data points plotted on OpenStreetMap.

along the segment of each matched coordinate. With this information, each recorded coordinate of the bus can be converted to a distance value relative to the starting location of the trip.

Suppose there are n raw timestamped coordinate data points recorded for a bus trip. For each record i, where $i = 1, 2, ..., n$, the timestamp at which the data is recorded is ts_i and the raw coordinate is C_i. Denote the time into trip of the ith data point by t_i, then

$$t_i = ts_i - ts_1. \tag{2.1}$$

As shown in Table 2.1, for each raw coordinate C_i, the map-matching process returns the matched coordinate, M_i, the ID of the road segment that the matched coordinate belongs to, r_i, as well as its normalized position along the segment, p_i. The normalized position value p_i is given so that $0 \leq p_i \leq 1$. $p_i = 0$ means the coordinate M_i is at the beginning of the segment r_i, while $p_i = 1$ means the end. In addition, the process provides a road segment lookup table which contains the length information L_j of each segment R_j as shown in Table 2.2.

Denote the index of road segment r_i by id_i (i.e. $r_i = R_{id_i}$), then the distance into trip of the ith coordinate by d_i is

$$d_i = \sum_{j=1}^{id_i-1} L_j + L_{id_i} * p_i - L_{id_1} * p_1, \tag{2.2}$$

where L_j is the length of road segment R_j. Equation 2.2 states that the distance into trip of the ith coordinate is equal to the sum of the lengths of all road segments upstream of that of the ith point plus the distance into the segment that it lies within, shifted with respect to the first data point.

Table 2.1: An example map-matched coordinate table.

i	TS	C	M	r	p
0	2022-04-25 08:24:45	(42.372642, -71.119048)	(42.372660, -71.119108)	R_0	0.639
1	2022-04-25 08:24:50	(42.372365, -71.119241)	(42.372373, -71.119267)	R_0	0.940
2	2022-04-25 08:24:57	(42.372246, -71.119324)	(42.372252, -71.119319)	R_2	0.463
3	2022-04-25 08:25:00	(42.372206, -71.119230)	(42.372215, -71.119224)	R_3	0.116
4	2022-04-25 08:25:04	(42.372129, -71.119030)	(42.372139, -71.119023)	R_3	0.452
...					

Table 2.2: An example of a road segment lookup table.

j	R	L (m)
0	R_0	67
1	R_1	7
2	R_2	5
3	R_3	55
...		

Combining the time into trip and distance into trip information calculated using Equations 2.1 and 2.2, a discrete series of time-distance data can be obtained from the map-matched coordinates and is shown in Table 2.3.

Table 2.3: An example of time-distance data converted from map-matched coordinates.

i	$T(sec)$	$D(m)$
0	0	0
1	5	20.167
2	12	33.502
3	15	42.567
4	19	61.047

2.4 Trajectory Smoothing

In this section, four alternative methods for bus trajectory smoothing are discussed, and the advantages and disadvantages of each are presented.

2.4.1 Linear Interpolation

The simplest method to construct a continuous curve on the time-space diagram using the discrete time-distance data obtained from the previous steps is through linear interpolation, or equivalently connecting adjacent data points using line segments (LSEG).

Algorithm 1 Linear Interpolation (LSEG)

Input: an $1 \times N$ vector T of timestamps $t_1 < t_2 < ... < t_n$, an $1 \times n$ vector D of distances $d_1 \leq d_2 \leq ... \leq d_N$ for $i = 1, ..., n$. Denote the continuous trajecto function to be constructed by $f(t)$, then $f(t_i) = d_i$.

1: **for** $i = 1, ..., n-1$ **do**
2: $\quad f(t) = \frac{d_{i+1}-d_i}{t_{i+1}-t_i} * t + d_i$
3: **end for**
4: return $f(t)$

An example snippet of the trajectory obtained from linear interpolation is shown in Figure 2-2. Considering the three properties of an ideal bus vehicle trajectory introduced above, it can be oberved that although the trajectory is continuous and monotonic, it is not smooth and therefore is not differentiable. Intuitively, the consequence of such drawback is that the speed profile calculated cannot be guaranteed to be continuous.

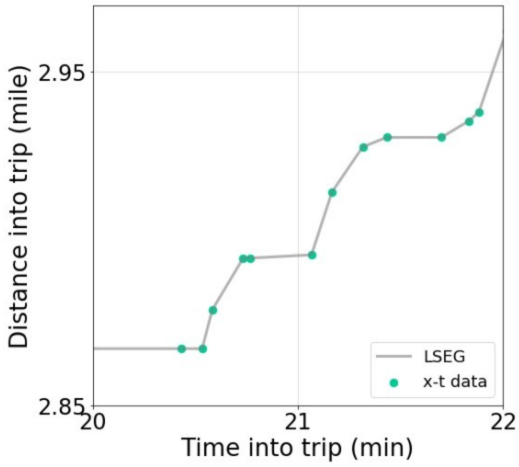


Figure 2-2: Trajectory constructed from Linear interpolation.

2.4.2 Polynomial Cubic Interpolation

To overcome the non-differentiability issue presented by linear interpolation, the Piecewise Cubic Hermite Interpolant (PCHIP) algorithm developed by Fritsch et al. is explored [15]. The PCHIP algorithm connects adjacent data points in T and D with a cubic polynomial function and guarantees that the trajectory is smooth by enforcing that the first derivative evaluated at each data point is equal for the connecting polynomial functions to the left and right of the data point [16]. Therefore, the algorithm maintains the cubic polynomial and monotonic properties of all sections of the trajectory.

An example snippet of the trajectory obtained from PCHIP is shown in Figure 2-3. It's worth noting that although the PCHIP algorithm guarantees a continuous first derivative, it does not guarantee a continuous second derivative, meaning that a continuous speed profile can be retrieved from the PCHIP trajectory, but not a continuous acceleration profile.

2.4.3 Local Regression

Although the PCHIP algorithm allows for the construction of a trajectory that satisfies all three properties of an ideal trajectory, one of the fundamental assumptions that has to be made is that all discrete distance into trip data points that serve as

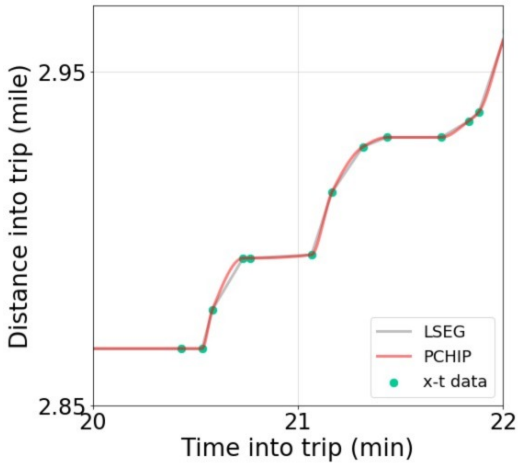


Figure 2-3: Trajectory constructed from PCHIP.

the input to the algorithm are the true distances, and that there is no error associated with each distance value. Such an assumption may not be valid, however, if the measurement error of the onboard device is taken into consideration.

A local regression process (LOCREG) can be used to estimate the true distance into trip of the bus vehicle at each recorded timestamp by weighing the importance of nearby data points based on their distance from the timestamp of interest, consequently smoothing out the transit vehicle trajectory [17].

Suppose the true distance of the vehicle at each time point t_i is x_i such that

$$x_i = d_i + \varepsilon_i, \tag{2.3}$$

where d_i is the measured distance at time t_i, and ε_i is the measurement error for d_i. The objective of the local regression algorithm is to find a function $f : t \to x$ such that it solves the following minimization problem at each data point (t_i, d_i):

$$\min \sum_{j=1}^{n} w_{i,j}(d_i - f(t_i))^2, \tag{2.4}$$

where the value of each weight term $w_{i,j}$ is determined by the selected bandwidth and kernel function. In this study, the tricube kernel is used in the local regression algorithm due to its computational efficiency and its low sensitivity to outliers which

could be present in the time-distance data because of the difficulty faced by GPS devices in locating bus vehicles in an urban setting. Since the average data frequency of heartbeat data is smaller than 10 seconds per record, a bandwidth of 20 data points are selected empirically, following the reasoning that the true distance is informed by other distance values measured one minute before or after the time point for which the distance value is estimated.

An example snippet of the trajectory obtained from local regression is shown in Figure 2-4. LOCREG is superior to LSEG and PCHIP in that it produces a continuous trajectory that takes into account the measurement error of each measured distance, but it provides little guarantee for differentiability or monotonicity. This can result in negative speed values when taking the first derivative of the LOCREG trajectory, and also present challenge in deriving speed and acceleration profiles.

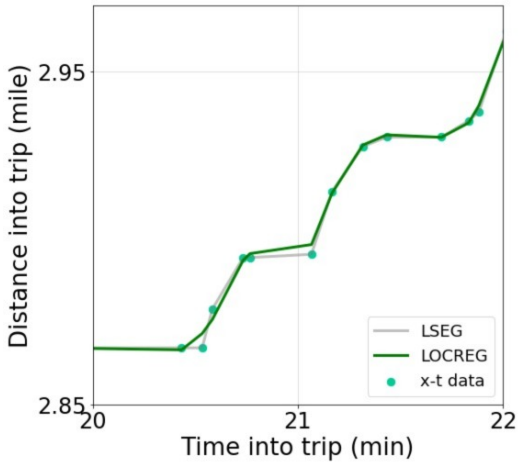


Figure 2-4: Trajectory constructed from Local Regression.

2.4.4 Interpolation and Regression

To address the issue with monotonicity from using local regression, an exploratory algorithm is proposed to remove the non-monotonic sections of the LOCREG data, and fill in the now missing intervals by passing the remaining data through a monotonic interpolation algorithm (e.g. PCHIP).

Combining the LOCREG algorithm with PCHIP, a complete workflow that starts

from data smoothing and ends with filling in data to obtain a smooth and monotonic trajectory is detailed in Algorithm 2.

Algorithm 2 LOCREG-PCHIP

Input: an $1 \times n$ vector T of time into trip values $t_1 < t_2 < ... < t_n$, an $1 \times n$ vector D of distance into trip values $d_1 \leq d_2 \leq ... \leq d_n$ for $i = 1, ..., n$ and $n > 2$.

X = []
$f_{locreg} = LOCREG(T, D)$
for i = 1, 2, ..., n **do**
    $x_i = f_{locreg}(t_i)$
    if $i > 1$ and $x_i < x_{i-1}$ **then**
        $x_i = x_{i-1}$
    end if
    $X.append(x_i)$
end for
$f = PCHIP(T, X)$
return f

An example snippet of the trajectory obtained from the LOCREG-PCHIP algorithm is shown in Figure 2-5. The trajectory obtained using LOCREG-PCHIP preserves the true distance values estimated from local regression at all time points except for those where the monotonicity principle is violated, in which case the distance value is chosen to be equal to the largest distance value observed prior the said time points. Since the algorithm ends with constructing a PCHIP function using the modified distance values, the resultant trajectory function is both monotonic and differentiable.

2.5 Goodness of Algorithm Evaluation

A way to quantify the goodness of the trajectory smoothing algorithm is desirable. The challenge with this task is that the ground truth, i.e. the actual historical location, speed and acceleration of the transit vehicle, is not a data source that could be easily obtained. There are a few limited strategies, however, that can be used to evaluate the quality of the speed and acceleration profiles. The following sections describe the details of these strategies and offer some discussion regarding reasons for

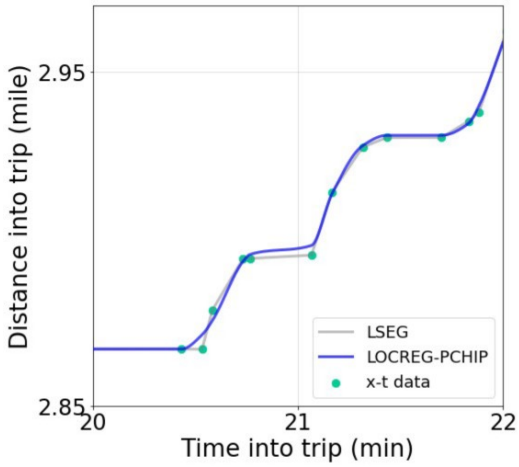


Figure 2-5: Trajectory constructed from LOCREG-PCHIP.

choosing one algorithm over others.

2.5.1 Validation of the Speed Profile

Since the AVL data provides historical information regarding the bus stopping and door open activities at bus stops, one way to validate the accuracy of the speed profile is to take advantage of the stop-level AVL data and check if the speed from the calculated trajectory is indeed zero when the transit vehicle was recorded as having doors open at a bus stop. The speed profiles obtained from all methods mentioned above (LSEG, LOCREG, LOCREG-PCHIP, PCHIP) can be displayed over color bands of door-open intervals recorded in the stop-level AVL data collected for the same bus trip as shown in Figure 2-6.

To measure how well the speed trajectory aligns with the AVL stop events, we calculate among all integer seconds within AVL door-open intervals how many of these seconds correspond to a zero speed on the speed profile. The notion of "zero speed" corresponds to the speed of the vehicle with its doors open while dwelling at a bus stop, which intuitively should be 0 mph. However, the 0-mph threshold could be relaxed to a speed below which a vehicle is considered not traveling to account for GPS data instability. As summarized in Table 2.4, if the "stop speed" is defined as below 5 mph, the speed at over 90% of AVL door-open integer timestamps are

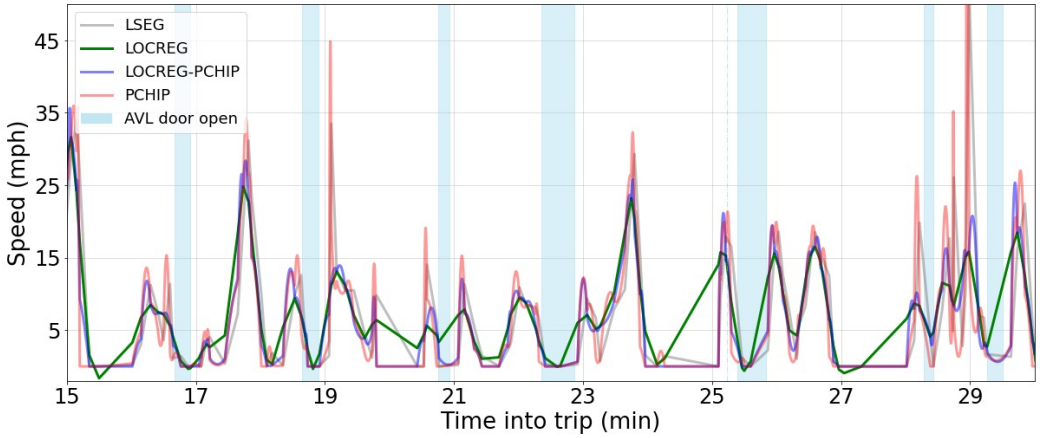


Figure 2-6: Comparison of all speed profiles.

correctly captured using the LOCREG, LOCREG-PCHIP and PCHIP algorithms. A more intuitive plot comparing the percentage of stop speed correctly captured by each algorithm is shown in Figure 2-7, where it shows that the performance of the LOCREG and LOCREG-PCHIP algorithms are almost identical to each other and very close to that of PCHIP at stop-speed threshold above 0.5 mph.

Table 2.4: Percentage of AVL door-open integer timestamps at which speeds are correctly captured by each algorithm.

Algorithm	Stop-Speed Threshold (mph)		
	0	3	5
LSEG	0	77	86
LOCREG	7	77	100
LOCREG-PCHIP	0	84	92
PCHIP	0	92	98

2.5.2 Validation of the Acceleration Profile

Another way to check for the goodness of the algorithm is to examine the percentage of accelerations that are beyond a reasonable threshold. Based on the research from Kirchner et al., the maximum bus vehicle acceleration is $0.17g = 3.7$ mphps and maximum deceleration is $-0.24g = -5.3$ mphps [18]. As indicated by the acceleration profiles shown in Figure 2-8, LOCREG-PCHIP produces a more reasonable

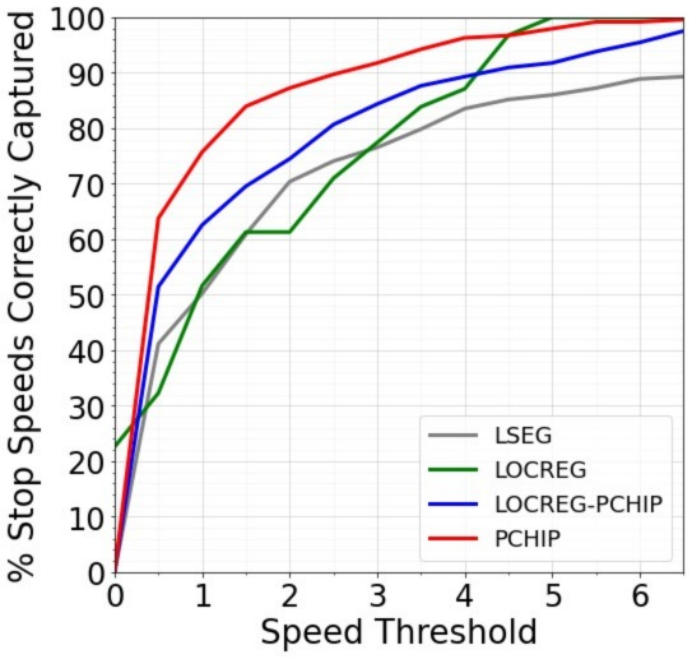


Figure 2-7: Percentage of AVL door-open integer timestamps at which speeds are correctly captured vs. the "stop speed" threshold.

acceleration profile than all other methods. This conclusion is verified by the percentage of unreasonable accelerations summarized in Table 2.5, which shows that only 1.3% of the accelerations from LOCREG-PCHIP are beyond the reasonable threshold, whereas 5.3% of accelerations from PCHIP are unreasonable.

Table 2.5: Percentage of trajectory data of which accelerations are beyond the $[-5.3, 3.7]$ mphps threshold.

Algorithm	% unreasonable acceleration
LSEG	53.6
LOCREG	0.0
LOCREG-PCHIP	1.3
PCHIP	5.3

Given the discussion above, the trajectory produced by the LOCREG-PCHIP algorithm is the best trajectory overall, since it preserves all three necessary properties of an ideal transit vehicle trajectory, while performing well when validating against stop-level AVL data and reasonable acceleration thresholds.

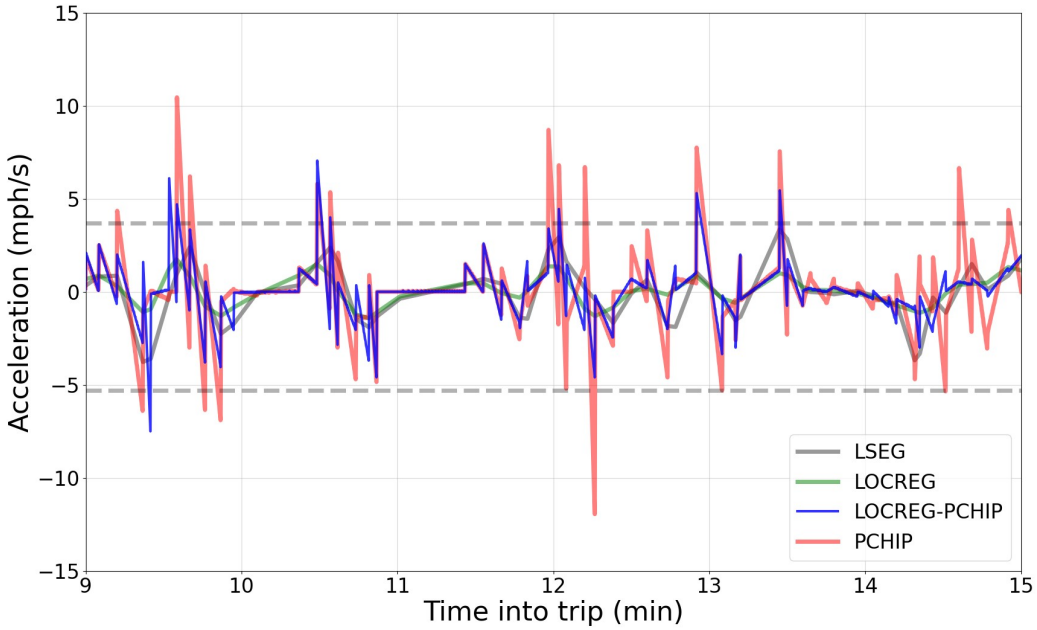


Figure 2-8: Comparison of all acceleration profiles.

2.5.3 Selection of the Best Algorithm

The "best algorithm" should produce trajectories not only satisfy all three characteristics of an ideal trajectory, but also perform reasonably well in the evaluation of speed and acceleration profiles. As shown in Table 2.6, the algorithms that satisfy these criteria are LOCREG-PCHIP and PCHIP.

Table 2.6: Evaluation of algorithms.

Algorithm	Evaluation Criteria					
	MONO[1]	CUBIC[2]	DIFF[3]	AVL[4] (%)	ACC[5] (%)	Best
LSEG	✓	✗	✗	86	46	
LOCREG	✗	✗	✓	100	100	
LOCREG-PCHIP	✓	✓	✓	92	99	✓
PCHIP	✓	✓	✓	98	95	

[1] MONO: the trajectory is non-decreasing
[2] CUBIC: the trajectory is made up of cubic polynomials
[3] DIFF: the trajectory is once differentiable
[4] AVL (%): the percentage of AVL dwell activities correctly captured
[5] ACC (%): the percentage of accelerations within a reasonable threshold

Comparing the LOCREG-PCHIP algorithm with PCHIP, LOCREG-PCHIP is

preferable because it is able to predict the true distances from observed distances using information provided by adjacent data points by taking advantage of the LOCREG algorithm, rather than trusting each observation a hundred percent. Therefore, LOCREG-PCHIP is used as the basis for reconstructing trajectories in the remainder of the thesis.

2.6 Case Study

To illustrate how the trajectory produced by the LOCREG-PCHIP algorithm can be used for bus operational analysis, a trajectory color-coded by vehicle speed is shown in the time-space diagram in Figure 2-9. Since the trajectory is built upon time-distance data returned by the map-matching process, which also provides information about the OpenStreetMap way segment IDs associated with each data, one can identify the exact way segment and distance along the segment of each data point. Therefore, the exact location of other road infrastructure such as bus stops, traffic signals, and pedestrian crossings with respect to the trajectory can also be identified and plotted in the same figure.

The combination of vehicle trajectory, vehicle speed, locations of bus stops, signals, and crossings, as well as the door-open time recorded in AVL data can provide a wealth of information regarding the bus vehicle operations and lend insight into the maneuver of the bus vehicle throughout its trip.

Several observations of the operations of the sample bus trip are described below to showcase the information that a complete bus vehicle trajectory can provide.

2.6.1 Stop Dwelling Activities

The complete bus vehicle trajectory allows for a straightforward visualization of stopping activities at bus stops. As shown by the sections of the trajectory labeled A in Figure 2-9, the trajectory captures the bus stopping at the far-side bus stop at Massachusetts Avenue & Albany St after it waits at the traffic signal just upstream of the bus stop.

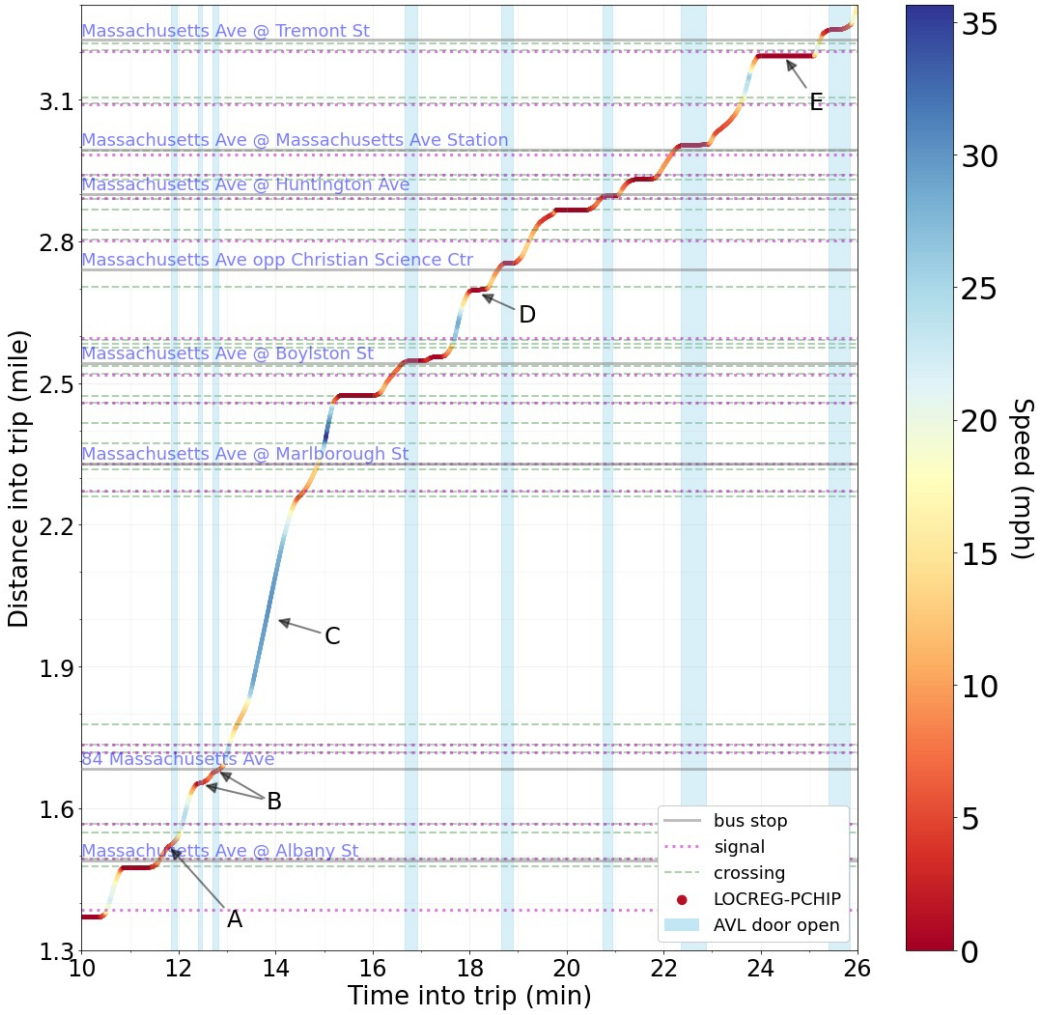


Figure 2-9: LOCREG-PCHIP trajectory used for the operational analysis of a sample weekday PM inbound trip of Route 1 operated by the MBTA.

The sections labeled *B* in Figure 2-9, on the other hand, show that the bus opened its door twice at the bus stop at 84 Massachusetts Avenue, once upstream of the stop and once at the bus stop. Interestingly, the stopping activity upstream of the stop could be due to the Massachusetts State regulation which requires bus drivers to stop and open the door before a railroad track [1]. Both of these stopping activities are validated by the door-open intervals recorded in the AVL data.

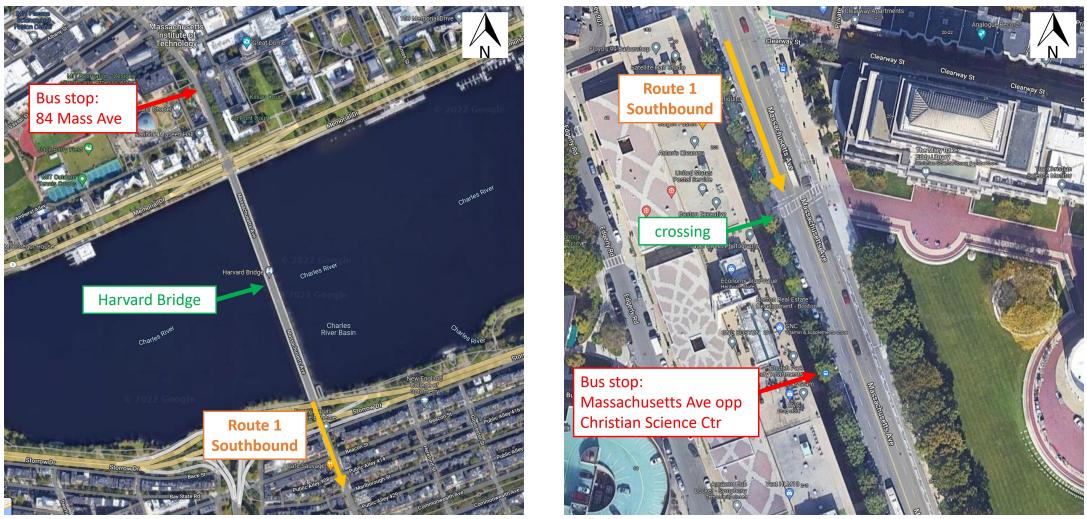
[1]Regulation regarding the operation of motor buses near railroad crossings as required by the Massachusetts State Department of Public Utilities: `https://www.mass.gov/files/220_cmr_155.00_final_8_7_09.pdf`.

2.6.2 Vehicle Speed

The section labeled C in Figure 2-9 shows that the bus traveled at a speed of approximately 25-30 mph 1.9-2.2 miles into the trip, faster than its speed at most other sections of the trip. This is because, within the 1.9-2.2-mile segments of the trip, the bus is traveling on the Harvard Bridge shown in Figure 2-10a, unimpeded by any signal or other source of delay.

2.6.3 Stopping at a Pedestrian Crossing

From the section labeled D in Figure 2-9, one can see that the vehicle stopped just upstream of a pedestrian crossing. An inspection of the satellite view of the road segments near this location as shown in Figure 2-10b, reveals that there is indeed a pedestrian crossing north of the bus stop at Massachusetts Avenue opposite the Christian Science Center. The trajectory shows that the vehicle stopped at the pedestrian crossing before stopping at the bus stop.



(a) Satellite view of the Harvard Bridge.

(b) Satellite view of the road segment near the stop at Mass Ave opp Christian Science Ctr.

Figure 2-10: Satellite images of the road segments associated with sections of the trajectory labeled C and D.

2.6.4 Stopping at a Traffic Signal

The section labeled E in Figure 2-9 shows that the bus stopped at the traffic signal upstream of the far-side stop at Massachusetts Avenue & Tremont St. Besides clearly showing the existence of the stopping activity at the traffic signal, the trajectory also shows the length of duration that the bus stopped at the signal. Such information about how long buses stop at a signal can be valuable for the decision making of transit signal priority projects.

2.7 Conclusion

This chapter delved into the methodologies used to reconstruct continuous, monotonic and differentiable bus vehicle trajectories from heartbeat data that contains noisy coordinates recorded at inconsistent frequencies. Raw coordinates can be realigned with the road networks through a map-matching process, and the result of such a process can be used to convert timestamped coordinates to a series of discrete time-distance data that serves as the input for the trajectory smoothing algorithm. Several trajectory smoothing algorithms are explored, and their goodness compared, in the selection of an algorithm that could produce a trajectory that not only satisfies the ideal properties but also performs well when validated against the expected speed and acceleration data. The best algorithm selected is LOCREG-PCHIP, which is essentially an algorithm that first predicts the true distances into trip of the bus at each recorded timestamp using nearby observed values using local regression, then interpolates for the distance at unobserved timestamps using the Piecewise Cubic Hermite Interpolant.

The continuous bus vehicles trajectories constructed using the methodologies described in this chapter allow for the extraction of the bus location, speed and acceleration at any point in time into its trip. The information contained in these trajectories will allow for further analysis of transit performance at the route, corridor and network levels, of multiple trajectories of the same route or of different routes traveling the same corridor are aggregated together, which is the focus of the following chapter.

Several limitations are present with the methodologies that could be addressed in future research. First, the methodologies presented in this chapter are developed using the heartbeat data with a frequency ranging between 1 and 10 seconds. Further research may be needed to determine how the algorithm will perform if the data frequency is much lower. Secondly, none of the smoothing algorithms presented in this chapter guarantees that the trajectory is twice-differentiable, therefore the acceleration profile is not guaranteed to be smooth. Although a check of the acceleration values illustrates that 99% of the accelerations produced by LOCREG-PCHIP are within the reasonable threshold, an algorithm that could guaranteed twice-differentiability could be useful if one wishes to obtain information such as driver aggressiveness. Lastly, the validation method used in this chapter to evaluate the goodness of algorithms is limited to comparing the speed and acceleration data with the expected values. An ideal method, however, is to compare the location, speed and acceleration of vehicles with real-world observations using high-precision measurement devices.

Chapter 3

Transit Delay Analysis

3.1 Introduction and Motivation

As discussed in the literature review, many researchers have attempted to use the historical GPS records of bus vehicles contained in the heartbeat data to understand various operational aspects of buses, such as stopping behavior at bus stops, signal delays, queuing delay, passenger in-vehicle travel time, etc. A common strategy used by researchers when attributing delays to different sources such as bus stops and traffic signals is by making certain assumptions, such that any delay activities observed near location of stops and signals are categorized as the corresponding delay. These assumptions do not always hold true when considering the real-world operations of bus vehicles. For example, if one were to specify a signal distance and state that only delays incurred within such distance are signal delays, then they would be assuming that the queue at any traffic signal would not exceed the signal distance, which is hardly justifiable. Similarly, if one were to specify a stop zone surrounding a bus stop and state that any stopping activity incurred within such zone is due to passenger pickup and drop off at the bus stop, they are assuming that the bus vehicle will always stop within the zone, which is unrealistic especially at bus stops with frequent bus bunching. Furthermore, existing research have mostly focused on analyzing the operations of one single bus route, but rarely mentions the possibility of analyzing multiple routes traveling on the same corridor.

To that end, this chapter presents methodologies that tackle the task of decomposing the travel time of bus vehicles by taking advantage of the continuous bus trajectories that can be reconstructed from heartbeat data. The author proposes methods for delineating traffic signal delay from the observed travel time following the logic from classic signal delay models and demonstrates ways to identify locations of bus dwell activities using a probability estimation algorithm. The methodologies presented in this chapter can be directly used to quantify the magnitude of various types of transit delay experienced by a bus route or observed along a bus corridor.

3.2 Analyzing Multiple Transit Vehicle Trajectories

In the previous chapter, a method was proposed to reconstruct a continuous, monotonic and differentiable transit vehicle trajectory, $x = f(t)$, from a series of discrete time-distance data T, D obtained from processing the raw heartbeat data of an individual vehicle trip. The trajectory constructed using this method always starts from the first recorded location given in the heartbeat data. For the trajectories of different trips to be comparable, however, a common spatial frame of reference is required to ensure that all trips depart from the same terminal stop and pass by the known facilities at the same distance into trip. This section provides details about how one can expand on the method for analyzing a single trajectory to perform such batch analysis of multiple bus trajectories, which will then serve as the basis for the delay analysis of bus routes and transit corridors.

3.2.1 Determining the Passing Times and Locations of Known Facilities

As shown in Table 3.1, the map-matching and trajectory smoothing processes not only associate each heartbeat record i with a road segment ID (r_i) and position along the segment (p_i), but also converts the distance d_i into a point on the smoothed trajectory $x_i = f(t_i)$. Conversely, one can take records of location points with known road seg-

ment ID and position along the segment, determine the distances along the smoothed trajectory, and back-calculate the corresponding timestamps when the vehicle would have moved by these locations.

Table 3.1: An example time-distance data table before the records of road infrastructure are processed.

Record Type	i	ts	t (sec)	r	p	d (m)	x (m)
stop	ST01			R_0	0.05		
heartbeat	0	2022-04-25 08:24:45	0	R_0	0.639	0	0
heartbeat	1	2022-04-25 08:24:50	5	R_0	0.940	20.167	12.015
signal	INT01			R_2	0.15		
heartbeat	2	2022-04-25 08:24:57	12	R_2	0.463	33.502	29.743
heartbeat	3	2022-04-25 08:25:00	15	R_3	0.116	42.567	40.236
heartbeat	4	2022-04-25 08:25:04	19	R_3	0.452	61.047	63.008
...	...	...	...	...	...	...	...

To illustrate how the back calculation can be done, consider two records that can be added to Table 3.1, one being the first stop (ST01) of the route, another being the first signalized intersection (INT01) that the bus encounters after starting its trip. Since the bus runs on a predetermined route, the location of all facilities and infrastructure such as bus stops, signalized intersections and pedestrian crossings are known to the analyst. Therefore, the road segment ID and position along segment information of both records are known. With this information, one can calculate the time and distance into trip at which the bus arrived at each facility.

Using Equation 2.2, the distance into trip of stop ST01 and signal INT01 can be calculated as follows:

$$\begin{aligned} x_{ST01} &= L_{r_{ST01}} * p_{ST01} - L_{r_0} * p_0 \\ &= L_{R_0} * p_{ST01} - L_{R_0} * p_0 \\ &= L_0 * 0.05 - L_0 * 0.639 = -39.463; \end{aligned} \tag{3.1}$$

$$
\begin{aligned}
x_{INT01} &= \sum_{j=0}^{id_{INT01}-1} L_{R_j} + L_{r_{INT01}} * p_{INT01} - L_{r_0} * p_0 \\
&= \sum_{j=0}^{2-1} L_{R_j} + L_{R_2} * p_{INT01} - L_{R_0} * p_0 \\
&= L_0 + L_1 + L_2 * 0.15 - L_1 * 0.2 \\
&= 31.937.
\end{aligned} \tag{3.2}
$$

Since the trajectory function $f(t)$ is monotonically increasing (i.e. non-decreasing), the mapping from the time space t to the distance space x has a one-to-many relationship, meaning that for any $x_i \in \mathbb{R}_{\geq 0}$, there could be more than one t where $f(t) = x_i$. When analyzing the impact of various facilities on bus movements, the "impact area" of the facility is defined as the area stretching from just downstream of the previous facility to just downstream of the facility of interest. Therefore, if the bus dwells at a facility at distance x_i and during the time interval $[t_a, t_d]$ for which $x_i = f(t_a) = f(t_d)$, the departing time from the facility, i.e., t_d, is the more appropriate value to be used in the calculation of the segment travel time. The distance into trip of each facility can then be back-calculated as follows:

$$
t_i = \max\{t : t = f^{-1}(x_i)\}. \tag{3.3}
$$

Therefore, the time and distance into trip information of known facilities can be calculated using both Equations 3.3 and 2.2. After shifting the time and distance into trip data of all records with respect to that of the first bus stop, a complete time-distance data table can be obtained as shown in Table 3.2. Note that the time and distance columns contain values that satisfy the trajectory function $f(t)$, and although only records related to heartbeat data and certain facilities are shown, the table can contain as many time-distance data points as the time resolution used in the analysis permits, because the function $f(t)$ is continuous.

Table 3.2: An example of a complete time-distance data table with records of road infrastructure added.

Record Type	i	t_i (sec)	x_i (m)	t_i (sec) shifted	x_i (m) shifted
stop	ST01	-10	-39.463	0	0
heartbeat	0	0	0	10	39.463
heartbeat	1	5	12.015	15	51.478
signal	INT01	9	31.937	19	71.400
heartbeat	2	12	33.743	22	73.206
heartbeat	3	15	40.236	25	79.699
heartbeat	4	19	63.008	29	102.471
...	...	...	...		

3.2.2 Aligning the Trajectories of Multiple Bus Trips

For trips that serve the same route pattern, bus vehicles are expected to pass by the same set of fixed facilities such as bus stops, traffic signals and pedestrian crossings. Therefore, such set of facilities can be used as a frame of reference when aligning the trajectories of multiple different bus trips.

Consider the example trip given in Table 3.2, the list of facilities and their distances into trip are calculated from the information about the corresponding road segment and position along the segment, which is a property of the route pattern and does not vary with trips. Therefore, all trips that serve the same pattern would share the same list of facilities as well as the distance into trip values at each facility, while the times at which the vehicle arrives at each facility are different. In other words, the trajectory function $f(t)$ of each trip would satisfy the following condition:

$$\exists t \in \mathbb{R}_{\geq 0} : f(t) = x_f, \forall x_f \in \mathbf{X}_f, \tag{3.4}$$

where: $\mathbb{R}_{\geq 0}$ = all non-negative real values;

t = a time into trip value;

x_f = the distance into trip value of a known fixed facility along the route;

$\mathbf{X}_f$ = the set of all distance into trip values of all fixed facilities along the route shared by all trips that serve the same route pattern.

Therefore, the same calculation procedure detailed in the previous section can be repeated on multiple trips serving the same route pattern, and the outcome of the calculations would be a set of unique trajectory functions $f(t)$'s, all of which satisfy the condition in Equation 3.4. Table 3.3 provides an example showing the time-distance data of multiple trips aligned to the same route pattern.

Table 3.3: An example of selected time-distance data of multiple trips that serve the same route pattern.

Record Type	i	x_i (m)	t_i (sec)		
			trip 1	trip 2	trip 3
stop	ST01	0	0	0	0
signal	INT01	71.40	19	22	21
signal	INT02	189	55	60	61
stop	ST02	...	...	...	...
crossing	CRS01	...	...	...	...

The methodology described here for aligning multiple trips to the same time-distance reference frame not only makes it easy to analytically compare the performance of buses across trips, but also allows for the trips to be displayed on the same time-space diagram for visual inspection.

3.2.3 Segmentation of Bus Routes

As mentioned above, a key concept when analyzing the impact of each facility on bus movement is that of the "impact area" of a facility, which is defined as the area stretching from just downstream of the previous facility to just downstream of the facility of interest. For example, as illustrated in Figure 3-1, to evaluate the impact of traffic signal INT02 on bus movements, the impact area can be defined as the stretch of the road between just downstream of INT01 and just downstream of INT02. Any delay observed within the segment INT01-INT02 can be attributed to one of the following: the signal delay caused by INT02, dwell caused by a bus stop between INT01 and INT02, or traffic congestion within the segment.

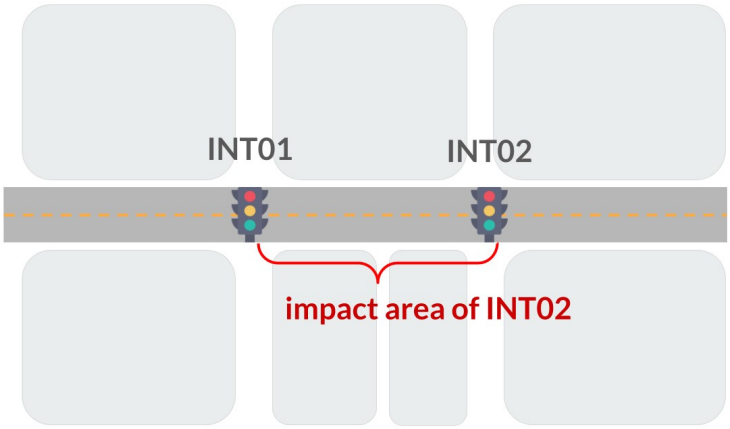


Figure 3-1: Illustration of a traffic signal impact area.

3.3 Quantifying Sources of Transit Delay

A segment-based approach is taken to analyze the impact of various roadway facilities on bus movements. This section details how the travel time and each delay component can be analyzed for individual segments, and how the segment metrics can be aggregated together when analyzing the performance of an entire route or a specific corridor.

3.3.1 Travel Time Decomposition

According to TCQSM, the observed travel time of buses is comprised of unimpeded running time T_u and additional running time losses T_l [19],

$$T_{obs} = T_u + T_l. \tag{3.5}$$

In particular, the unimpeded travel time is made up of free-flow travel time, dwell time, acceleration/deceleration time and bus slowdowns near the bus stop due to congestion. The running time losses, on the other hand, include additional travel times attributed to traffic signal delay and interference from other vehicles. These terms involved in the calculation of bus running time can be generalized into a few

categories, namely free-flow travel time, dwell time, signal delay, congestion delay and loss time. Therefore, Equation 3.5 can be rewritten as follows:

$$T_{obs} = T_{ff} + T_{dwell} + D_{signal} + D_{congestion} + Loss, \tag{3.6}$$

where: T_{obs} = observed travel time;
T_{ff} = free-flow travel time;
T_{dwell} = dwell time at bus stops;
D_{signal} = traffic signal delay;
$D_{congestion}$ = congestion delay (e.g. interference from other vehicles, slowdown near bus stops);
$Loss$ = loss time (e.g. acceleration, deceleration, etc.).

Equation 3.6 offers a method to decompose the total travel time of each bus trip into individual components. If such decomposition is performed on a large sample of trips, then one can take advantage of the distribution of each travel time component to obtain statistics such as the mean value and variance, and determine the significance that each source of delay carries in delaying bus movements. The expected value of the observed travel time can be written as:

$$\mathbb{E}[T_{obs}] = \mathbb{E}[T_{ff}] + \mathbb{E}[T_{dwell}] + \mathbb{E}[D_{signal}] + \mathbb{E}[D_{congestion}] + \mathbb{E}[Loss]. \tag{3.7}$$

In the following sections, details are provided on how one can calculate each term in Equation 3.7 in order to quantify the different types of delay.

3.3.2 Observed Travel Time

The observed travel time of a segment bounded by two facilities such as traffic signals, e.g. INT01-INT02, is measured as the difference between the time a bus departs from the upstream signal and the downstream signal.

$$T_{obs} = T_{INT02} - T_{INT01}. \tag{3.8}$$

With a large sample of trips, the observed travel time of a segment, T_{obs}, can be calculated for each trip, leading to a probability distribution of the metric. The expected value of the observed travel time can therefore be calculated as follows:

$$\mathbb{E}[T_{obs}] = \frac{\sum_{s \in \mathbf{S}} T_{obs,s}}{|\mathbf{S}|}, \tag{3.9}$$

where: s = a bus trip that traverses between the segment of interest;
$\mathbf{S}$ = the set of all bus trips that traverse between the segment of interest;
$T_{obs,s}$ = the travel time along the analyzed segment observed from trip s; and
$|\mathbf{S}|$ = the size of the set $\mathbf{S}$.

3.3.3 Free-Flow Travel Time

By definition, the free-flow travel time of buses in a segment is the total time it takes for a bus to travel from the beginning to the end of the segment without any delays. In practice, analysts usually use the travel times recorded during off-peak periods such as midnight as an estimator for the average free-flow travel time [20].

The expected value of the free-flow travel time obtained using the empirical estimator can then be formulated as follows:

$$\mathbb{E}[T_{ff}] = \mathbf{T}_{ff(5)}, \tag{3.10}$$

where: T_{ff} = the random variable for free-flow travel time;
$\mathbf{T}_{ff(n)}$ = the nth percentile free-flow travel time, or equivalently the $\lfloor \frac{n*N}{100} \rfloor$th (smallest) order statistic of the random sample of free-flow travel times, where N is the sample size.

3.3.4 Dwell Time

Dwell time is defined as the length of time a bus spends dwelling at a bus stop to pick up and drop off passengers. Since the onboard GPS device used to record heartbeat

data is often the same device used to record AVL data, it is reasonable to combine the two data sources together to support the analysis of dwell times. Specifically, the timestamps at which doors are open and closed on a bus vehicle at each stop recorded in the AVL data can be used to discern dwell activities from other stopping activities.

For each record k in the AVL data, where $k = 1, 2, ..., s$ and s is the total number of stops visited by a vehicle in its trip, the coordinate of the bus stop is $C_k = (lat_k, lon_k)$, the timestamp at which the first door opening activity occurs is denoted by DO_k, and the last door closing activity DC_k. An example of the layout of the AVL data is shown in Figure 3.4.

Table 3.4: An example of AVL data records.

trip_id	stop_id	C	door open time	door close time
TRIP01	ST01	(lat_1, lat_1)	DO_1	DC_1
TRIP01	ST02	(lat_2, lat_2)	DO_2	DC_2
...	...	...		

Since the heartbeat data is recorded at an inconsistent frequency every a few seconds, the data may not capture precisely the exact times when the vehicle doors are open and closed. One way to overcome the issue of imperfect dwell activity data is to append the data given by AVL to the heartbeat data. Using Equations 2.1 and 2.2, the timestamps DO_k and DC_k as well as the stop coordinate in each AVL record can be converted to time into trip and distance into trip values to_k, tc_k and x_k. The result of the conversion is shown in Table 3.5.

Table 3.5: An example of AVL data records converted to time and distance into trip values.

Record Type	k	t_k (sec) shifted	x_i (m) shifted
avl	ST01	to_1	x_1
avl	ST01	tc_1	x_1
avl	ST02	to_2	x_2
avl	ST02	tc_2	x_2
...	...	...	

The data in Table 3.5 can be appended to Table 3.2, and the combined data can

then be sent to the smoothing algorithm to produce a trajectory that can correctly reflect the dwell activities recorded in the AVL data.

The expected value of the dwell time at each bus stop can be estimated as follows:

$$\mathbb{E}[t_{dwell}] = \frac{\sum_{s \in \mathbf{S}} t_{dwell,s}}{|\mathbf{S}|}, \tag{3.11}$$

where: s = a bus trip that dwells at the bus stop of interest;
$\mathbf{S}$ = the set of all bus trips that dwell at the stop of interest;
$t_{dwell,s}$ = the dwell time at the analyzed stop observed from trip s;
$|\mathbf{S}|$ = the size of the set $\mathbf{S}$.

It may seem that since the dwell time information at each stop is already provided by the AVL data, there is no point in perfecting vehicle trajectories or trying to extract the same information from the trajectories. In fact, the purpose of improving the vehicle trajectory with AVL data is not to extract dwell time from it, but rather to make it possible to discern dwell time from other types of stopping activities such as signal delays and crossing delays, as shown in Figure 3-2.

3.3.5 Signal Delay

Signal delay is perhaps the most difficult to quantify, yet the most crucial in the operational analysis of buses on urban roads. The classical methods for modeling traffic signal delay often generalize vehicle delays caused by traffic signals into three categories, namely uniform delay, random delay and overflow delay [21].

As explained in the classical Webster model, uniform delay is the delay encountered by vehicles that arrive at a non-saturated signalized intersections at a constant rate; random delay is the additional delay experienced by vehicles with arrival rates following a probability (e.g. Poisson) distribution, where the random arrival of vehicles results in occasional over-saturation of the signal and leads to further delay; and overflow delay is the additional delay experienced by vehicles joining an ever-lasting queue at a signal that is over capacity for an extended period of time [22]. The delays

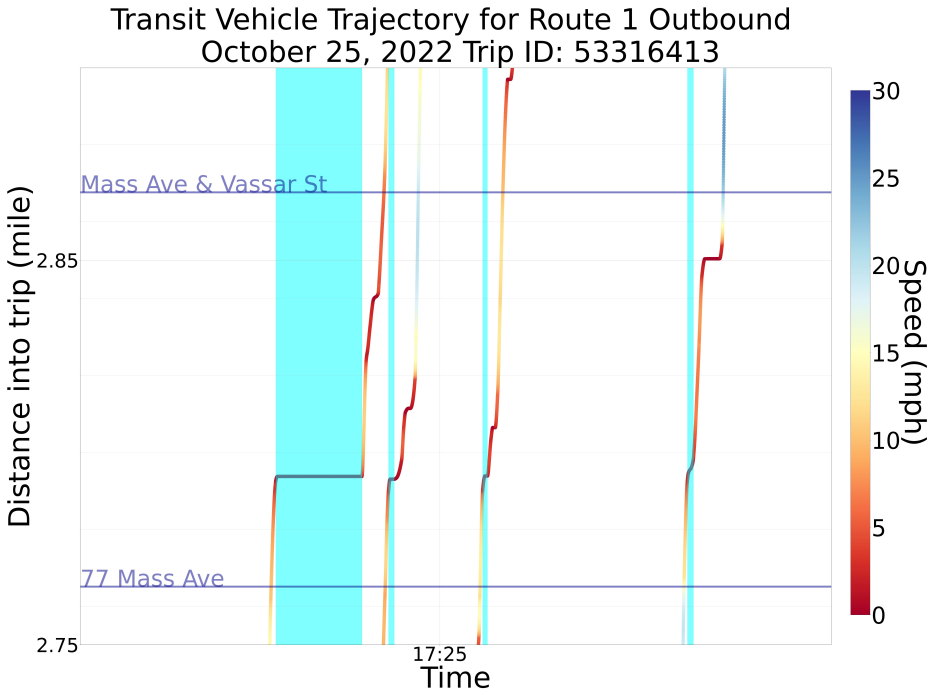


Figure 3-2: Example trajectories constructed using heartbeat data in combination wth AVL data. Blue bands are door-open times recorded in the AVL data.

calculated from these models describe the average delay that vehicles can experience at an intersection but does not offer insight into what each vehicle might experience individually.

From the perspective of an individual vehicle, only uniform delay and overflow delay are essential in the analysis of its movement through the signal. Regardless of arrival rate, a vehicle experiences uniform delay if it spends time waiting in a standing queue that can be served by the upcoming green phase, and overflow delay if it waits for the downstream queue to clear up before standing in the queue that can be cleared by the upcoming green phase. Random delay, in this case, is the result of vehicles experiencing uniform delay and an occasional overflow delay.

To simplify the use of terminologies, a queue that can be served by the upcoming green phase will be called a "typical queue"; on the other hand, a queue that cannot be fully served by the upcoming green phase, leaving leftover vehicles to be served by following cycles, will be called a "long queue".

Following the logic provided in the classical models, the signal delay of each vehicle

can be separated into two terms as follows:

$$d_{signal} = d_{uniform} + d_{overflow}, \tag{3.12}$$

where: d_{signal} = random variable for signal delay;
$d_{uniform}$ = random variable for uniform delay;
$d_{overflow}$ = random variable for overflow delay.

Specifically, as shown in Figure 3-3, buses may encounter different types of signal delay depending on the type of queue they end up joining. Some vehicles experience no signal delay since they either are the first to arrive at a green light or can join a moving platoon through the signal without stopping. Some only experience uniform delay, since they either are the first to arrive at a red light or encounter a typical queue without experiencing any cycle failure. Others would encounter a long queue that cannot be dissipated within one green phase, therefore experience both uniform delay and overflow delay.

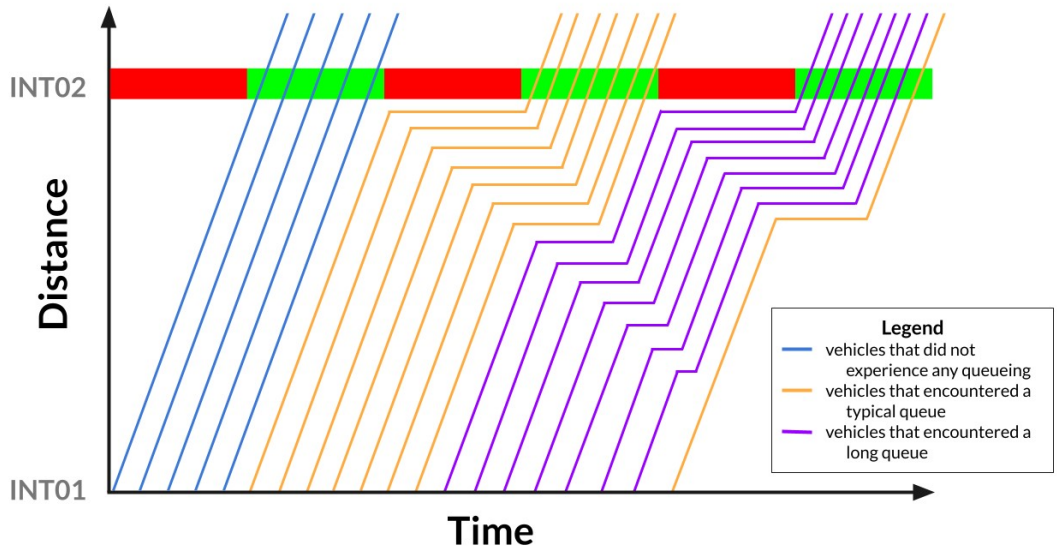


Figure 3-3: Different scenarios of buses passing through a signalized intersection.

Such categorization of bus trips by the type of queue a vehicle joins allows for the sample space of all bus trips to be divided into three sub spaces. Based on the

relationship given in Equation 3.12 and the Law of Total Expectation, the expected value of signal delay can be derived as follows:

$$\begin{aligned}
\mathbb{E}[d_{signal}] &= \mathbb{E}[d_{uniform} + d_{overflow}] \\
&= \mathbb{E}[d_{uniform} + d_{overflow}|\text{bus did not queue}] * \mathbb{P}[\text{bus did not queue}] \\
&\quad + \mathbb{E}[d_{uniform} + d_{overflow}|\text{bus stopped in a typical queue}] \\
&\quad * \mathbb{P}[\text{bus stopped in a typical queue}] \\
&\quad + \mathbb{E}[d_{uniform} + d_{overflow}|\text{bus stopped in a long queue}] \\
&\quad * \mathbb{P}[\text{bus stopped in a long queue}].
\end{aligned} \tag{3.13}$$

A few characteristics of vehicle movements in each scenario described above could help simplify Equation 3.13. Intuitively, buses that did not stop in a queue experienced no signal delay. In addition, buses encountering a typical queue would stop at the back of the queue and proceed to clear the intersection without experiencing overflow delay. Lastly, buses encountering a large queue would stop at the back of the queue, move forward when part of the downstream vehicles are cleared by one or more green phases, before clearing the intersection when the large queue ahead of the vehicle reduces to a typical queue. Therefore, Equation 3.13 can be rewritten as follows:

$$\begin{aligned}
\mathbb{E}[d_{signal}] &= \mathbb{E}[d_{uniform}|\text{bus stopped in a typical queue}] \\
&\quad * \mathbb{P}[\text{bus stopped in a typical queue}] \\
&\quad + \mathbb{E}[d_{uniform} + d_{overflow}|\text{bus stopped in a long queue}] \\
&\quad * \mathbb{P}[\text{bus stopped in a long queue}].
\end{aligned} \tag{3.14}$$

With the simplified equation in 3.14, The problem of figuring out the average signal delay then becomes the problem of finding out the probability that a bus joined each type of queue as well as the corresponding uniform and overflow delays.

The length of each type of signal delay is reflected by the duration of each stopping activity, or equivalently the time span of each horizontal section of the bus trajectory curve. Therefore, by observing the pattern and locations of all stopping activities from each bus's trajectory, one can determine the length and type of signal delays a vehicle experiences. As shown in Figure 3-4, such observation would be easy to make if one could obtain complete information of the trajectories of all vehicles traveling along the road segment and records of the signal phasing history for the time period of analysis. However, information regarding non-transit vehicles and traffic signals is not usually available to transit analysts, and it is only fair to assume that analysts have access to nothing beyond the transit heartbeat data and GTFS data.

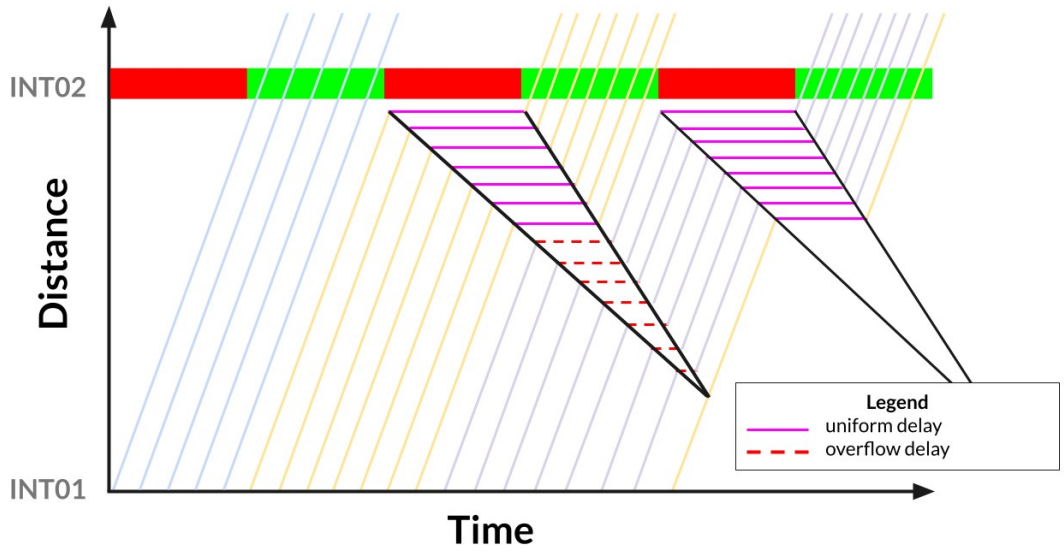
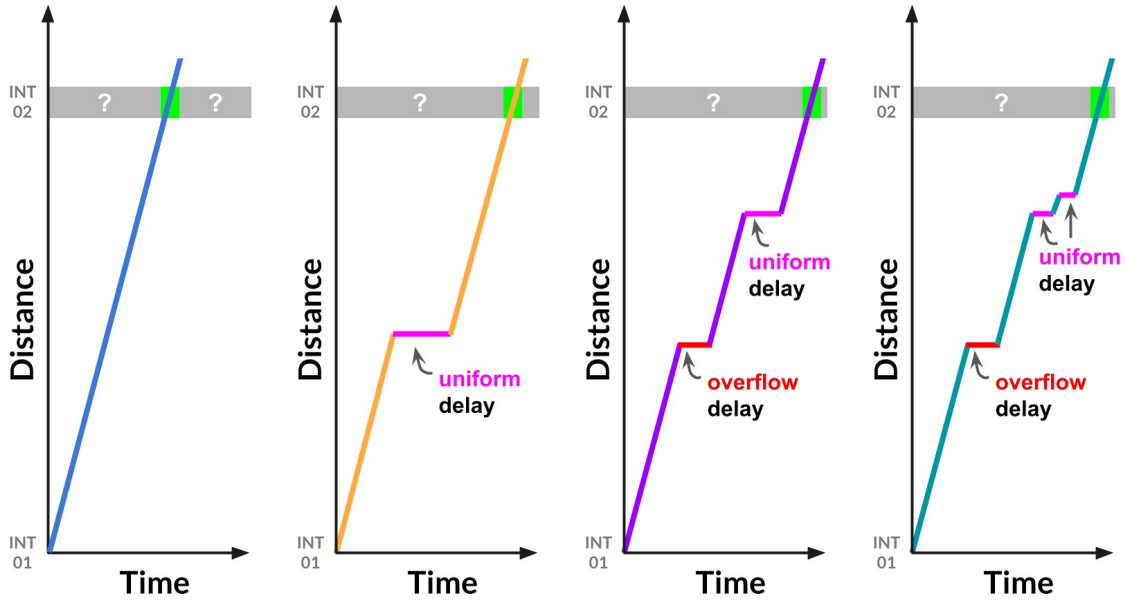


Figure 3-4: Identification of signal uniform and overflow delay.

In absence of historical signal phasing records and information about non-transit vehicle trajectories, the only data that is observable by transit analysts are bus vehicle trajectories through signalized intersections as illustrated in Figure 3-5. Rather than determining signal delays from the movements of vehicle platoons, one can discern uniform and overflow delay from the trajectory of an individual bus vehicle. If no stopping activity is observed from the trip as shown in Figure 3-5a, then the bus can be considered as passing without queuing, and therefore experienced no signal

delay. If only one stopping activity is observed as shown in Figure 3-5b, then the delay observed is uniform delay, and the bus can be considered as having joined a typical queue. If multiple stopping activities are observed at different locations of the road segment as shown in Figure 3-5c, then the delay observed closest to the signal is uniform delay while all others are overflow delay, indicating that the bus joined a long queue.



(a) Trip not delayed by signal.

(b) Trip passed through within one cycle.

(c) Trip waited for more than one cycle.

(d) Trip waited for more than one cycle with creeping behavior.

Figure 3-5: Different scenarios of buses moving through a traffic signal.

It is worth adding a few caveats to the arguments above. First, vehicles that experience uniform delay would in theory clear the intersection within the upcoming green phase, thus should not be delayed for longer than the duration of the red phase. Therefore, uniform delay at a traffic signal should be capped by the duration of the red phase. Second, it is common for buses in the back of a standing queue to exhibit "creeping" behavior, where vehicles near the stop bar leave the queue by turning on red, leading other vehicles in the queue to fill in the space by slightly moving forward. To take this behavior into account, stopping activities that take

place within a threshold distance of each other can be grouped together into the same category as shown in Figure 3-5d.

Since the mean uniform delay and overflow delays of all trips can be measured from the trajectories of individual bus trips, and the probability that a trip belongs to each queue category can be determined by counting the number of each type of trips, Equation 3.14 can be rewritten as follows:

$$\begin{aligned}\mathbb{E}[d_{signal}] &= \frac{\sum_{s\in\mathbf{S_{typical}}} d_{uniform,s}}{|\mathbf{S_{typical}}|} * \frac{|\mathbf{S_{typical}}|}{|\mathbf{S}|} \\ &\quad + \frac{\sum_{s\in\mathbf{S_{long}}} d_{uniform,s} + d_{overflow,s}}{|\mathbf{S_{long}}|} * \frac{|\mathbf{S_{long}}|}{|\mathbf{S}|} \\ &= \frac{\sum_{s\in\mathbf{S}} d_{uniform,s} + d_{overflow,s}}{|\mathbf{S}|}\end{aligned} \tag{3.15}$$

where: s = a bus trip that passes through the signal of interest;

$\mathbf{S}$ = the set of all bus trips that pass through the signal of interest;

$\mathbf{S_{typical}}, \mathbf{S_{long}}$ = the set of bus trips that encountered typical queue and long queue, respectively;

$d_{uniform,s}, d_{overflow,s}$ = the uniform and overflow delay of trip s, respectively;

$|\mathbf{S}|$ = the size of the set $\mathbf{S}$.

Consequently, the average uniform delay and overflow delay can be calculated as follows:

$$\begin{aligned}\mathbb{E}[d_{uniform}] &= \frac{\sum_{s\in\mathbf{S_{typical}}} d_{uniform,s} + \sum_{s\in\mathbf{S_{long}}} d_{uniform,s}}{|\mathbf{S}|}, \\ \mathbb{E}[d_{overflow}] &= \frac{\sum_{s\in\mathbf{S_{long}}} d_{overflow,s}}{|\mathbf{S}|}.\end{aligned} \tag{3.16}$$

3.3.6 Duration of Red Signal Phase

As discussed above, the uniform signal delay should be capped by the length of the red signal phase. Information about the red phase duration can be obtained from traffic signal timing plans, which unfortunately are not always available to transit

analysts. Therefore, a method for estimating the red phase duration using only bus heartbeat data is provided here.

As shown in Figure 3-4, of all stopping activities observable from vehicle trajectories, the duration of ones taking place closest to the analyzed signal match the length of the red phase. In other words, since vehicles that are stopped by the signal at the stop bar would experience longer red time than all other vehicles further upstream in the queue, the duration of stop activities at the stop bar can be a good estimator for the length of the red phase.

$$R = \{d_a : a \in \mathbf{A}_{stop\_bar}\}_{(95)}, \tag{3.17}$$

where: R = estimator for the red phase duration;
d_a = duration of stopping activity a;
$\mathbf{A}_{stop\_bar}$ = the set of all stopping activities that take place near the stop bar;
$\{d\}_{(n)}$ = the nth percentile stopping duration.

The choice of the percentile value to be used is up to the analyst and can be selected after comparing vehicle stopping duration with real signal timing values. It is worth noting that the 95th percentile value is used here instead of the maximum stopping duration to avoid choosing outliers which indicate longer-than-normal stopping time at the stop bar caused by downstream congestion.

3.3.7 Crossing Delay

For segments that end with pedestrian crossings, or railroad crossings which could also impose variable delays to buses, rather than traffic signals, the signal delay term, $\mathbb{E}[d_{signal}]$, in Equation 3.7 can be replaced by the crossing delay term $\mathbb{E}[d_{crossing}]$. The observed travel times of vehicles traveling through these types of crossing-bounded segments can therefore be decomposed as follows:

$$\mathbb{E}[t_{obs}] = \mathbb{E}[t_{ff}] + \mathbb{E}[t_{dwell}] + \mathbb{E}[d_{crossing}] + \mathbb{E}[d_{congestion}] + \mathbb{E}[loss]. \tag{3.18}$$

Since all other terms in Equation 3.18 are the same as in Equation 3.7 except for the crossing delay term, details for how crossing delays can be calculated will be discussed in this section, but signal-bounded segments will continue to serve as the focus of the discussion in the paper.

The calculation of crossing delay is similar to that of signal delay, but much simpler. Since pedestrians arrive at the crossing stochastically, the length of time a crosswalk becomes occupied is also stochastic and is not under control as are traffic signals. Therefore, there is no meaning in distinguishing uniform delay from overflow delay at crossings, and crossing delay is simply the sum of lengths of all stopping activities observed upstream of a pedestrian crossing. Formulaically, the expected crossing delay at a pedestrian crossing can be calcualted as follows:

$$\mathbb{E}[d_{crossing}] = \frac{\sum_{s \in \mathbf{S}} \sum_{a \in \mathbf{A}_s} d_{crossing,a}}{|\mathbf{S}|} \tag{3.19}$$

where: s = a bus trip that passes through the crossing of interest;

$\mathbf{S}$ = the set of all bus trips that pass through the crossing of interest;

a = a stopping activity;

$\mathbf{A}_s$ = the set of all stopping activities experienced by trip s;

$d_{crossing,a}$ = the duration of stopping activity a that takes place upstream of the crossing of interest;

$|\mathbf{S}|$ = the size of the set $\mathbf{S}$.

3.3.8 Loss Time

Loss times include when vehicles decelerate from cruising speed to standstill or accelerate back to travelling speed when approaching or departing from facilities such as bus stops, traffic signals and pedestrian crossings. Intuitively, the more stopping activities there are at these facilities, the larger the loss time would be. Technically, the total loss time of a trip can be measured from the stopping activities recorded in the vehicle trajectory, and the average loss time can be calculated as follows:

$$\mathbb{E}[loss] = \frac{\sum_{s \in \mathbf{S}} \sum_{a \in \mathbf{A}_s} acc_a + del_a}{|\mathbf{S}|}, \tag{3.20}$$

where: s = a bus trip that passes through the signal of interest;

$\mathbf{S}$ = the set of all bus trips that pass through the signal of interest;

a = a stopping activity;

$\mathbf{A}_s$ = the set of all stopping activities experienced by trip s;

acc_a = the time for bus to accelerate from standstill to travelling speed after stopping activity a;

dec_a = the time for bus to decelerate from travelling speed to standstill before stopping activity a;

$|\mathbf{S}|$ = the size of the set $\mathbf{S}$.

Compared to all the other delay terms and the total travel time of the trip, loss time is insignificant in magnitude. Although analysts can choose to explicitly calculate the loss time as formulated in Equation 3.20, the term is neglected in the analysis presented in this thesis for simplicity.

3.3.9 Congestion Delay

Congestion delay can be viewed as times that buses spend traveling below free-flow speed that are not due to stopping activities such as dwelling at bus stops or stopping at traffic signals and pedestrian crossings.

With all other terms in Equation 3.7 identified, the average congestion delay can be calculated as:

$$\mathbb{E}[d_{congestion}] = \mathbb{E}[t_{obs}] - \mathbb{E}[t_{ff}] - \mathbb{E}[t_{dwell}] - \mathbb{E}[d_{signal}] - \mathbb{E}[loss]. \tag{3.21}$$

3.4 Case Study 1: Route 1 Analysis

The previous sections detailed how the trajectories of individual bus trips can be analyzed together to quantify bus vehicle delays of various types. This section aims

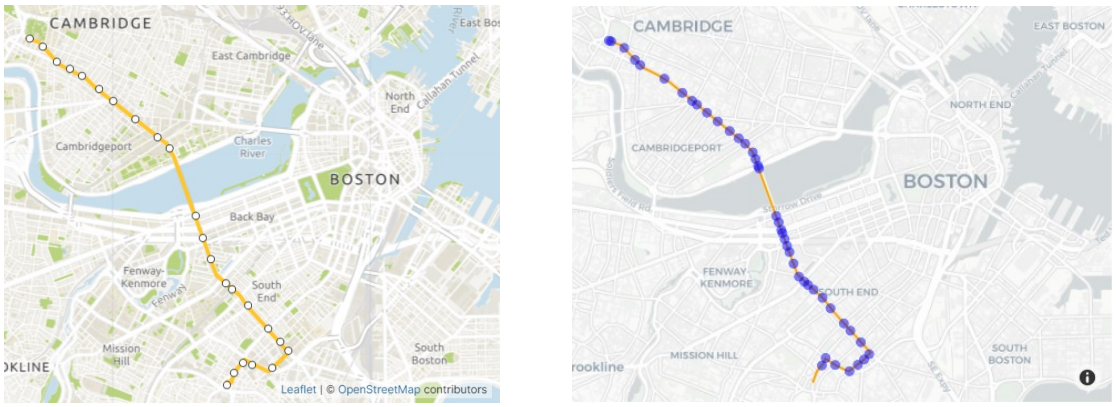
to illustrate how the methodologies outlined above can be used in the delay analysis of transit route and corridors using real-world examples.

The first case study focuses on analyzing and quantifying the different types of delay impacting the outbound direction of Route 1 operated by the MBTA.

3.4.1 Study Area

The study area is spanned by all Route 1 stops and signals that the route passes through, as shown in Figure 3-8.

Typical analysis of transit routes that uses AVL data usually defines study segments by bus stops as shown in Figure 3-6a, since AVL data only contains information down to the stop level. With bus heartbeat data, on the other hand, the study area can be divided into segments defined by traffic signals using methods described in Section 3.2.1 as shown in Figure 3-6b.



(a) Route 1 outbound marked by stops.

(b) Route 1 outbound marked by signals.

Figure 3-6: Study area of Route 1 outbound.

3.4.2 Study Time Period

The study analyzes the operations of Route 1 outbound (northbound) on 12 weekdays, Tuesdays through Thursdays, in a 5-week period between October 11, 2022 and November 3, 2022. Bus trips are grouped and aggregated by four time periods - Midnight (10 pm - 1 am), AM Peak (6 am - 9 am), Midday (11 am - 2 pm) and PM

Peak (4 pm - 7 pm). For simplicity, the following discussion only focuses on the bus operations during the PM Peak.

3.4.3 Data Source

The bus vehicle heartbeat data used in this case study is the GTFS-RT data, static GTFS data, and the AVL data of the analyzed Route 1 provided by the MBTA.

3.4.4 Time-Space Diagram of Multiple Bus Trajectories

The trajectory of multiple bus trips can be constructed and aligned to the same time-distance reference frame as described in Section 3.2.2. The result of aligning bus trajectories in the same time-space diagram is shown in Figure 3-7, where the trajectory of each trip is plotted by displaying the time into trip on the x-axis, and the distance into trip on the y-axis, and colored using the the speed of the bus that can be easily calculated by taking the first derivative of the trajectory function.

A few observations about the operations Route 1 can be made from the time-space diagrams. First of all, the coloring of trajectories by bus traveling speed allows the analyst to visually identify the specific locations along the route where buses are consistently slowed down or traveling freely, thus leading to easier identification of intersections and corridors needing examination and improvements. Secondly, by displaying the trajectories of the actual bus trips in parallel with those of the scheduled trips given by the GTFS data, the analyst can not only identify how late a trip departed from the first stop, but also inspect the stops, signals or crossings that exacerbate the reduced on-time performance of bus trips. Lastly, the diagram offers a easy visualization of the bus bunching phenomenon and allows analysts to have a preliminary understanding of where exactly along the route the headway between bus trips becomes substandard.

[1]Bus stops are not displayed to avoid overcrowding of graphics.

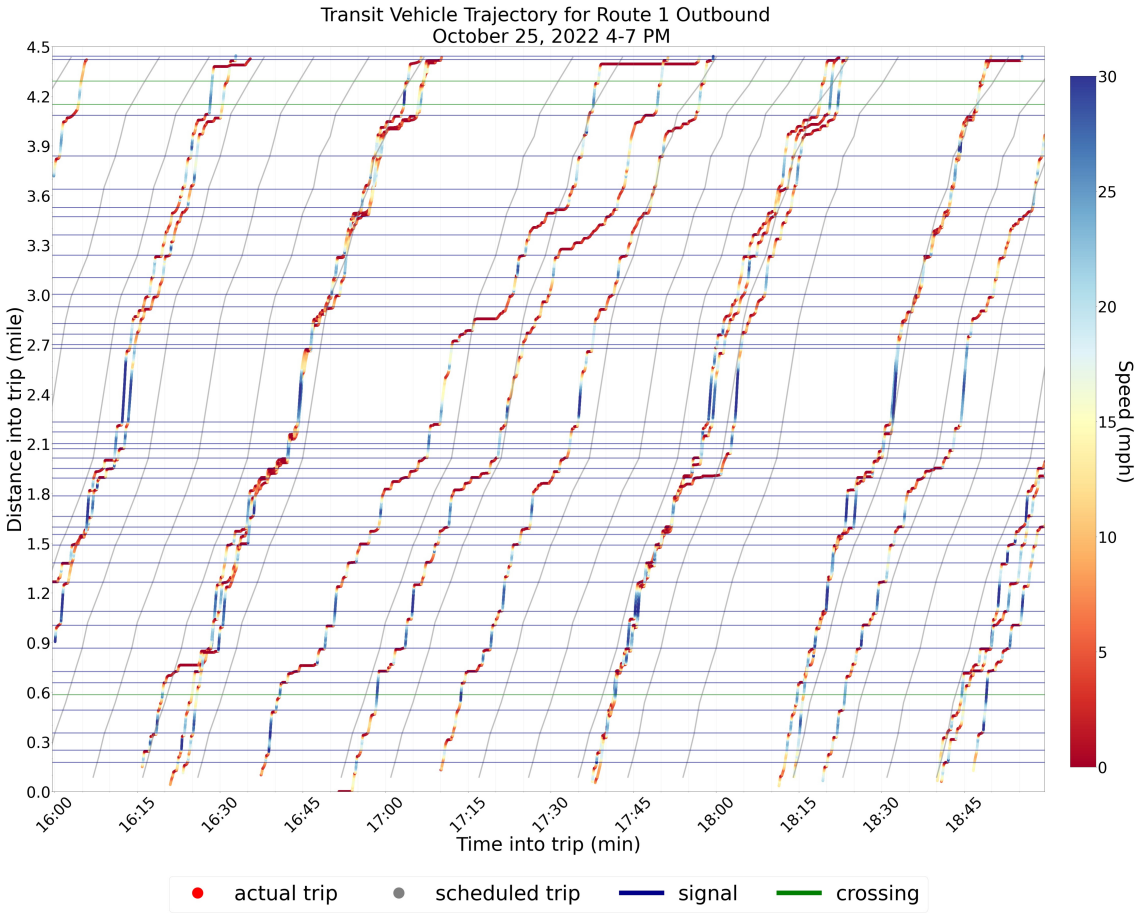


Figure 3-7: A time-space diagram showing multiple trajectories serving Route 1 outbound[1].

3.4.5 Delay Analysis

Besides making qualitative conclusions about transit operations by inspecting the time-space diagram, analysts can also quantitatively evaluate the different types of delay that transit vehicles experience and their relative impact on the overall bus operations using the methodologies presented in this research.

As explained in Section 3.3, the delay experienced by a bus trip can be analyzed using a segment-based approach, where the route is divided up into segments bounded by traffic signals and selected pedestrian crossings (e.g. signalized crossings, mid-block crossings, etc.). The northbound Harvard Bridge segment is used as an example to demonstrate how delay analysis can be conducted for a particular road segment. The location of the segment is shown on the map in Figure 3-8a, where it can be seen that

the segment is bounded by the signal at Massachusetts Avenue & Beacon Street on the south end and another traffic signal on the north end of the bridge.

A sample trajectory of a trip that served outbound Route 1 during the study time period is shown in Figure 3-8b. One can observe from the trajectory of this sample trip that while traveling on northbound Harvard Bridge, the bus was running at cruising speed until reaching the traffic signal, stopped for the signal to turn green, and then continued through the intersection. In this case, the bus encountered some amount of uniform signal delay and no overflow delay. Such observation can be made for all trips that traversed through this segment, and quantitative analysis can be carried out to determine the average delays of each type.

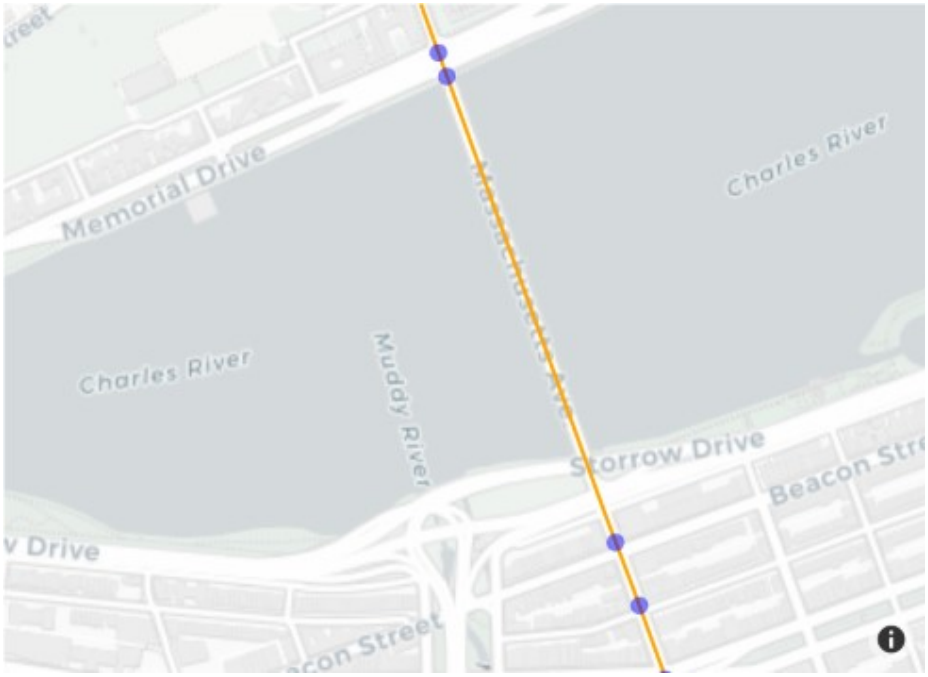
While inspecting a single road segment, the absolute time at which buses enter and leave the segment is less important than the relative time. Therefore, the time axis of each trajectory can be shifted so that time is 0 seconds at the beginning of the segment. Doing so will yield the plot of stacked trajectories as shown in Figure 3-9a, where trips displayed to the left of others experienced less delay, and the ones that are further to the right in the time-space diagram experienced more significant delays.

Observed Travel Time

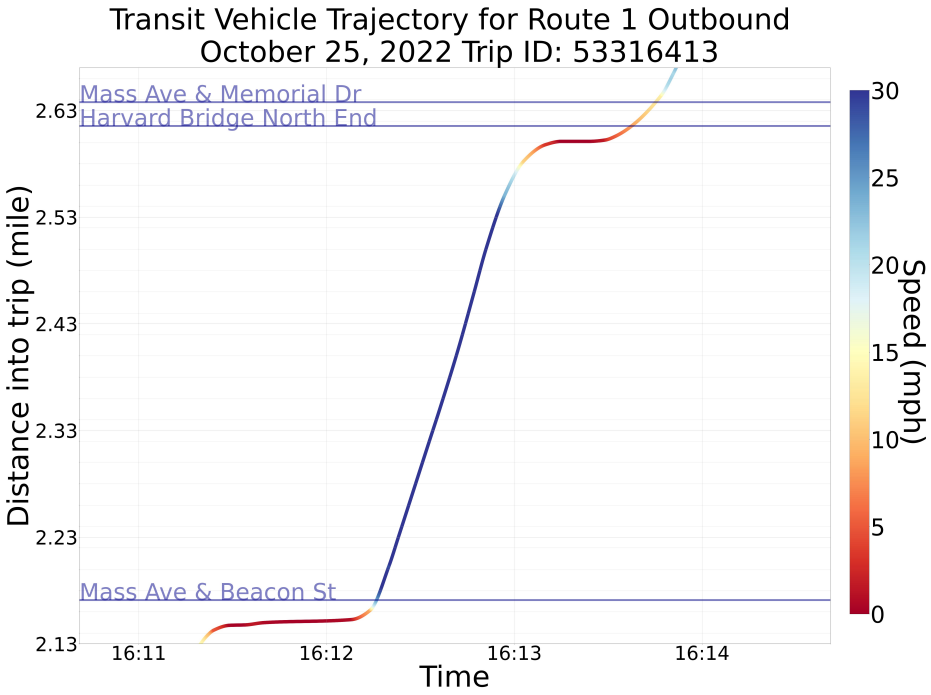
The observed travel time of each trip during the PM Peak is the time at which the bus leaves the segment. Of 298 outbound Route 1 trips recorded during the PM peak of the study time period, 166 trips are considered valid as these trips did not contain large data gaps within the analyzed segment, nor did the map matching algorithm fail to match the recorded coordinates to the predefined bus route. The distribution of the observed PM-peak travel times is shown in Figure 3-9b.

As formulated in Equation 3.9, the average observed travel time is the mean of the observed travel time of each trip, and can therefore be calculated as follows:

$$\mathbb{E}[t_{obs}] = \frac{\sum_{s \in \mathbf{S}} t_{obs,s}}{|\mathbf{S}|} = \frac{104.3 + 56.4 + 110.1 + 159.0 + ... + 77.3}{161} = 130 \text{ sec.} \quad (3.22)$$

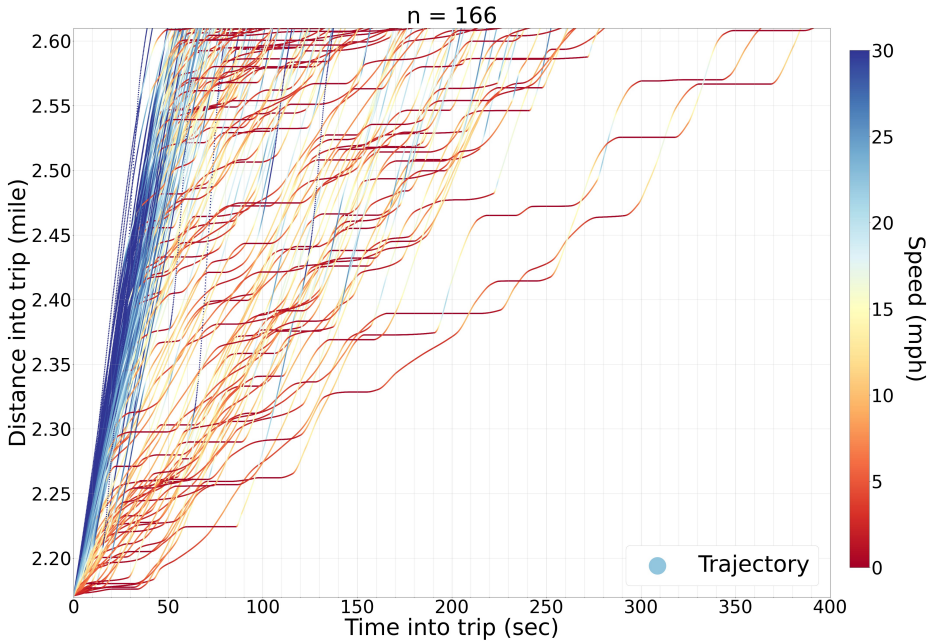


(a) Northbound Harvard Bridge shown on a map.

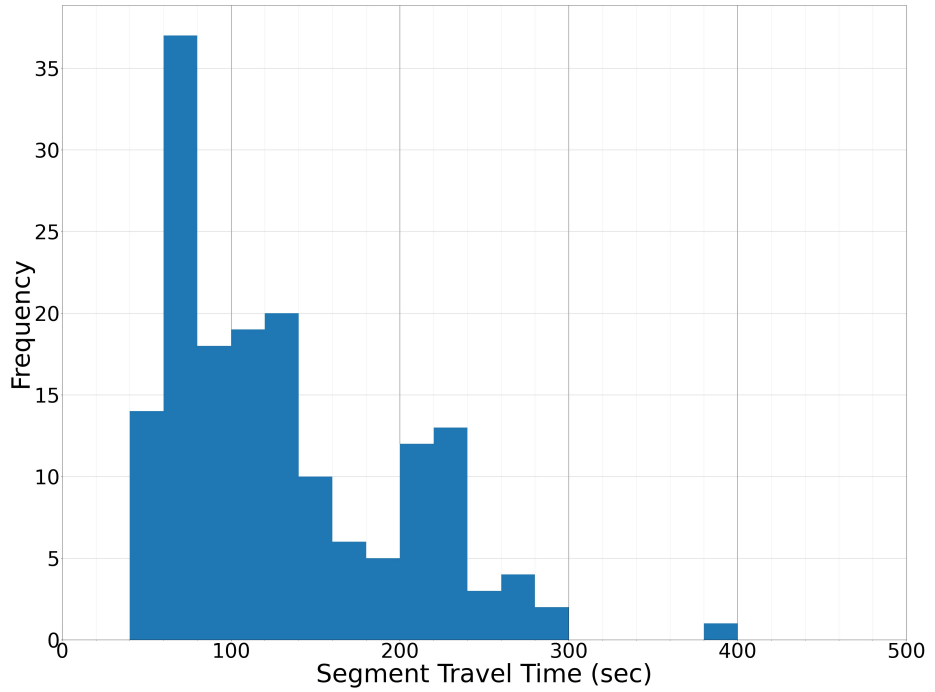


(b) A sample trajectory passing through the northbound Harvard Bridge segment.

Figure 3-8: The northbound Harvard Bridge segment.



(a) Stacked trajectories of PM-peak trips traversing through northbound Harvard Bridge.



(b) Histogram of observed travel times during the PM peak.

Figure 3-9: Observed travel times on northbound Harvard Bridge during the PM peak.

Free-Flow Travel Time

The free-flow travel time is derived from observations of 207 midnight trips, of which 161 are considered valid. The stacked trajectories and distribution of travel times of midnight trips are shown in Figure 3-10.

Using Equation 3.23, the expected free-flow travel time can be estimated using the $5th$ percentile travel time observed from midnight trips as follows:

$$\mathbb{E}[t_{ff}] = t_{ff(5)} = 42 \text{ sec.} \tag{3.23}$$

Dwell Time

For segments that do not contain any bus stops, such as the analyzed northbound Harvard Bridge segment, dwell time is zero and does not need to be taken into account when decomposing the observed travel times along the segment.

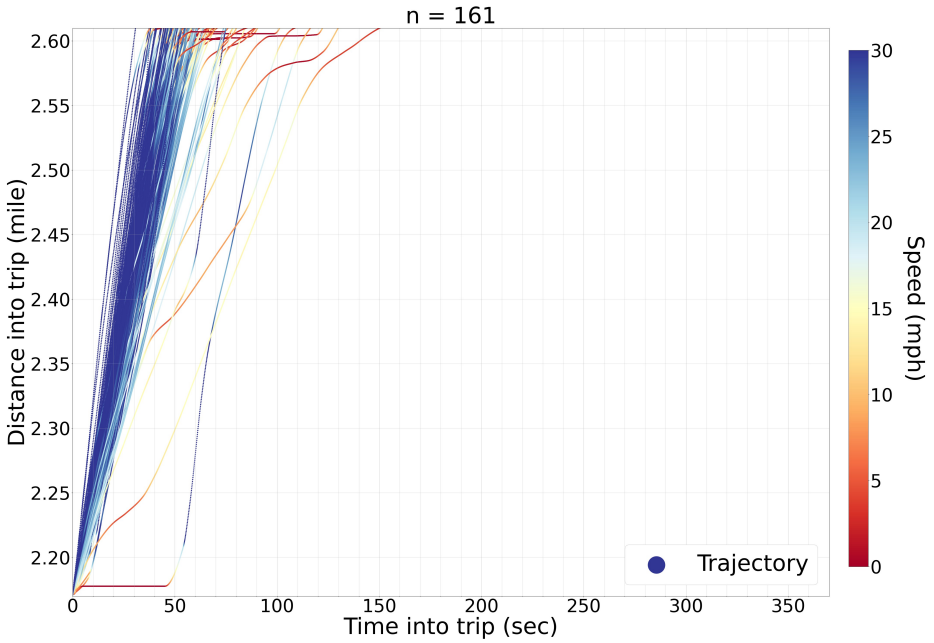
An illustration of dwell times is shown in Figure 3-11a, where the stopping activities of all trips that took place within the Vassar-Albany segment are plotted in the scatter plot with x-axis being the normalized distance along the segment, and y-axis the duration of the stopping activity. The dwell activities at the near-side Albany Street bus stop are colored red, and mostly take place near the stop bar as expected. The dwell times recorded from each bus trip in the Vassar-Albany segment are distributed as shown in the histogram in Figure 3-11b.

Using Equation 3.11, the expected dwell time at the Albany Street stop can be obtained as follows:

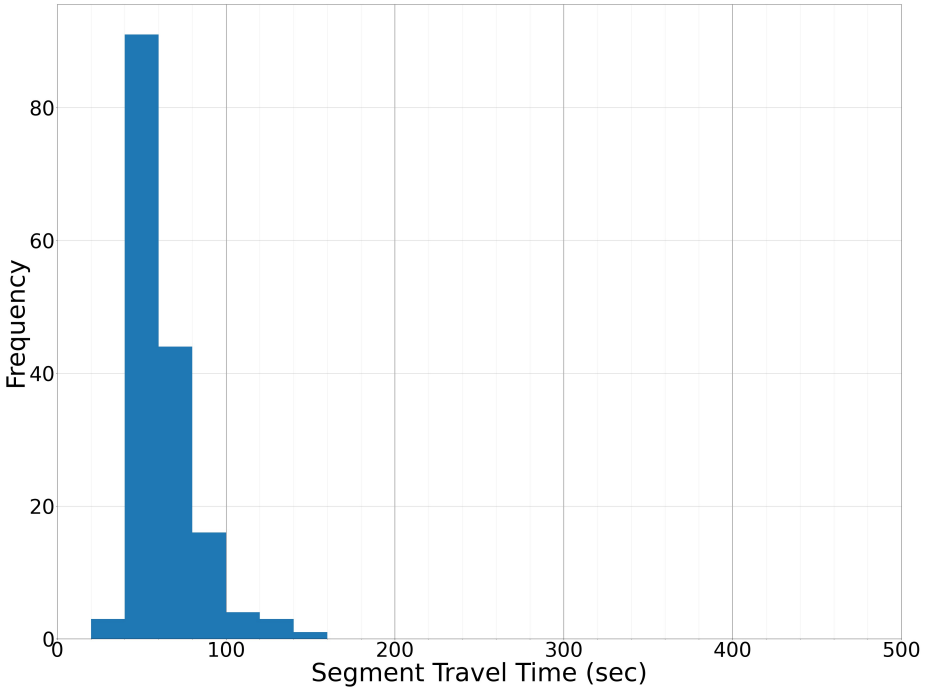
$$\mathbb{E}[t_{dwell}] = \frac{\sum_{s \in \mathbf{S}} t_{dwell,s}}{|\mathbf{S}|}, = \frac{0 + 9.4 + 0 + 46 + ... + 39.8}{159} = 26.7 \text{ sec.} \tag{3.24}$$

Duration of Red Signal Phase

As discussed in Section 3.3.6, the duration of the red phase of a traffic signal can be estimated using the duration of stopping activities that occurred close to the stop

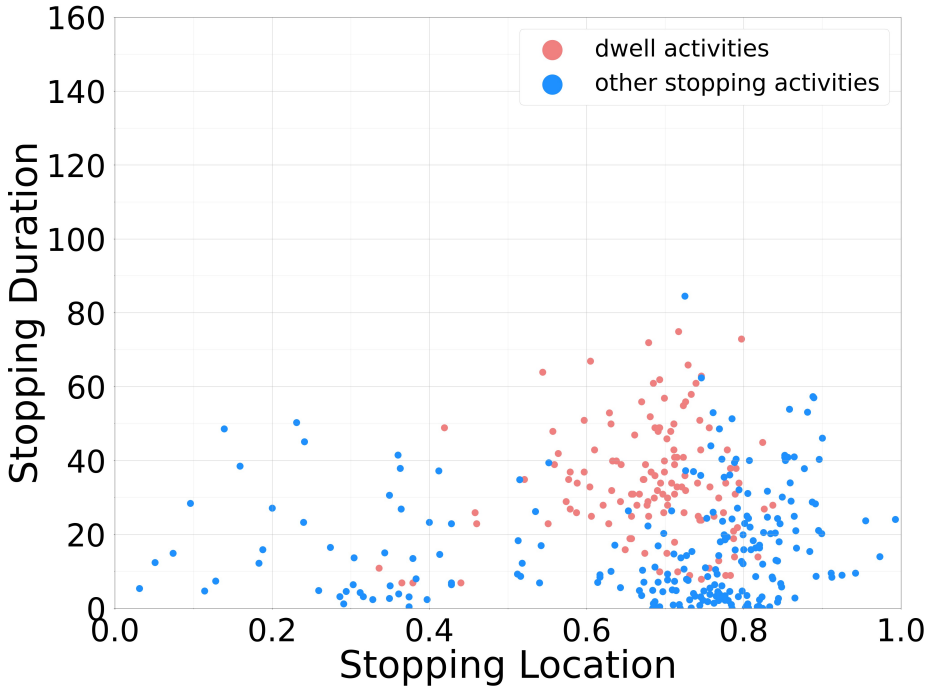


(a) Stacked trajectories of midnight trips.

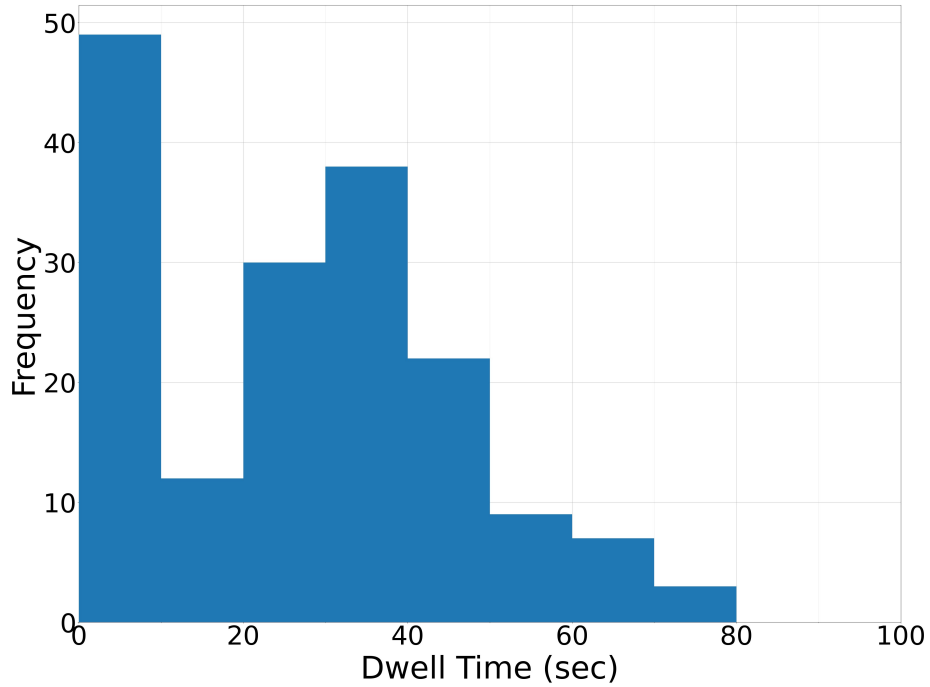


(b) Histogram of observed travel times during the midnight.

Figure 3-10: Observed travel times on northbound Harvard Bridge during midnight.



(a) Scatter plot of stopping activities within the Vassar-Albany segment during the PM peak.



(b) Histogram of dwell times during the PM peak at the Albany St stop.

Figure 3-11: Visualization of stopping activities and histogram of dwell times in the Vassar-Albany segment.

bar. For this case study, the authors chose to define an estimator for the red phase as the $90th$ percentile duration of all stopping activities that take place in the last 20% of the segment length.

Take the northbound Harvard Bridge segment as an example - if the length of the segment is normalized so that 0 represents the Massachusetts Avenue & Beacon St signal, and 1 represents the signal at the north end of the bridge as shown in Figure 3-12a, then the estimated red phase length is the $90th$ percentile duration of all stopping activities that took place between 0.8 and 1 segment length.

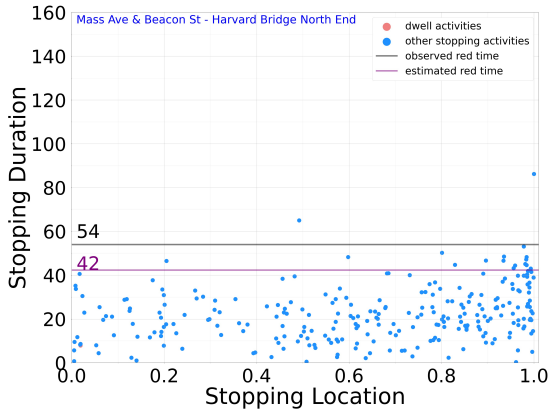
For all six signalized intersections in Figure 3-12, field observations were made to compare the estimated red time to the observed red time. The result of the comparison is detailed in Table 3.6, where it can be seen that out of six signals, the estimated red times of five are within 12 seconds of the observed signal length or on average with a 5-second difference, except for the Landsdowne signal.

Table 3.6: Comparison of Estimated and Observed Red Times

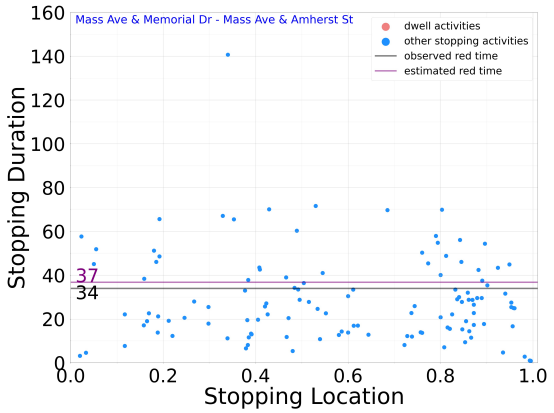
Location	Est. Red Time	Obs. Red Time	Diff [1]	% Diff [2]
Harvard Bridge North End	42	54	12	22%
Amherst	37	34	-3	-9%
Vassar	61	59	-2	-3%
Albany	44	43	-2	-5%
Landsdowne	17	43	26	60%
Brookline	42	50	8	16%

[1] Diff = Obs. Red Time - Est. Red Time
[2] % Diff = Diff / Obs. Red Time * 100 %

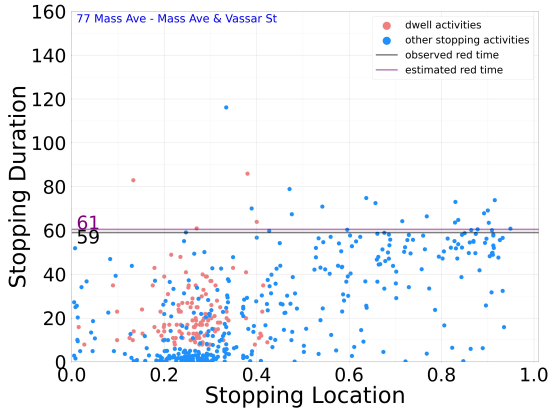
Further examination of the Landsdowne signal reveals that the large difference between the estimated and observed red times is in fact caused by the timing offset with the upstream Albany signal. The visualization of stopping activities upstream of the Landsdowne signal, as displayed in Figure 3-12e, shows that most stopping activities near the stop bar have lengths much shorter than the observed red time, indicating that a portion of the red signal phase is never experienced by bus vehicles. This observation can be explained by the illustration in Figure 3-13. From field observation, it was noticed that the upstream signal at Albany St turns green when



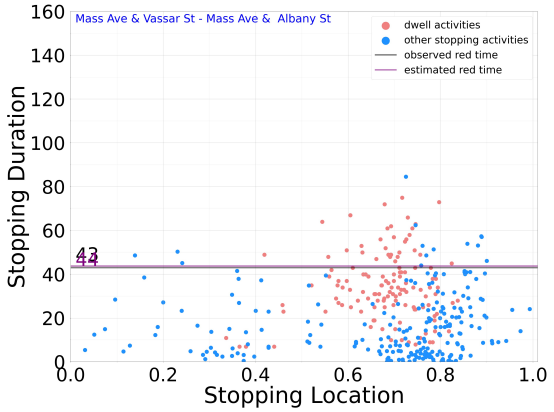
(a) Northbound Harvard Bridge.



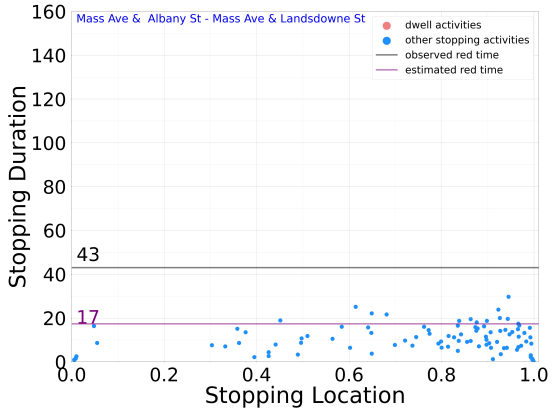
(b) Mass Ave & Beach St to northbound Harvard Bridge to Mass Ave & Amherst St.



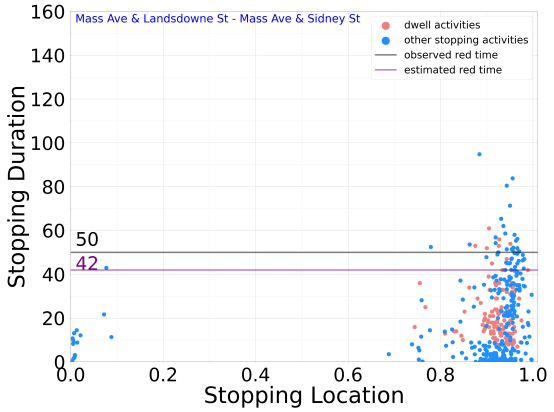
(c) 77 Mass Ave to Mass Ave & Vassar St.



(d) Mass Ave & Vassar St to Albany St.



(e) Mass Ave & Albany St to Landsdowne St.



(f) Market Central Crossing to Mass Ave & Sidney St.

Figure 3-12: Visualization of the length and location of stopping activities in different segments.

the signal at Landsdowne St turns red. As a result, the length of red time at the Landsdowne signal that can be experienced by any vehicle traveling northbound along

Massachusetts Avenue, such as all vehicles serving outbound Route 1, will be shorter than the actual red time at Landsdowne due to the time it takes for vehicles released from the Albany to travel to the stop bar at Landsdowne. Therefore, the red time estimated using methods described in Section 3.3.6 is an estimate of the "effective" red time, i.e. the maximum possible red time that can be experienced by the bus vehicles, and is not necessarily equal to the length of the entire red phase.

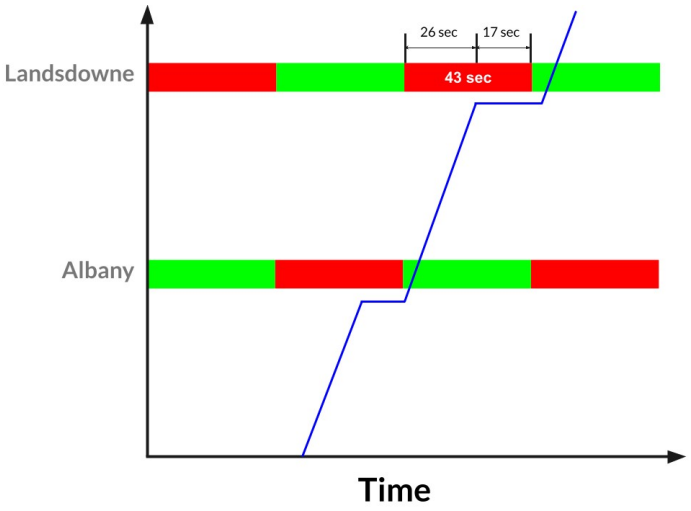


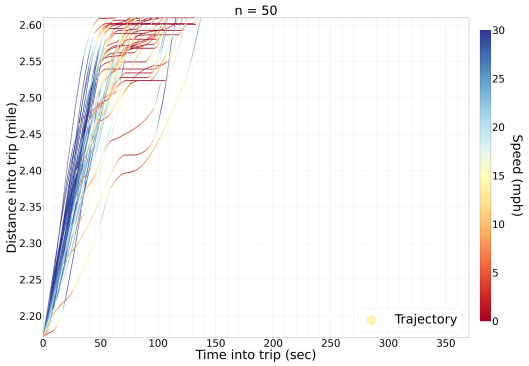
Figure 3-13: Illustration of the signal timing at the Albany and Landsdowne signals.

It is also worth noting that the estimated red time at Harvard Bridge North End and Sidney St are 22% and 16% different from the observed red time. Without signal time plans provided by the agency operating traffic signals, it would be difficult for transit analysts to further improve the estimate using only heartbeat data. If analysts can manage to obtain the timing plans of signals that buses travel through, the actual red time can be used in place of the estimated red time in the signal delay analysis. When using red times obtained from field timing plans, it is important to consider the offset with the upstream signal and use the "effective red time" as opposed to the maximum red time of coordinated signals.

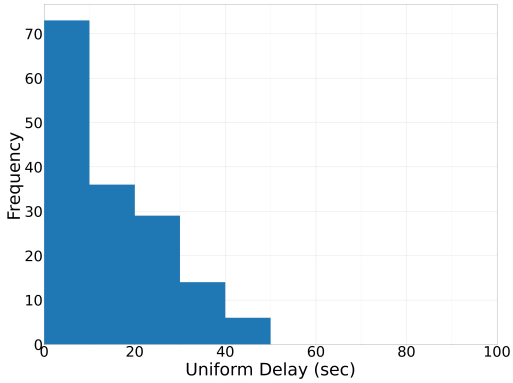
Signal Delay

As described in Section 3.3.5, signal delay can be categorized into uniform delay and overflow delay. Of all 166 trips that traversed through the northbound Harvard Bridge

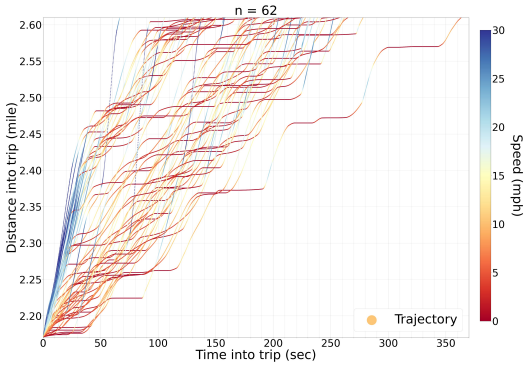
segment shown in Figure 3-9a, 50 experienced uniform delay only (Figure 3-14a), and 62 experienced both uniform and overflow delay (3-14c).



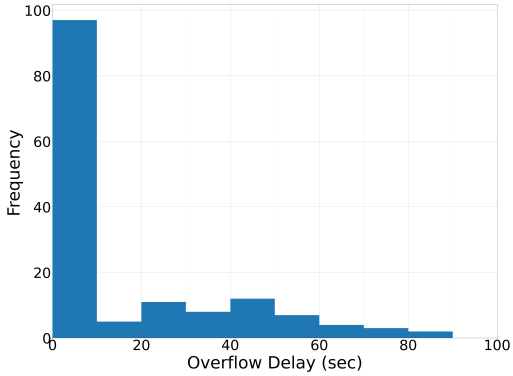
(a) Stacked trajectories of PM-peak trips that only experienced uniform delay.



(b) Histogram of uniform delays during the PM peak.



(c) Stacked trajectories of PM-peak trips that experienced uniform and overflow delay.



(d) Histogram of overflow delays during the PM peak.

Figure 3-14: Stacked trajectories and histograms of delays of PM-peak trips on northbound Harvard Bridge that experienced uniform delay only as well as uniform plus overflow delay.

Using Equation 3.16, the expected uniform delay and overflow delay can be calculated as follows:

$$\begin{aligned}\mathbb{E}[d_{uniform}] &= \frac{\sum_{s\in\mathbf{S_{typical}}} d_{uniform,s} + \sum_{s\in\mathbf{S_{long}}} d_{uniform,s}}{|\mathbf{S}|} \\ &= \frac{22.4 + 23.4 + 0 + ... + 17.8}{161} = 13.6 \text{ sec}, \\ \mathbb{E}[d_{overflow}] &= \frac{\sum_{s\in\mathbf{S_{long}}} d_{overflow,s}}{|\mathbf{S}|} = \frac{0 + 0 + 45.9 + 53.6 + ... + 0}{161} = 20.83 \text{ sec}.\end{aligned} \tag{3.25}$$

Crossing Delay

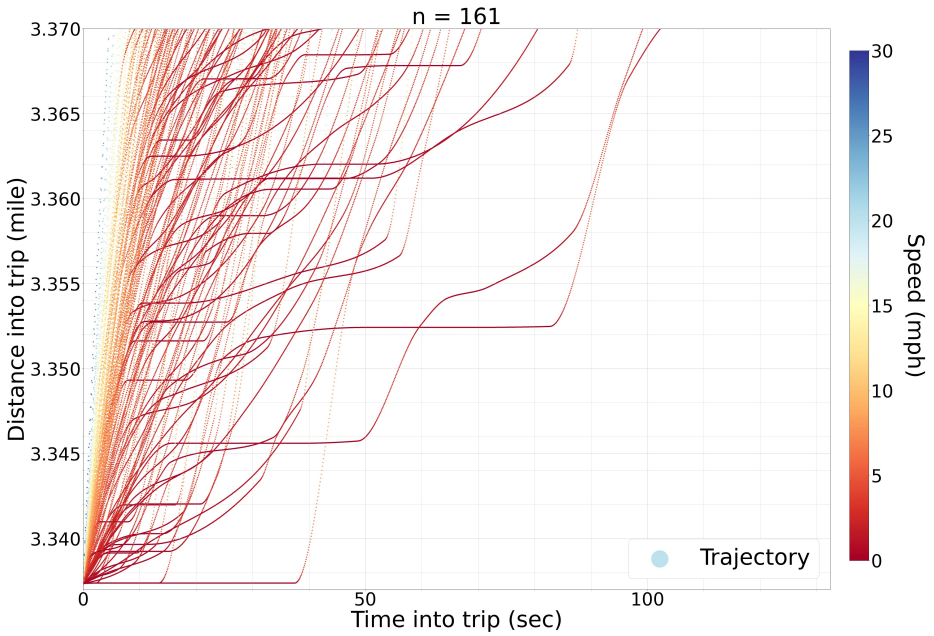
Crossing delay can be calculated for segments that end with a marked pedestrian crossing. Such analysis can be helpful for analysts who wish to understand the impact of crossings on transit operations. For example, the Norfolk-Pearl segment is bounded by the mid-block crossing at Massachusetts Avenue & Pearl St, which is enhanced with Rectangular Rapid Flashing Beacons (RRFB). Therefore, it may be of interests to analysts to know how often or how long on average buses stop at the crossing due to pedestrian crossing activities. The stacked trajectories of valid bus trips that traveled along the Norfolk-Pearl segment is shown in Figure 3-15a, and the distribution of crossing delays is shown in Figure 3-15b.

Using Equation 3.19, the expected crossing delay at the Pearl St crossing can be obtained as follows:

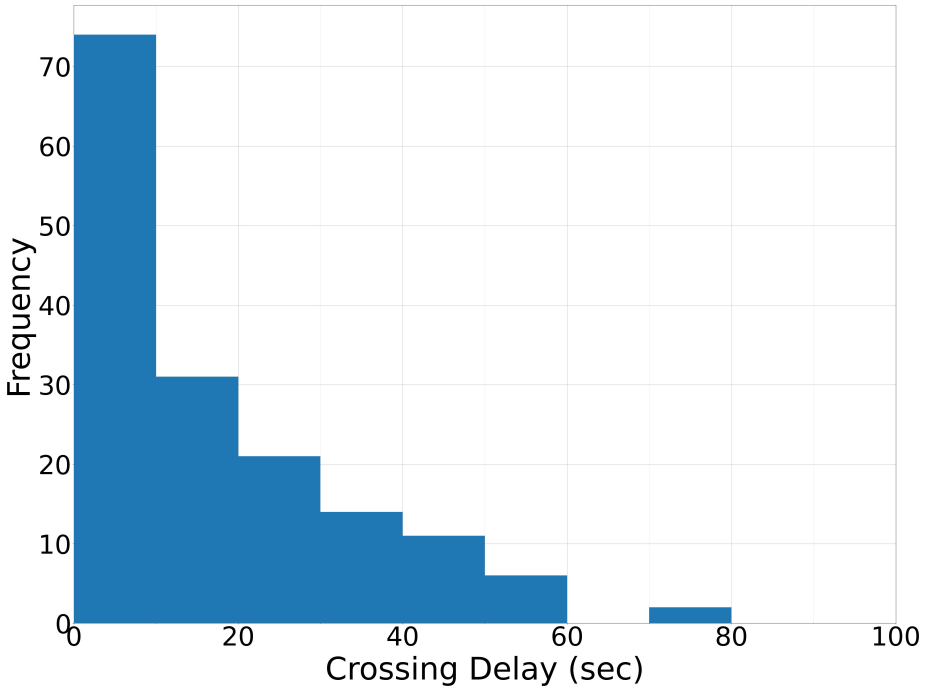
$$\mathbb{E}[d_{crossing}] = \frac{\sum_{s\in\mathbf{S}} \sum_{a\in\mathbf{A}_s} d_{crossing,a}}{|\mathbf{S}|} = \frac{0 + 47.9 + 33.6 + 7.9 + ... + 9.0}{161} = 17.4 \text{ sec}. \tag{3.26}$$

3.4.6 Results

Following the procedure above, delay analysis can be carried out for each segment along the studied outbound Route 1, and the results of the delay analysis for selected segments during the AM and PM peak periods are shown in Figure 3-16.



(a) Stacked trajectories of PM-peak trips.



(b) Histogram of crossing delays during the PM peak.

Figure 3-15: Stacked trajectories and observed crossing delays of trips through the Norfolk-Pearl segment.

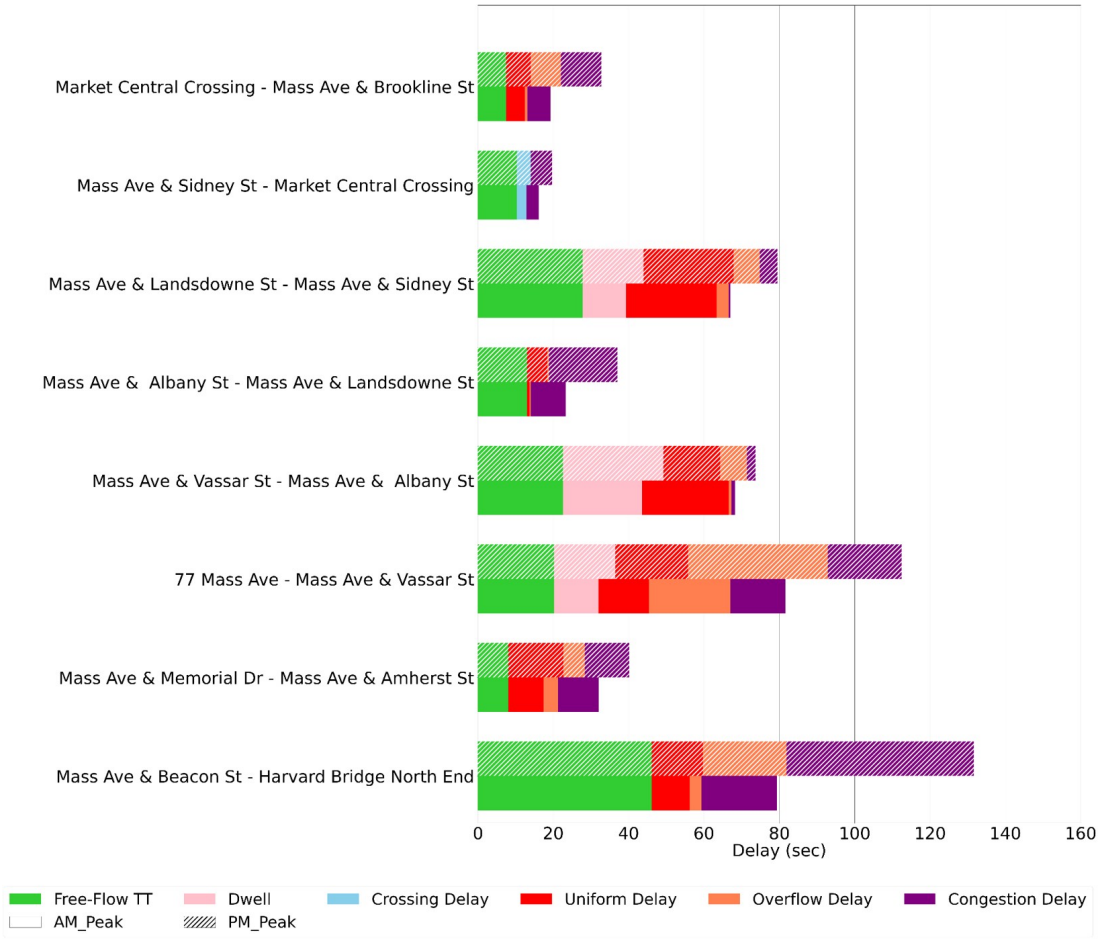


Figure 3-16: Result of the delay analysis for outbound Route 1 during the AM and PM peak periods (segment names are shown from south to north bottom up).

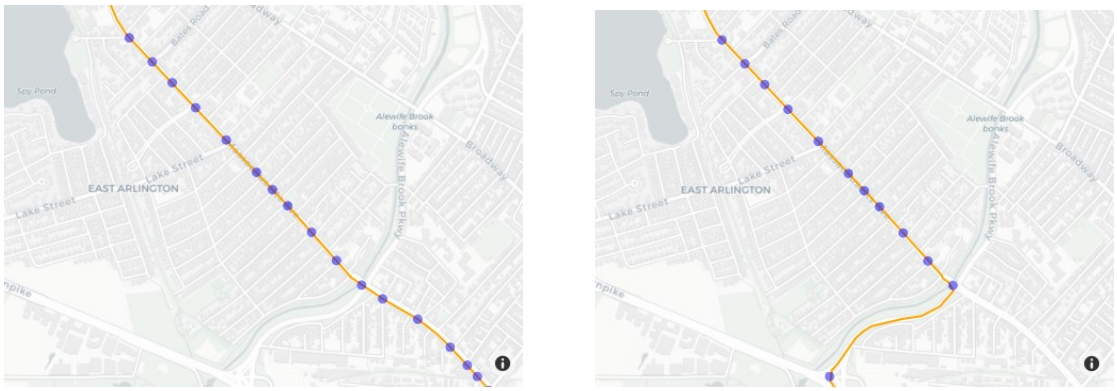
3.5 Case Study 2: Arlington Massachusetts Avenue Analysis

A second case study is provided as an example for how the methodology proposed in this study can be used to quantify different types of bus delay for not just a single route, but also multiple routes traveling the same corridor.

3.5.1 Study Area

The study area for this case study is southbound Massachusetts Avenue between Wyman St in Arlington and Alewife Brook Parkway served by inbound (southbound) Route 77 and Route 350 operated by the MBTA. The overlapping area served by

both routes are shown in Figure 3-17.



(a) Inbound Route 77 marked by intersections.

(b) Inbound Route 350 marked by intersections.

Figure 3-17: Inbound Route 77 and Route 350 shown on maps.

3.5.2 Study Time Period

The study analyzes the operations of Route 77 and 350 inbound (southbound) on 24 weekdays, Tuesdays through Thursdays, in an 8-week period between September 6, 2022 and October 27, 2022. Bus trips of both routes that served the study corridor during the AM Peak period, defined as the 3-hour window from 6 am to 9 am, are used for the delay analysis.

3.5.3 Data Source

The data used in this case study is the same as that used in the previous case study, namely the GTFS-RT data, static GTFS data and AVL data.

3.5.4 Delay Analysis Results

When conducting delay analysis of a transit corridor served by multiple bus routes, the procedure is almost identical to that of a single route. The main differences lie within the fact that corridor analysis can take advantage of a larger number of trips completed by more than one route, thus allowing the analyst to obtain better estimates of delays by using a richer set of vehicle trajectories.

As shown in Table 3.7, if only Route 350 is used in analyzing the southbound transit corridor on Massachusetts Avenue, then data from only 181 trips is available during the AM peak. By aggregating the performance of Route 350 trips with those of Route 77 that serves exactly the same corridor in a part of its route, the number of trips available for analysis is increased to over 1,000.

Table 3.7: Number of bus trips available for delay analysis

Time Period	Route 350 Only	Route 77 Only	Both Routes Combined
Midnight	41	521	562
AM Peak	181	855	1036
Midday	115	671	786
PM Peak	194	940	1134

To exemplify the impact of the increased sample size, the following discussion is provided. A portion of the study corridor on southbound Massachusetts Avenue, between Thorndike St and just before the Alewife Brook Parkway intersection, underwent a transit improvement project where a bus-only lane is installed in the curb lane. The bus-only lane operates on weekdays from 6 am to 9 am only and is open to general traffic in other hours of the day.

Suppose for some reason an analyst wants to understand the effectiveness of the bus-only lane by comparing the delay results between the AM and PM peak periods. Without discussing the validity of this method, the point of this discussion is to illustrate how the increased sample size offered by multiple routes can provide better insight into the operations of the transit corridor. As shown in Figure 3-18, where the total travel time and different types of delay are calculated and presented for each route separately, it is unclear how the AM operations within the Lafayette-Alewife segment compares with those in the PM. Results from Route 77 shows that the average AM travel time through the segment is shorter than the PM, but results from Route 350 show otherwise.

By aggregating all trips completed by both route 77 and 350, the delay analysis results shown in Figure 3-19 illustrate that the segment travel time, and in particular,

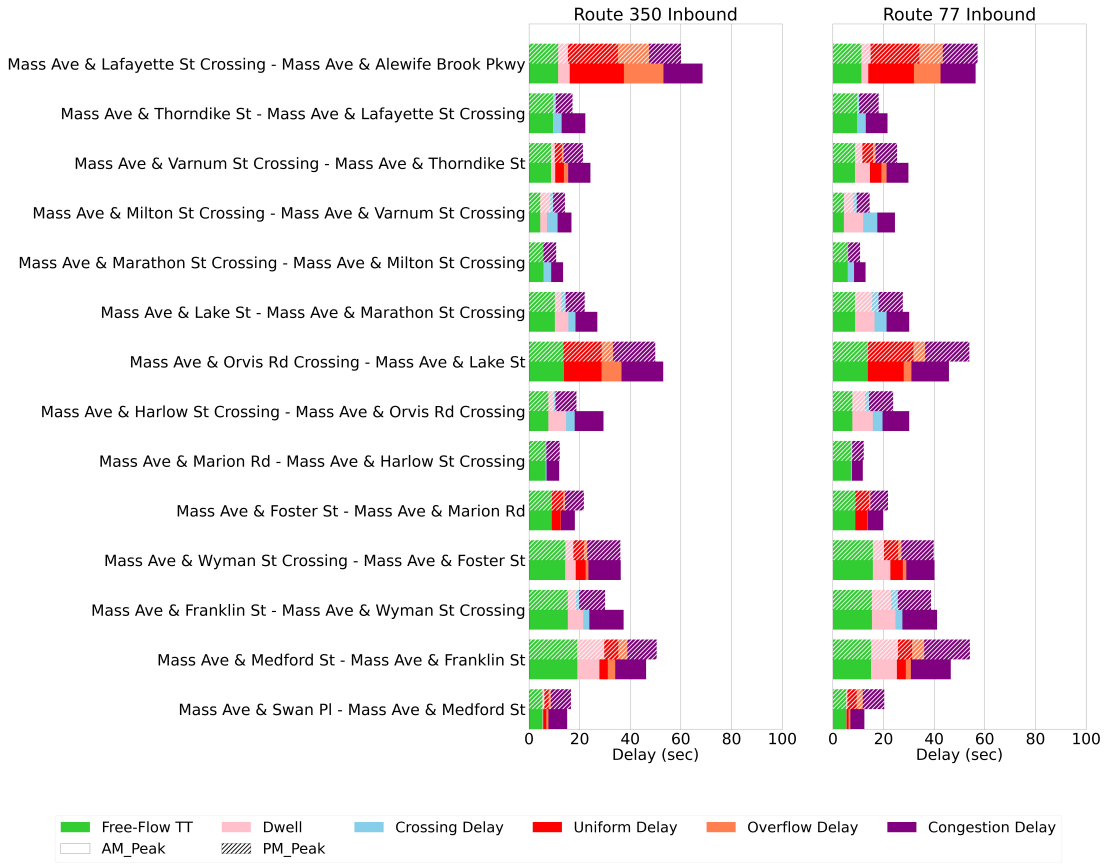


Figure 3-18: Delay analysis results of inbound Route 77 and Route 350 trips along the study corridor calculated separately.

signal delay and congestion delay in the segment, are comparable between the AM and PM peak periods.

In fact, an examination of the AM and PM travel times decomposed into each category as alternatively displayed in Figure 3-20a reveals that the main factors that resulted in the greater AM travel time than PM within the Lafayette-Alewife segment are overflow signal delay and congestion delay, while dwell time and uniform signal delays observed during the AM peak are less than or similar to those in the PM peak. In theory, bus lanes should have helped reduce overflow delay and congestion delay, so it is inconclusive whether or how much the bus lane helped improve congestion within this segment by simply comparing delays between the AM and PM peaks. If the analyst conducts similar analyses using data collected before and after the installation of the bus lane, then the change in time of AM congestion delay would

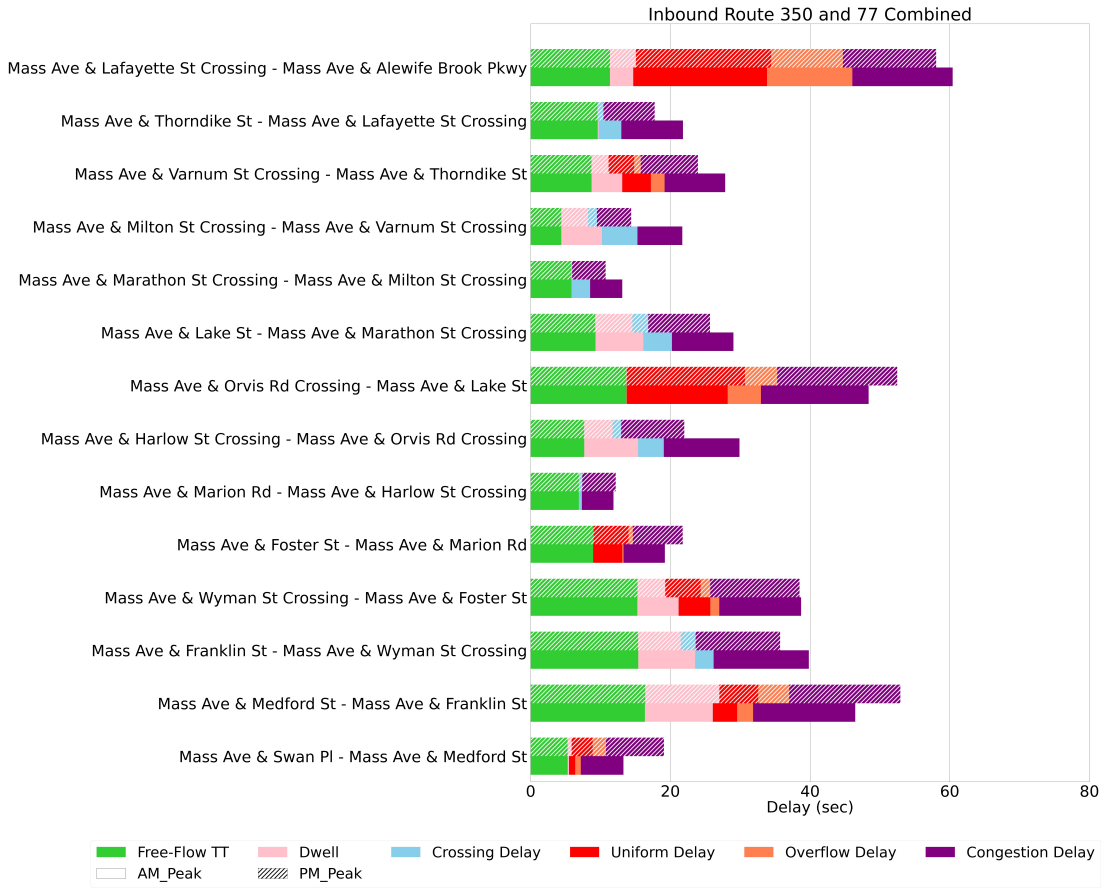
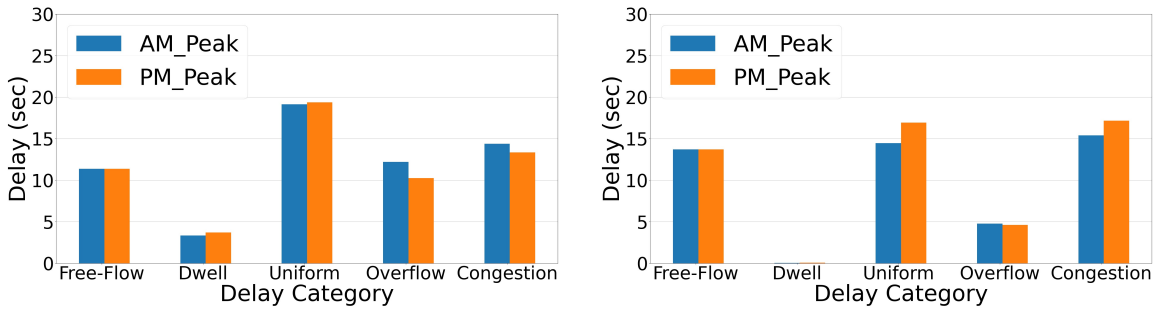


Figure 3-19: Delay analysis results of aggregating inbound Route 77 and Route 350 trips along the study corridor.

help quantify how much the bus-only lane had helped reduce congestion delay within the Lafayette-Alewife segment. Such a study is discussed in detail in the following chapter regarding before-after studies of bus improvement projects.

Another example given by the Orvis-Lake segment as shown in Figure 3-20b reveals that the main contributor to the greater PM travel time than AM are uniform delay and congestion delay, while the overflow delay observed during the two peak periods are very similar. This result suggests that the agency could investigate strategies such as retiming of signals or transit signal priority to target reducing the uniform delay during the PM peak.



(a) Lafayette-Alewife (b) Orvis-Lake

Figure 3-20: AM and PM travel times along the segment decomposed into different categories.

3.6 Conclusion

In this chapter, detailed methods are proposed to conduct quantitative analysis of bus delays by taking advantage of vehicle trajectories built from bus heartbeat data. By aligning the complete vehicle trajectories constructed from data of multiple bus trips of one or many routes onto the same spatial-temporal reference frame, vehicle travel times along the route can be decomposed into different types of delay, including dwell time at bus stops, crossing delay at pedestrian crossings, signal uniform delay, signal overflow delay, and congestion delay.

There are several limitations with the research presented in this chapter. First, the categorization of dwell time heavily relies on the availability of AVL data. Although it is often the case that the AVL data and heartbeat data are recorded by the same device and thus should be simultaneously available, this might not be true for cases where heartbeat data comes from GPS devices that are not associated with the AVL system. In such cases, the task of delineating dwell activities from other stopping activities could become challenging at near-side stops, where the dwell activities can be easily characterized as signal delay. Secondly, the categorization of dwell time and signal delay does not take into account the impact of traffic signal on dwell times. At near-side bus stops where buses can be stopping for the red signal while waiting for passengers to finish alighting and boarding, it is possible that part of signal delay is contained in the observed dwell time. Thirdly, although the methodology

allows for the analysis of multiple bus routes serving the same corridor, the boarding and alighting patterns of different bus routes may be substantially different due to the route pattern. For analysis that rely on precise categorization of dwell patterns on a bus corridor, further considerations of such difference need to be taken into account. Lastly, the analysis of signal delay does not account for situations when the road segment downstream of a traffic signal is blocked thus preventing vehicles from proceeding even when the signal is green. In this case, the delay incurred upstream of the traffic signal should be attributed to downstream congestion, but would be misclassified as signal delay instead. This limitation, however, can be overcome if historical traffic signal time is available.

The methodology proposed in this chapter enables analysts to inspect the operations of not just any single transit route, but also any transit corridor served by multiple routes. The result of the delay analysis offers insight into the magnitude of different types of delay encountered by bus vehicles in each time period, therefore allowing analysts to strategize over the type and scale of improvement projects to target specific delay. In addition to the decomposed travel times, the segment-based analysis also allows for the visualization of stopping activity scatter plots, which can be a useful tool for understanding the temporal and spatial distribution of queues. Furthermore, the decomposed travel times offers the unique opportunity to conduct more in-depth studies that could unveil how different operational characteristics contribute to different types of delay.

Chapter 4

Evaluating Transit Improvement Strategies

4.1 Introduction and Motivation

The management of public transit systems is often an iterative process, as agencies constantly plan, operate, evaluate, and adjust elements of the network to meet the ever-changing passenger demand while providing satisfying service. The preceeding chapters delved into the mechanisms of using heartbeat data to extract information about bus vehicle operations. This chapter focuses on how one can use the operational information of buses to evaluate the effectiveness of transit improvement strategies.

Common strategies that agencies deploy to improve the quality and reliability of bus services include the installation of bus-only lanes, transit signal priority, queue jump lanes, stop relocation, off-board fare collection, etc. Depending on the targeted issue being addressed, one or a combination of strategies are chosen by the agency, who would analyze and communicate the costs and benefits as well as the impacts of the project to the community. Following project completion, the agency would then conduct before-after analyses to evaluate the effectiveness of the chosen strategies.

In many before-after studies of transit improvement projects aiming at improving transit speed and reliability, analysts often quantify the effectiveness of the strategy by analyzing the change of specific metrics such as travel time, on-time performance,

or headway adherence from before to after the treatment strategies are implemented. These before-after studies are observational as opposed to experimental because improvement projects are usually carried out at selected locations experiencing suboptimal service by agencies with available funding. As such, it is impractical to conduct randomized control trials for any particular type of transit treatment, and so there is no one-size-fits-all formula that can be derived to generalize the effectiveness of a treatment. To understand the benefit of an improvement project, analysts must conduct before-after studies on a case-by-case basis and rely on the limited data collected during specific time intervals before and after project completion.

The conventional method that is widely used by transit analysts in before-after studies is the naive method, which involves simply taking the difference or percent difference between the selected metric (e.g., travel time) observed in the before and after periods. This method, however, is valid only if certain assumptions are true, but such constraints are not often clearly stated in the studies or understood by analysts. In the following section, discussions are provided regarding the naive method as well as its assumptions. In addition, a second method, namely the "comparison group method", which relies on more relaxed assumptions, is presented.

4.2 Methodologies

Both the naive method and the comparison group method are discussed thoroughly in the classic book in the field of road safety authored by Ezra Hauer [10]. In principle, the observational before-after studies in road safety are very similar to those in traffic operations, as in both cases, the task at hand is to compare the factual with the counterfactual. In road safety, analysts evaluate how a certain treatment improves or exacerbates safety by comparing the observed accident rates at study locations after treatment with the prediction of what the accident rates would have been had the treatment not been implemented. In transit operations, analysts quantify the effectiveness of transit improvement projects by comparing the observed metric (e.g., travel time) after treatment with the prediction of what the metric would have been

had the improvement strategy not been in place.

To keep the discussion relevant to transit operations, the following discussion will use "travel time" as the metric of interest, although the methodologies could be applied similarly to many other metrics.

4.2.1 Basic Elements of Before-After Studies of Travel Time

The Definition of "Travel Time"

Before diving into the technical discussion of the naive method, it is important to clarify what "travel time" means. Depending on the type of treatment applied to the transit corridor, the definition could vary. For example, in queue jump and transit signal priority projects, analysts may be interested in knowing how much travel time is saved in traveling between stops - in this case, "travel time" is defined as the time from vehicles departing the upstream bus stop of the signal to arriving at the downstream bus stop. Alternatively, in bus-lane projects where stretches of bus-only lanes are installed along the bus route, analysts may wish to know the change in travel time through the bus-lane corridor, so "travel time" in this case could mean the time between the vehicle reaching the beginning and the end of the bus lane.

In the latter case of bus-lane analysis, analysts should decide whether dwell time at bus stops should be excluded from travel time. Since bus lanes are installed to improve vehicles' travelling speed in between stops, some may argue that dwell time should be excluded from the analyzed travel time, and only improvements in running time should be used in the evaluation of bus-lane effectiveness. Others, however, may argue that improving bus speeds also leads to improved bus reliability, which consequently impacts dwell activities at bus stops, so dwell times should be included in the analysis. In any case, it would be critical for analysts to clearly define what "travel time", or any other chosen metric, means both spatially and temporally, before carrying out before-after studies and reporting the results to the public.

The Factual and the Counterfactual

As mentioned previously, the essence of observational before-after studies is to compare the factual with the counterfactual. By definition, the factual is the realized or observed travel time collected from the study location after the treatment is put in place. The counterfactual, on the other hand, is the expected or predicted travel time that buses would have experienced had the original configuration stayed and no treatment was installed.

Assume that if one were to collect a large number of travel times from bus trips that operated along a defined section of the route in a particular time period, then each travel time value follows a distribution in the location-scale family parameterized by mean μ and variance σ^2, i.e. travel time $T \sim Dist(\mu, \sigma^2)$. Let $\mathbf{K} = [K_1, K_2, ..., K_{n_k}]^T$ be a random sample of travel times from a distribution with mean κ and variance σ_k^2 collected at the study location before treatment (i.e., in the "before" period), $\mathbf{P} = [P_1, P_2, ..., P_{n_p}]^T$ be a random sample ($\sim Distr(\pi, \sigma_p^2)$) of predicted travel times completed by trips in the "after" period had there been no treatment (i.e., the counterfactual), and $\mathbf{L} = [L_1, L_2, ..., L_{n_l}]^T$ be a random sample ($\sim Distr(\lambda, \sigma_l^2)$) of travel times collected after the treatment is put in place (i.e., the factual). In evaluating the effectiveness of a treatment, the task at hand is to evaluate how the treatment changed the expected travel time at the study location. Therefore, the most relevant statistics in the analysis are the expected values of travel times, which can be estimated using the random samples collected from the information provided by the heartbeat data. Furthermore, the variance of each estimator for the expected travel times can also be estimated from the observed data. The notation for the different types of travel times, their expected values, as well as the estimators for expected values and their variances are summarized in Table 4.1.

"Before" and "After" Periods

The terms "before" period and "after" period are used to describe the time periods before and after the treatment is installed in which bus travel time data is collected for

Table 4.1: Random samples of travel times, their expected values, estimators for expected values and the variances of estimators for each time period

Time Period	Travel Time at the Study Location	Expected Travel Time	Estimator of expected travel time	Variance of estimator
Before	**K**	κ	$\hat{\kappa}$	$\text{Var}[\hat{\kappa}]$
After (prediction)	**P**	π	$\hat{\pi}$	$\text{Var}[\hat{\pi}]$
After	**L**	λ	$\hat{\lambda}$	$\text{Var}[\hat{\lambda}]$

the purpose of before-after studies. The selection of these time periods should follow the principle that within each time period, the travel time of each trip is a random variable from the same distribution. For example, in a before-after study evaluating the effectiveness of a bus lane in reducing bus travel time during the AM peak, the "before" period should be such that the travel times of all AM bus trips within the study corridor in this time period are independently and identically distributed. The selection of time periods could either follow the rule of thumb, e.g. by selecting only weekday trips within the same season of the year, or be determined using statistical tests that verify whether multiple groups of travel time distributions are identical at a certain confidence level.

Measures of Effectiveness (MOEs)

It is obvious that only **K** and **L** can be observed from the real-world operations of bus vehicles, while **P** is unknown and needs to be estimated. The main difference between the naive method and the comparison group method is exactly this, namely how the statistics around **P** can be obtained. Regardless of the method used, the effectiveness of a treatment can be measured using the following metrics, as proposed by Hauer [10] for evaluating road safety projects and adapted by the author to fit in the context of transit improvement projects:

1. The reduction in the expected travel time during the "after" period (i.e. the difference between the predicted travel time had there been no treatment and

the expected travel time after the treatment has been put in place), is

$$\delta = \pi - \lambda. \tag{4.1}$$

Use $\hat{\pi}$ and $\hat{\lambda}$ to estimate π and λ, δ can then be expressed as:

$$\hat{\delta} = \hat{\pi} - \hat{\lambda}. \tag{4.2}$$

2. Under the assumption that the travel times observed in the "after" period are independent from those in the "before" period, the variance of $\hat{\delta}$ is

$$\mathrm{Var}[\hat{\delta}] = \mathrm{Var}[\hat{\pi}] + \mathrm{Var}[\hat{\lambda}]; \tag{4.3}$$

3. The ratio of the expected travel time after treatment to the predicted travel time in the same "after" period had there been no treatment is

$$\theta = \lambda/\pi. \tag{4.4}$$

An unbiased estimator for θ, obtained from Taylor series expansion evaluated at the means of $\hat{\lambda}$ and $\hat{\pi}$, is

$$\hat{\theta} = \frac{\hat{\lambda}}{\hat{\pi}}/(1 + \frac{\mathrm{Var}[\hat{\pi}]}{\pi^2}). \tag{4.5}$$

4. The variance of $\hat{\theta}$ is

$$\mathrm{Var}[\hat{\theta}] \approx (\frac{\lambda}{\pi})^2(\frac{\mathrm{Var}[\hat{\lambda}]}{\lambda^2} + \frac{\mathrm{Var}[\hat{\pi}]}{\pi^2})/(1 + \frac{\mathrm{Var}[\hat{\pi}]}{\pi^2})^2. \tag{4.6}$$

5. The percent reduction in travel time is then

$$\gamma = 100 \times (1 - \theta). \tag{4.7}$$

Intuitively, the treatment is effective if $\delta > 0$, or $\theta < 1$.

4.2.2 The Naive Method

Assumptions

In the naive method, the key assumption is that the operations of buses during the "after" period had there been no treatment would be exactly the same as in the "before" period. In other words, the counterfactual expected travel time during the "after" period, if no treatment was applied to the study location, can be predicted using the expected travel time during the "before" period.

An illustration of this assumption is shown in Figure 4-1. No matter what time trend there might have been prior to the treatment placement, the expected travel time in the "after" period without treatment is estimated as being exactly the same as the travel time "before" treatment, or $\hat{\pi} = \hat{\kappa}$.

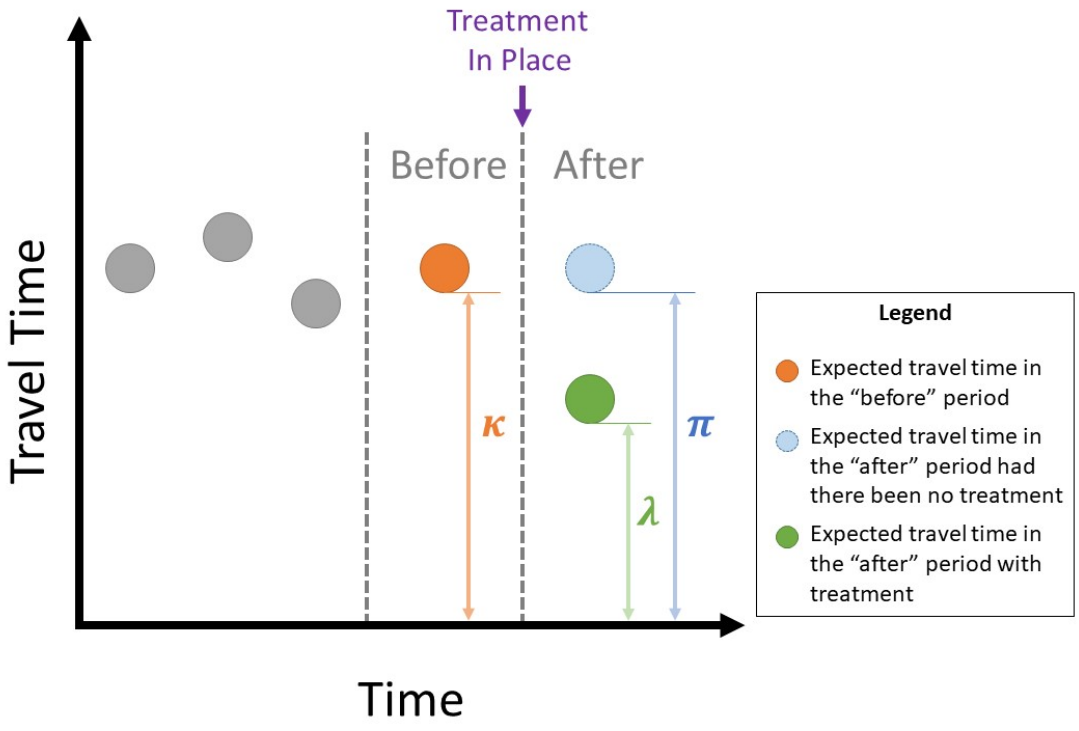


Figure 4-1: Illustration of the naive method.

Calculating MOEs

In order to calculate the MOEs listed in Section 4.2.1, samples of travel times need to be collected from the "before" and "after" periods. Suppose the analyst obtained the travel times of n_k bus trips during the "before" period, namely $k_1, k_2, ..., k_{n_k}$, and of n_l trips during the "after period, namely $l_1, l_2, ..., l_{n_l}$. For each $i = 1, 2, ..., n_k$, k_i is the realization of the random variable K_i; similarly for each $j = 1, 2, ..., n_l$, l_j is the realization of L_j.

Since the expected value of the sample mean is equal to the population mean, an unbiased estimator for the expected travel time in each period is the mean of all travel times recorded in the corresponding random sample. Take the travel time in the "before" period as an example, it can be trivially shown that the sample mean of all K_i's, denoted by $\bar{K}$, is an unbiased estimator for the population mean κ:

$$\begin{aligned} \mathrm{E}[\bar{K}] &= \mathrm{E}[\frac{K_1 + K_2 + ... + K_{n_k}}{n_k}] \\ &= n_k \frac{\mathrm{E}[K_1]}{n_k} \\ &= \mathrm{E}[K_1] = \kappa. \end{aligned} \tag{4.8}$$

Therefore,

$$\hat{\kappa} = \bar{K} = \frac{K_1 + K_2 + ... + K_{n_k}}{n_k}. \tag{4.9}$$

The variance of $\hat{\kappa}$ calculated using the unbiased estimator for σ_k^2, i.e. $s^2[K]$, is

$$\begin{aligned} \mathrm{Var}[\hat{\kappa}] &= \mathrm{Var}[\frac{K_1 + K_2 + ... + K_{n_k}}{n_k}] \\ &= \frac{\mathrm{Var}[K_1] + \mathrm{Var}[K_2] + ... + \mathrm{Var}[K_{n_k}]}{n_k^2} \\ &= \frac{s^2[K]}{n_k} \\ &= \frac{\sum_{i=1}^{n_k}(K_i - \bar{K})^2}{n_k * (n_k - 1)}. \end{aligned} \tag{4.10}$$

Similarly, the estimator for λ and its variance can be expressed as:

$$\hat{\lambda} = \bar{L} = \frac{L_1 + L_2 + ... + L_{n_l}}{n_l}. \tag{4.11}$$

$$\text{Var}[\hat{\lambda}] = \frac{\sum_{i=1}^{n_l}(L_i - \bar{L})^2}{n_l * (n_l - 1)}. \tag{4.12}$$

Since the naive study assumes that $\hat{\pi} = \hat{\kappa}$, the four metrics presented in Equation 4.2, 4.3, 4.5, 4.6 can be calculated as follows:

1. The reduction in the expected travel time

$$\begin{aligned} \hat{\delta} &= \hat{\pi} - \hat{\lambda} \\ &= \bar{K} - \bar{L}; \end{aligned} \tag{4.13}$$

2. The variance of reduction of the expected travel time

$$\begin{aligned} \text{Var}[\hat{\delta}] &= \text{Var}[\hat{\pi}] + \text{Var}[\hat{\lambda}] \\ &= \text{Var}[\hat{\kappa}] + \text{Var}[\hat{\lambda}] \end{aligned} \tag{4.14}$$

3. The ratio of the expected travel times

$$\hat{\theta} = \frac{\bar{L}}{\bar{K}} / (1 + \frac{\text{Var}[\hat{\kappa}]}{\bar{K}^2}) \tag{4.15}$$

4. The variance of the ratio of the expected travel times

$$\text{Var}[\hat{\theta}] \approx (\frac{\bar{L}}{\bar{K}})^2 (\frac{\text{Var}[\hat{\lambda}]}{\bar{L}^2} + \frac{\text{Var}[\hat{\kappa}]}{\bar{K}^2}) / (1 + \frac{\text{Var}[\hat{\kappa}]}{\bar{K}^2})^2. \tag{4.16}$$

4.2.3 The Comparison Group Method

Assumptions

The comparison group method relaxes the strong assumption in the naive method that the counterfactual travel time in the "after" period is exactly the same as in

the "before period". The method takes into account the change in travel time due to factors not related to the treatment through the use of a comparison group.

While the treatment group is the stretch of the bus route that has been retrofitted with a certain type of treatment, the comparison group is a different portion of the route along which buses operate almost identically as along the portion in the treatment group prior to the treatment placement. As shown in Figure 4-2, the travel time of the treatment group changes in the same way as the comparison group. Therefore, the ratio of travel time of the comparison group during the "after" period to the "before" period is used to predict the counterfactual travel time of the treatment group.

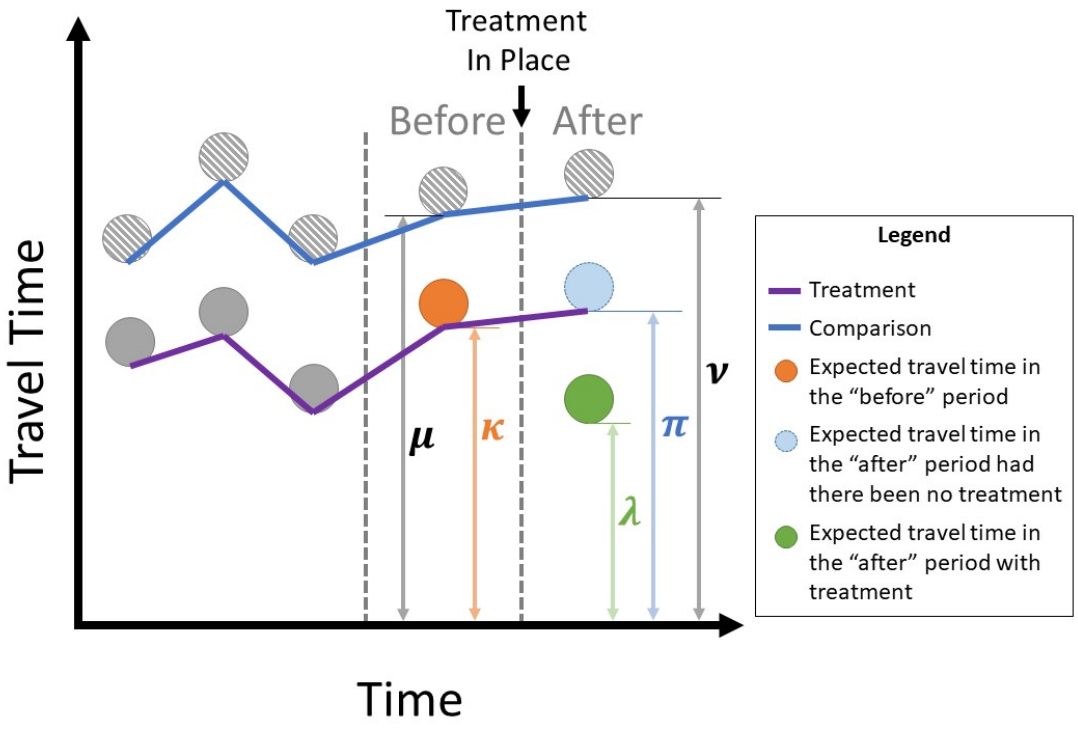


Figure 4-2: Illustration of the comparison group method.

Let $\mathbf{M} = [M_1, M_2, ..., M_{n_m}]^T$ be a random sample of travel times from a distribution with mean μ and variance σ_m^2 collected at the locations in the comparison group during the "before" period, and $\mathbf{N} = [N_1, N_2, ..., N_{n_n}]^T$ be a random sample ($\sim Distr(\nu, \sigma_n^2)$) of travel times collected of the comparison group in the "after" pe-

riod. The notation for the random sample and expected values of both the treatment and the comparison groups are summarized in Table 4.2.

Table 4.2: Travel times random samples and the corresponding population means used in the comparison group method

Time Period	Treatment Group	Comparison Group
Before	**K**, κ	**M**, μ
After (prediction)	**P**, π	-
After	**L**, λ	**N**, ν

Estimating μ, ν and variances of the estimators

Just as shown in the naive method, the estimators for the expected travel times of the comparison group can be obtained from the corresponding sample mean:

$$\hat{\mu} = \bar{M} = \frac{M_1 + M_2 + ... + M_{n_m}}{n_m}; \tag{4.17}$$

$$\hat{\nu} = \bar{N} = \frac{N_1 + N_2 + ... + N_{n_n}}{n_n}. \tag{4.18}$$

The variance of each estimator can be calculated as:

$$\text{Var}[\hat{\mu}] = \frac{\sum_{i=1}^{n_m}(M_i - \bar{M})^2}{n_m * (n_m - 1)}; \tag{4.19}$$

$$\text{Var}[\hat{\nu}] = \frac{\sum_{i=1}^{n_n}(N_i - \bar{N})^2}{n_n * (n_n - 1)}. \tag{4.20}$$

Selection of Comparison Group

As defined previously, the travel time ratio is the ratio of the expected travel time in the "after" period to that in the "before" period. Therefore, the travel time ratio for the comparison group is:

$$r_C = \frac{\nu}{\mu}; \tag{4.21}$$

Similarly, the travel time ratio for the treatment group is

$$r_T = \frac{\pi}{\kappa}. \tag{4.22}$$

The goal in selecting a comparison group is to find one such that $r_C = r_T$. If one were to evaluate a time series of comparison ratio Ω's, where $\Omega = r_C/r_T$, then by definition,

$$\Omega = \frac{r_C}{r_T} = \frac{\kappa\nu}{\lambda\mu}. \tag{4.23}$$

Let $\mathbf{\Omega} = [\Omega_1, \Omega_2, ..., \Omega_{n_o}]^T$ be a sequence of comparison ratios calculated from the expected travel times, namely κ, λ, μ and ν, observed in a series of "before" and "after" time intervals defined by a sliding time window. The sliding time window essentially divides the time series into $n_o + 1$ smaller intervals as it moves along, creating n_o pairs of "before" and "after" time intervals. For each "before" and "after" pair t ($t = 1, 2, ..., n_o$) in the time series, the comparison ratio can be expressed as:

$$\Omega_t = \frac{\kappa_t\nu_t}{\lambda_t\mu_t}. \tag{4.24}$$

Since the expected travel times in Equation 4.23 are not known, an unbiased estimator $\hat{\Omega}_t$ for Ω_t that uses the estimators for the expected travel times is found as:

$$\hat{\Omega}_t = \frac{\hat{\kappa}_t\hat{\nu}_t}{\hat{\lambda}_t\hat{\mu}_t}/(1 + \frac{\mathrm{Var}[\hat{\lambda}_t]}{\lambda_t^2} + \frac{\mathrm{Var}[\hat{\mu}_t]}{\mu_t^2}). \tag{4.25}$$

In the context of estimating Ω_t's, each Ω_t represents the true comparison ratio between the travel time ratios of the comparison group and the treatment group for the time interval t. Since the true value of Ω_t can not be known, $\hat{\Omega}_t$ is used an as unbiased estimator for the corresponding Ω_t. For each pair of "before" and "after" time intervals, t, there would be an Ω_t that is unknown but can be estimated by $\hat{\Omega}_t$. Furthermore, each Ω_t in $\mathbf{\Omega}$ is not just unknown, but also follows a distribution with unknown parameters mean ω and variance σ_ω^2.

Formulaically, the population mean of each Ω_t, ω, can be estimated using the

sample mean:

$$\hat{\omega} = \mathrm{E}[\mathbf{\Omega}] = \frac{\hat{\Omega}_1 + \hat{\Omega}_2 + ... + \hat{\Omega}_{n_o}}{n_o} = \bar{\Omega}. \tag{4.26}$$

Through differential analysis and the argument that the sample variance $s^2[\mathbf{\Omega}]$ is a combination of the randomness in each expected travel time term and σ_ω^2 it self, Hauer [10] established that

$$\hat{\sigma_\omega^2} = \max\{0, s^2[\mathbf{\Omega}] - (\frac{\mathrm{Var}[\hat{\kappa}]}{\kappa^2} + \frac{\mathrm{Var}[\hat{\lambda}]}{\lambda^2} + \frac{\mathrm{Var}[\hat{\mu}]}{\mu^2} + \frac{\mathrm{Var}[\hat{\nu}]}{\nu^2})\}. \tag{4.27}$$

The goal in selecting a good comparison group is to find one for which the distribution of Ω_t's centers around 1 with little variance, i.e., $\hat{\omega} \to 1$ and $\hat{\sigma_\omega^2} \to 0$.

Estimating r_T, π and $\mathrm{Var}[\hat{\pi}]$

Once a "before" period and an "after" period are selected, following the rule that the travel time of all trips in each period are independent and identically distributed, the value of μ, ν and λ can be immediately obtained by calculating the sample mean of travel times collected in each time period for each group.

In the process of selecting a comparison group, an estimate of the comparison ratio ω given by $\bar{\Omega}$ is obtained. By the definition of r_T,

$$r_T = \frac{\nu}{\mu\omega}. \tag{4.28}$$

The value of r_T can then be estimated using the unbiased estimator

$$\hat{r_T} = \frac{\nu}{\mu\omega}/(1 + \frac{\mathrm{Var}[\hat{\mu}]}{\mu^2} + \frac{\sigma_\omega^2}{\omega^2}). \tag{4.29}$$

It is worth noting that, as was discussed above, an ideal comparison group would have ω close to 0 and σ_ω close to 1. As $\omega \to 1$ and $\sigma_\omega \to 0$, the unbiased estimator $\hat{r_T} = \frac{\nu}{\mu\omega}/(1 + \frac{\mathrm{Var}[\hat{\mu}]}{\mu^2} + \frac{\sigma_\omega^2}{\omega^2}) \to \frac{\nu}{\mu}/(1 + \frac{\mathrm{Var}[\hat{\mu}]}{\mu^2})$, which is an unbiased estimator for $r_C = \frac{\nu}{\mu}$. In other words, the travel time ratio of an ideal comparison group can be used to estimate the travel time ratio of the treatment group, or $\hat{r_T} = \hat{r_C}$.

The variance of the estimator $\hat{r_T}$ is

$$\begin{aligned} \hat{\text{Var}}[\hat{r_T}] &= r_T^2(\frac{\text{Var}[\hat{\nu}]}{\nu^2} + \frac{\text{Var}[\hat{\mu}]}{\mu^2} + \frac{\sigma_\omega^2}{\omega^2}); \\ \Rightarrow \frac{\text{Var}[\hat{r_T}]}{r_T^2} &= \frac{\text{Var}[\hat{\nu}]}{\nu^2} + \frac{\text{Var}[\hat{\mu}]}{\mu^2} + \frac{\sigma_\omega^2}{\omega^2}. \end{aligned} \tag{4.30}$$

The expected travel time in the "after" period had there been no treatment is then

$$\hat{\pi} = \hat{r_T}\hat{\kappa}. \tag{4.31}$$

The variance of the estimator $\hat{\pi}$ is

$$\hat{\text{Var}}[\hat{\pi}] = \pi^2(\frac{\text{Var}[\hat{\kappa}]}{\kappa^2} + \frac{\text{Var}[\hat{r_T}]}{r_T^2}). \tag{4.32}$$

Calculating MOEs

To calculate the MOEs, the task is still to estimate the difference or the ratio between the expected travel time before and after treatment and the corresponding variances, in other words, $\hat{\delta}$, $\text{Var}[\hat{\delta}]$, $\hat{\theta}$ and $\text{Var}[\hat{\theta}]$. The main difference between the MOEs calculated using the comparison group method and the naive method is how π and its variance are estimated.

The four metrics presented in Equation 4.2, 4.3, 4.5, 4.6 can be calculated as follows:

1. The reduction in the expected travel time

$$\begin{aligned} \hat{\delta} &= \hat{\pi} - \hat{\lambda} \\ &= \hat{r_T}\hat{\kappa} - \hat{\lambda}. \end{aligned} \tag{4.33}$$

2. The variance of the reduction in the expected travel time

$$\begin{aligned} \hat{\text{Var}}[\hat{\delta}] &= \hat{\text{Var}}[\hat{\pi}] + \text{Var}[\hat{\lambda}] \\ &= \pi^2(\frac{\text{Var}[\hat{\kappa}]}{\kappa^2} + \frac{\text{Var}[\hat{\nu}]}{\nu^2} + \frac{\text{Var}[\hat{\mu}]}{\mu^2} + +\frac{\sigma_\omega^2}{\omega^2}) + \text{Var}[\hat{\lambda}]. \end{aligned} \tag{4.34}$$

3. The ratio of the expected travel times

$$\begin{aligned}\hat{\theta} &= \frac{\hat{\lambda}}{\hat{\pi}}/(1+\frac{\text{Var}[\hat{\pi}]}{\pi^2}) \\ &= \frac{\hat{\lambda}}{\hat{\pi}}/(1+\frac{\text{Var}[\hat{\kappa}]}{\kappa^2}+\frac{\text{Var}[\hat{r_T}]}{r_T^2})\end{aligned} \tag{4.35}$$

4. The variance of the ratio of the expected travel times

$$\hat{\text{Var}}[\hat{\theta}] = (\frac{\lambda}{\pi})^2(\frac{\text{Var}[\hat{\lambda}]}{\lambda^2}+\frac{\text{Var}[\hat{\kappa}]}{\kappa^2}+\frac{\text{Var}[\hat{r_T}]}{r_T^2})/(1+\frac{\text{Var}[\hat{\kappa}]}{\kappa^2}+\frac{\text{Var}[\hat{r_T}]}{r_T^2})^2. \tag{4.36}$$

4.3 Case Study: Bus-Only Lanes

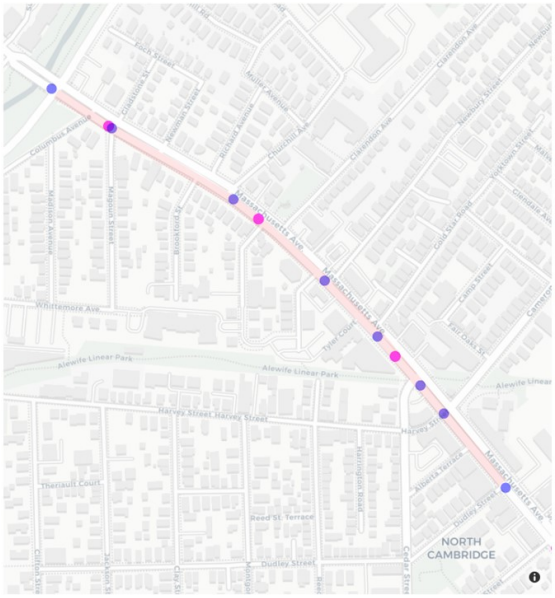
This case study aims at examining the effectiveness of a stretch of bus lanes installed on Massachusetts Avenue between Alewife Brook Parkway and Dudley Street in Cambridge, Massachusetts in both directions. Both the naive method and the comparison group methods will be used to evaluate how much travel time is saved by bus trips after the bus-lane treatment is put in place.

4.3.1 Study Area

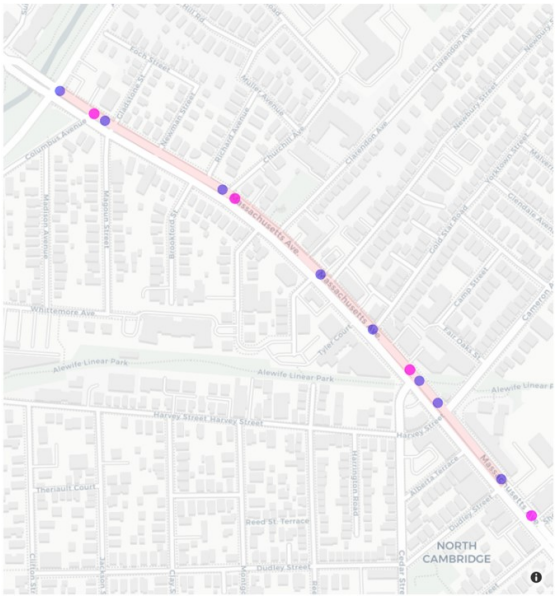
The study area is on Massachusetts Avenue between Alewife Brook Parkway and Dudley Street. Only one bus route, Route 77 operated by the MBTA, operates along this stretch of the roadway. Illustrations of the study area as well as traffic signals and bus stops along each direction of the study area are shown in Figure 4-3.

4.3.2 Study Time Period

The inbound bus lane only operates during the AM peak hours, while the outbound one operates all day. For the purpose of this study, the operations of outbound Route 77 during the PM peak (4 - 7 pm) will be examined. According to the City of Cambridge, the installation of the bus lane took place between November and December, 2021 [23]. To compare the travel times before and after the installation of



(a) Inbound bus lane marked by signals (blue) and stops (magenta)



(b) Outbound bus lane marked by signals (blue) and stops (magenta)

Figure 4-3: Bus lanes in both directions of Massachusetts Avenue between Alewife Brook Parkway and Dudley Street.

the bus lane, a two-month "before" period and two-month "after" period are selected. Specifically, the "before" period is defined as all Tuesdays through Thursdays in September and October of 2021, and "after" period the same days in September and October of 2022.

4.3.3 Data Source

To understand the change in performance of Route 77 before and after the bus-lane installation, travel times of each bus trip within the study area during the "before" and "after" period are collected. Different from traditional before-after studies, such as the one commissioned by the City of Cambridge [7], that use stop-to-stop AVL data as the data source for transit vehicle travel time, this study uses the trip travel time extracted from the heartbeat data. Although AVL data allows for extraction of travel time between bus stops, it does not offer analysts the ability to understand the travel time of buses between specific intersections where bus lanes usually begin and end. As discussed in the previous chapters, heartbeat data allows analysts to know

the distance into trip at any point in time, and equivalently the time into trip at any location along the bus route. Therefore, the travel time of each bus trip between the Alewife Brook Parkway and Dudley Street intersections during the study periods are extracted from the heartbeat data and used in the following analysis.

4.3.4 Outbound 77

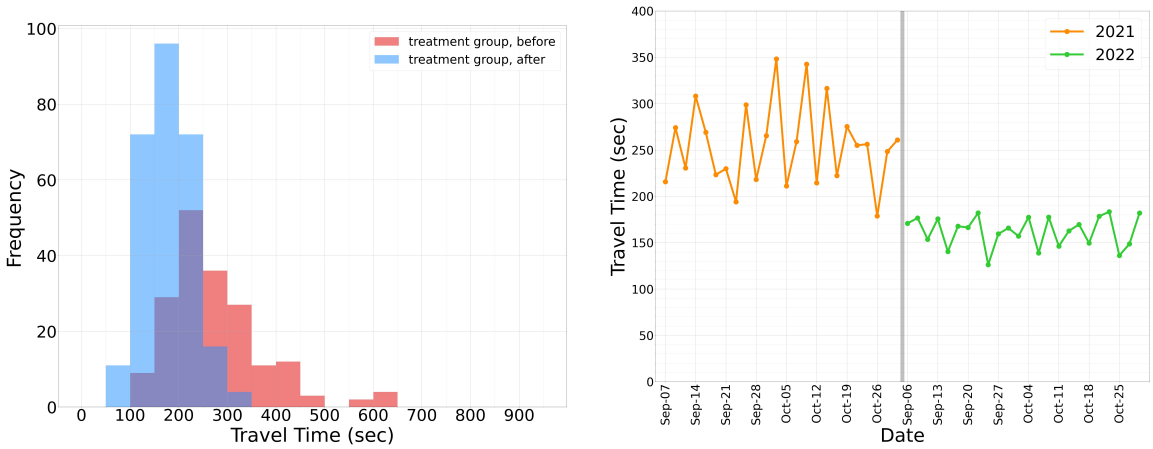
The Naive Method

In the two-month "before" period, the travel time of 191 outbound trips from Dudley Street to Alewife Brook Parkway is collected during the PM peak on Tuesdays, Wednesdays and Thursdays in September and October 2021. In the "after" period, the outbound travel time of 271 PM-peak trips is collected in the same study area on Tuesdays through Thursdays in September and October 2022.

The distributions of the "before" and "after" travel times are shown in Figure 4-4a. An alternative way to visualize the change of travel time is by plotting the average daily PM-peak outbound travel time, as shown in Figure 4-4b. From visual inspection of both figures, it is clear that the travel time in the "after" period is smaller than in the "before" period with reduced variability. The question to be answered remains how one can quantify this reduction in travel time.

Using the naive method, the parameters in Table 4.1 can be estimated as follows:

$$\begin{aligned}
\hat{\pi} = \hat{\kappa} = \bar{K} &= \frac{K_1 + K_2 + ... + K_{185}}{185} = 274 \text{ sec} \\
\text{Var}[\hat{\pi}] = \text{Var}[\hat{\kappa}] &= \frac{(K_1 - 274)^2 + (K_2 - 274)^2 + ... + (K_{191} - 274)^2}{185 * (185 - 1)} = 53.8 \\
\hat{\lambda} = \bar{L} &= \frac{L_1 + L_2 + ... + L_{271}}{271} = 180 \text{ sec} \\
\text{Var}[\hat{\lambda}] &= \frac{(L_1 - 180)^2 + (L_2 - 180)^2 + ... + (L_{271} - 180)^2}{271 * (271 - 1)} = 9.30
\end{aligned} \tag{4.37}$$



(a) Distribution of outbound trip travel times within the study area before and after the installation of the bus lane.

(b) Daily average outbound travel time within the study area in Sep & Oct 2021 and 2022.

Figure 4-4: Comparison of outbound travel times within the study area before and after the bus-lane treatment is put in place.

The MOEs can then be calculated as follows:

$$
\begin{aligned}
\hat{\delta} &= \hat{\pi} - \hat{\lambda} = 93.4 \text{ sec} \\
\sigma_{\hat{\delta}} &= \sqrt{\text{Var}[\hat{\pi}] + \text{Var}[\hat{\lambda}]} = \sqrt{53.8 + 9.30} = 7.95 \\
\hat{\theta} &= \frac{\hat{\lambda}}{\hat{\pi}} / (1 + \frac{\text{Var}[\hat{\pi}]}{\pi^2}) = \frac{180}{274} / (1 + \frac{53.8}{274^2}) = 0.658 \\
\sigma_{\hat{\theta}} &= \frac{\lambda}{\pi} / (1 + \frac{\text{Var}[\hat{\pi}]}{\pi^2}) \sqrt{\frac{\text{Var}[\hat{\lambda}]}{\lambda^2} + \frac{\text{Var}[\hat{\pi}]}{\pi^2}} \\
&= \frac{180}{274} / (1 + \frac{53.8}{274^2}) \sqrt{\frac{9.30}{180^2} + \frac{53.8}{274^2}} = 0.0209.
\end{aligned}
\tag{4.38}
$$

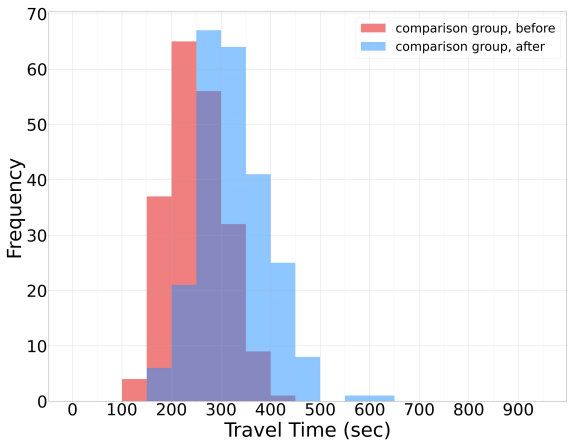
Therefore, the results from the naive method suggest that the outbound bus lane in the study area reduced outbound bus travel time by 93.4 ± 7.95 sec, or reduced the travel time by $34.2\% \pm 2.09\%$ during the PM peak.

The Comparison Group Method

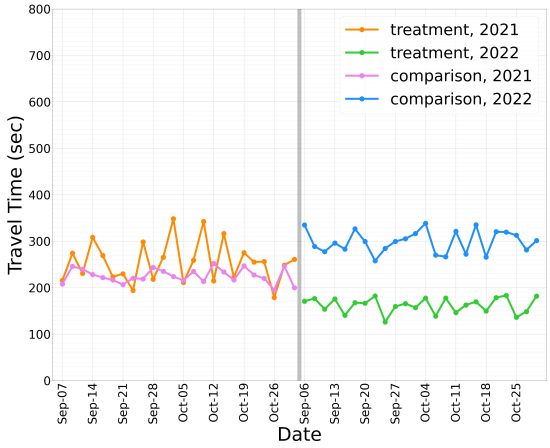
Initially, it was thought that a 0.45-mile stretch of northbound Massachusetts Avenue upstream of the 0.45-mile treatment (study) area should be selected as the comparison group area. However, by definition, the operations of buses within the comparison

group area should not be impacted by the treatment. Therefore, the comparison group is chosen from the offset (by 0.35 miles) section of the route upstream of the study area rather than immediately upstream or downstream to avoid any operational impact of the bus-lane treatment on the comparison group. In the "before" period, the outbound travel time of 204 PM-peak trips are collected in the comparison group area on Tuesdays through Thursdays in September and October 2021, and for "after" period, 234 trip travel times are collected in the same months in 2022.

The distributions of outbound trip travel times in the comparison group area during the "before" and "after" periods are shown in Figure 4-5a. The outbound trip travel times collected for PM-peak trips averaged by date are plotted as shown in Figure 4-5b for both the treatment and comparison groups.



(a) Distribution of outbound trip travel times in the comparison group area before and after the installation of the bus lane.



(b) Daily average outbound travel time within the treatment area and the comparison group area in Sep & Oct 2021 and 2022.

Figure 4-5: Distribution of outbound travel times within the area before and after the bus-lane treatment is put in place and the change in daily average PM-peak travel time of both groups.

From visual inspection of both figures, it appears that time series of travel times observed from the comparison group resembles that of the treatment group. Furthermore, the average travel time of the comparison group observed during the "after" period is larger than in the "before" period, indicating that the average travel time of the treatment group in the "after" period would follow the same trend. Such observation shows that the assumption used in naive method, that the counterfactual

travel time within the study area in the "after" period had there been no treatment is identical to the travel time in the "before" period, no longer holds.

Using the naive method, the parameters in Table 4.2 can be estimated as follows:

$$\begin{aligned}
\hat{\kappa} = \bar{K} &= \frac{K_1 + K_2 + ... + K_{185}}{185} = 274 \text{ sec} \\
\text{Var}[\hat{\kappa}] &= \frac{(K_1 - 274)^2 + (K_2 - 274)^2 + ... + (K_{191} - 274)^2}{185 * (185 - 1)} = 53.8 \\
\hat{\lambda} = \bar{L} &= \frac{L_1 + L_2 + ... + L_{271}}{271} = 180 \text{ sec} \\
\text{Var}[\hat{\lambda}] &= \frac{(L_1 - 180)^2 + (L_2 - 180)^2 + ... + (L_{271} - 180)^2}{271 * (271 - 1)} = 9.30 \\
\hat{\mu} = \bar{M} &= \frac{M_1 + M_2 + ... + L_{204}}{204} = 251 \text{ sec} \\
\text{Var}[\hat{\mu}] &= \frac{(M_1 - 251)^2 + (M_2 - 251)^2 + ... + (M_{204} - 251)^2}{204 * (204 - 1)} = 14.8 \\
\hat{\nu} = \bar{N} &= \frac{N_1 + N_2 + ... + N_{234}}{234} = 325 \text{ sec} \\
\text{Var}[\hat{\nu}] &= \frac{(N_1 - 325)^2 + (N_2 - 325)^2 + ... + (N_{234} - 325)^2}{234 * (234 - 1)} = 20.8
\end{aligned} \tag{4.39}$$

To evaluate how well the comparison group represents the treatment group, a sliding time window is applied to the time series of average daily PM trip travel times to estimate the mean and variance of the comparison ratio, i.e. ω and σ_ω.

Table 4.3: Average travel times in a time series of "before" and "after" intervals.

Date	$\hat{\kappa}$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\nu}$	$\hat{\text{Var}}[\lambda]$	$\hat{\text{Var}}[\mu]$	$\hat{\Omega}$
2021-09-07	216	274	208	246	14.3	3.00	0.93
2021-09-08	274	231	246	240	19.1	5.47	1.16
2021-09-09	231	308	240	228	14.7	12.8	0.712
2021-09-14	308	269	228	222	33.6	10.2	1.11
...	...	...	...	...	...	...	...
2021-10-21	256	179	220	195	2.95	3.25	1.27
2021-10-26	179	248	195	246	45.3	3.72	0.909
2021-10-27	248	261	246	200	37.6	12.6	0.772
mean	255	256	225	226	-	-	1.034
variance	1971	1989	244	241	-	-	0.03671

From Table 4.3, it is clear that the mean of $\hat{\Omega}$'s, i.e. ω, is close to 1 and variance

close to 0. Therefore, the change in travel time of the comparison group can be used to predict the change in travel time of the treatment group in the "after" period had there been no treatment.

The travel time ratio of the treatment group as well as its normalized variance are

$$\begin{aligned} \hat{r_T} &= \frac{\nu}{\mu\omega}/(1+\frac{\text{Var}[\hat{\mu}]}{\mu^2}+\frac{\sigma_\omega^2}{\omega^2}); \\ &= \frac{325}{251*1.034}/(1+\frac{14.8}{251^2}+\frac{0.03671}{1.034^2}) = 1.21 \end{aligned} \tag{4.40}$$

$$\begin{aligned} \frac{\text{Var}[\hat{r_T}]}{r_T^2} &= \frac{\text{Var}[\hat{\nu}]}{\nu^2}+\frac{\text{Var}[\hat{\mu}]}{\mu^2}+\frac{\sigma_\omega^2}{\omega^2} \\ &= \frac{20.8}{325^2}+\frac{14.8}{251^2}+\frac{0.03671}{1.034^2} = 0.0347. \end{aligned} \tag{4.41}$$

The expected travel time of the treatment group in the "after" period and the normalized variance can then be estimated as follows:

$$\begin{aligned} \hat{\pi} &= \hat{r_T}\hat{\kappa} \\ &= 1.21*274 = 330sec; \\ \frac{\hat{\text{Var}}[\hat{\pi}]}{\pi^2} &= \frac{\text{Var}[\hat{\kappa}]}{\kappa^2}+\frac{\text{Var}[\hat{r_T}]}{r_T^2} \\ &= \frac{53.8}{274^2}+0.0347 = 0.0354. \end{aligned} \tag{4.42}$$

The MOEs can then be calculated as follows:

$$\begin{aligned} \hat{\delta} &= \hat{\pi}-\hat{\lambda} = 330-180 = 150 \text{ sec} \\ \sigma_{\hat{\delta}} &= \sqrt{\text{Var}[\hat{\pi}]+\text{Var}[\hat{\lambda}]} = \sqrt{0.0354*330^2+9.30} = 62.2 \\ \hat{\theta} &= \frac{\hat{\lambda}}{\hat{\pi}}/(1+\frac{\text{Var}[\hat{\pi}]}{\pi^2}) = \frac{180}{330}/(1+0.0354) = 0.527 \\ \sigma_{\hat{\theta}} &= \frac{\lambda}{\pi}/(1+\frac{\text{Var}[\hat{\pi}]}{\pi^2})\sqrt{\frac{\text{Var}[\hat{\lambda}]}{\lambda^2}+\frac{\text{Var}[\hat{\pi}]}{\pi^2}} \\ &= \frac{180}{330}/(1+0.0354)\sqrt{\frac{9.30}{180^2}+0.0354} = 0.0997. \end{aligned} \tag{4.43}$$

Therefore, the results from the comparison group method suggests that the outbound bus lane in the study area reduced outbound bus travel time by 150 ± 62.2

sec, or reduced the travel time by $47.3\% \pm 9.97\%$ during the PM peak.

The MOE results from both the naive method and the comparison group method are summarized in Table 4.4.

Table 4.4: Summary of MOEs from the naive method and the comparison group method.

Method	Travel Time Savings (sec)		Percentage of TT Savings (%)	
	mean	std. dev.	mean	std. dev.
Naive	93.4	7.95	34.2%	2.09%
Comparison Group	150	62.2	47.3%	9.97%

4.3.5 Naive Method vs. Comparison Group Method

Comparing the MOEs calculated from the naive method with those from the comparison group method for the northbound bus lane utilized by outbound Route 77, it can be noted that the travel time saving calculated using the comparison group method is much larger than the naive method, and the variance of each MOE in the comparison group method is also greater than those in the naive method.

As discussed previously, the main difference between the two methods lies within the assumption made for the counterfactual travel time, i.e. what the travel time in the studied segment would have been in the "after" period had there been no treatment. In the comparison group method, an upstream section of the route considered representative of the treated section is chosen to account for the change in travel time from the "before" to the "after" period that is not related to the treatment. As shown in Figure 4-5, the travel time of the comparison group increased from September and October 2021 to the same months in 2022, thus indicating that the the travel time of the treatment group in the "after" period (i.e. the counterfactual travel time) would have been larger than those observed in the "before" period had there been no bus lane installed in 2022. The naive method, on the other hand, assumes that the counterfactual travel time in the "after" period is exactly the same as those in the "before" period. Therefore, it makes sense that the travel time saving calculated

using the comparison group method is larger than that from the naive method in the case of outbound Route 77.

Upon inspection of Equation 4.41, the dominant term is found to be $\frac{\sigma_\omega^2}{\omega^2}$, which captures the variance in the series of comparison ratios between the treatment and comparison groups calculated using a sliding time window. In other words, the closer σ_ω^2 is to 0 and ω is to 1, the smaller the term $\frac{\sigma_\omega^2}{\omega^2}$ is. Since the series of comparison ratio captures how closely the variation of travel times in the comparison group follows those in the treatment group, this observation aligns with the intuition that the more identical the comparison group is to the treatment group, the smaller the variance of the MOEs would be. Therefore, it is worth spending time selecting a comparison group that yields the smallest $\frac{\sigma_\omega^2}{\omega^2}$ in order to minimize the variances.

4.4 Conclusion

In this chapter, in-depth discussions are provided on conducting observational before-after studies of transit improvement projects. By adapting classical methods commonly used in road safety studies to fit in the context of transit analysis, the author presented detailed methodologies for how one can evaluate the effectiveness of transit improvement projects by comparing the travel time "before" and "after" the introduction of a treatment. In particular, the naive method and the comparison method are explored for calculating MOEs such as the absolute travel time saving and the percent change in travel time due to the treatment.

The main difference between the two methods lies within the assumption made about the counterfactual travel time, i.e. the travel time in the "after" period had there been no treatment. The naive method calculates MOEs by taking the difference between the observed travel time in the "before" period and that from the "after" period, as the method relies on the assumption that the observed travel time in the "before" period would be a perfectly good representation of the counterfactual travel time in the "after" period had there been no treatment. The comparison group method, on the other hand, assumes that the change in travel time from "before" to

"after" the treatment is solely due to factors unrelated to the treatment. The method estimates the counterfactual travel time in the treatment group by using a comparison group that adequately represent the operations of the treatment group in the before period. The change in travel time from "before" to "after" in the comparison group, which would be due to factors other than the treatment, is translated to the treatment group. The effectiveness of the treatment can then be calculated using the estimated counterfactual travel time.

A case study is presented that evaluates the effectiveness of a 0.45-mile bus-only lane on northbound Massachusetts Avenue in Cambridge, MA. The case study illustrated that the naive method and the comparison method can produce different MOEs and that the MOEs calculated from the comparison method are generally larger than those from the naive method. The question of which particular method the agency should use to evaluate and report the effectiveness of the improvement project should be based on consideration of whether the assumption made in the naive method holds. In cases where transit operations during the "before" period are known to be very different from those in the "after" period, such as when the "before" period is during the COVID-19 pandemic and the "after" period is after the pandemic, the assumption made by the naive method may no longer hold. In such cases, the comparison method should be used to trade variance for accuracy.

Several limitations are present with this research. One limitation is that the research only explored evaluating the effectiveness of the improvement project by comparing travel times, but did not exploring other metrics such as on-time performance. The effectiveness of a transit improvement project can be demonstrated by showing not just travel time savings, but also an increase in reliability. Another limitation is that the comparison groups method assumes that the operations of buses within the comparison group is not impacted by the treatment during the "after" period. Such assumption may not hold in cases where the comparison group is selected to be immediately adjacent to the treatment group, as the effects of the treatment could propagate upstream and downstream beyond the scope of the treatment itself. Further research is needed to address these limitations to improve the robustness of the

methodologies presented in this chapter.

Chapter 5

Conclusions and Recommendations

5.1 Summary

In this thesis, the author delved into the task of understanding transit operations using heartbeat data from three perspectives. First, a comprehensive approach is proposed to reconstructing transit vehicle trajectories. Then, a methodology is presented for plotting the trajectories of multiple trips in the same temporal-spatial reference frame for analyzing and categorizing transit delays. Lastly, two observational before-after study methods are proposed and compared for evaluating transit improvement strategies.

Chapter 2 discussed how to reconstruct the trajectories of bus vehicles from raw heartbeat data that is often noisy and recorded at inconsistent frequencies. The methodology developed allows for the creation of a continuous, monotonic, and differentiable trajectory that is validated against AVL data and a reasonable acceleration threshold. The LOCREG-PCHIP algorithm was found to be the most effective in achieving all three properties of an ideal vehicle trajectory. The resulting trajectories enable the extraction of bus location, speed, and acceleration at any point in time during a bus trip, offering valuable information for further transit performance analysis.

Chapter 3 introduced methods to conduct a quantitative analysis of bus delays using the constructed vehicle trajectories. By decomposing the travel times into

different types of delay, the methodology proposed in this chapter allows analysts to inspect operations on any individual single transit route or transit corridor served by multiple routes. This in-depth analysis offers insights into the magnitude and type of bus delay, aiding in strategic decision-making for targeted transit improvement projects.

Chapter 4 discussed methodologies for conducting before-after studies to evaluate the effectiveness of transit improvement projects using travel times extracted from heartbeat data. Two main methods, namely the naive method and the comparison method, with different assumptions about counterfactual travel times are explored and compared. While the naive and comparison methods for before-after studies each have their strengths and limitations, the choice between them should be made based on the specifics of the context and the nature of the transit operations during the "before" and "after" periods.

In conclusion, this thesis offers an integrated approach to better understand transit operations using bus heartbeat data. By developing methodologies for trajectory reconstruction, delay analysis, and before-after study for evaluating transit improvement projects, it provides a framework for transit authorities to identify operational issues, devise improvement strategies, and assess their effectiveness by leveraging the wealth of information provided by the heartbeat data.

5.2 Recommendations

Based on the research in this thesis, the following recommendations are proposed:

1. For transit agencies that have the capability of archiving heartbeat data recorded at an average frequency less than 10 seconds, the methodology proposed in Chapter 2 can be used to reconstruct a complete trajectory of every bus trip. The bus vehicle trajectory can be plotted on a time-space diagram for qualitative analysis of bus operations.

2. By combining complete vehicle trajectories constructed from heartbeat data

with GTFS data and AVL data, agencies can employ the methodology in Chapter 3 to identify and quantify various types of delay of transit routes or corridors. This would allow transit planners and decision-makers to prioritize projects that target specific types of bus delay.

3. Transit analysts can incorporate the methodologies presented in Chapter 4 for conducting before-after studies of transit improvement projects. The choice between the naive method and the comparison method should be made based on the specific context and nature of transit operations during the "before" and "after" periods. This would allow for a more accurate assessment of the effectiveness of implemented strategies and help guide future decision-making.

5.3 Limitations

Some limitations are present with the methodologies proposed in each chapter, and can be summarized as follows.

In reconstructing vehicle trajectory, the methodology heavily depends on the heartbeat data having an average of frequency of less than 10 seconds. The methodology also lacks guarantee for twice-differentiability to construct acceleration profiles. Moreover, the unavailability of historical speed and acceleration measurements from vehicles makes it difficult to validate the speed and accelerations calculated from the constructed trajectories.

In decomposing vehicle travel time and categorizing transit delays, the methodology relies on the availability of AVL data for dwell time categorization. Other limitations include the lack of consideration for the impact of traffic signal on dwell times, and the inability to account for downstream congestion when analyzing signal delay.

In evaluating transit improvement projects using before-after studies, the methodologies presented focus only on travel time comparisons without considering other metrics such as on-time performance or headway adherence. Furthermore, the comparison method, albeit it overcomes the limitation of not accounting for change in

travel time due to factors other than the treatment presented in the naive method, assumes that the operations of buses in the comparison group are unaffected by the treatment during the "after" period, which may not hold true in some cases.

5.4 Future Work

The research presented in this thesis builds a solid foundation for analyzing raw heartbeat data and extracting useful information about bus operations. It is expected that an analyst can follow the methodology presented in each chapter to completely replicate the result and apply to other transit routes or networks. Nonetheless, there are many opportunities to build upon this research and develop more applications for using heartbeat data.

Firstly, additional algorithms can be explored that improves upon the LOCREG-PCHIP algorithm for constructing ideal vehicle trajectories. Using the validation method presented in this thesis, researchers can explore algorithms that take in the time and distance into trip data and output trajectories that can be evaluated against the benchmark values comparing against AVL data.

Secondly, in-depth analysis can be carried out to help agencies better understand how each type of quantified delay can be translated to operational strategies targeting the specific delay. It is also worth identifying the costs associated with each type of delay as well as the entity or agency responsible for mitigating the delay. Further, the ability to retrieve a complete vehicle acceleration profile could allow for investigation of driver aggressiveness on various bus routes or corridors.

Thirdly, the detailed analysis of transit delays enabled by this research allows for the causal analysis of how various road and operational features are related to the different types of delay. In addition to the causal analysis, other methods for conducting observational before-after studies can be explored to address the limitation in the naive and the comparison group methods. Furthermore, before-after analysis of operational metrics other than travel time, such as reliability, speed, as well as the different type of delay obtained from travel time decomposition, can be conducted in

order to build a more robust and holistic framework for evaluating the effectiveness of transit improvement projects.

Bibliography

[1] Google Transit. Gtfs realtime overview. URL https://developers.google.com/transit/gtfs-realtime. [Date accessed: 04.19.2023].

[2] Randolph W Hall and Nilesh Vyas. Buses as a traffic probe: Demonstration project. *Transportation Research Record*, 1731(1):96–103, 2000.

[3] Frederick W Cathey and Daniel J Dailey. Transit vehicles as traffic probe sensors. *Transportation Research Record*, 1804(1):23–30, 2002.

[4] Christy Coghlan, Sina Dabiri, Brian Mayer, Mitch Wagner, Eric Williamson, Michael Eichler, and Naren Ramakrishnan. Assigning bus delay and predicting travel times using automated vehicle location data. *Transportation Research Record*, 2673(3):624–636, 2019.

[5] Eric Lind and Joseph Reid. Diagnosing obstacles to speed and reliability with high-resolution automatic vehicle locator data: bus time budgets. *Transportation research record*, 2675(12):464–474, 2021.

[6] Zack Aemmer, Andisheh Ranjbari, and Don MacKenzie. Measurement and classification of transit delays using gtfs-rt data. *Public Transport*, pages 1–23, 2022.

[7] City of Cambridge. Post-implementation travel time analysis of massachusetts avenue, . URL https://www.cambridgema.gov/-/media/Files/Traffic/2021/20221118_northmassavepostimplementationanalysis.pdf. [Date accessed: 04.10.2023].

[8] Kunihiro Sakamoto, Chandana Abhayantha, and Hisashi Kubota. Effectiveness of bus priority lane as countermeasure for congestion. *Transportation Research Record*, 2034(1):103–111, 2007. doi: 10.3141/2034-12. URL https://doi.org/10.3141/2034-12.

[9] Hyung Jin KIM. Performance of bus lanes in seoul: Some impacts and suggestions. *IATSS Research*, 27(2):36–45, 2003. ISSN 0386-1112. doi: https://doi.org/10.1016/S0386-1112(14)60142-4. URL https://www.sciencedirect.com/science/article/pii/S0386111214601424.

[10] E. (Ezra) Hauer. *Observational before–after studies in road safety : estimating the effect of highway and traffic engineering measures on road safety*. Pergamon, Oxford, OX, U.K. ;, 1st ed. edition, 1997. ISBN 0080430538.

[11] B. Nagy and A. Kelly. Trajectory generation for car-like robots using cubic curvature polynomials. In *Proceedings of 3rd International Conference on Field and Service Robotics (FSR '01)*, pages 479 – 490, June 2001.

[12] Mireille Elhajj and Washington Ochieng. Impact of new gps signals on positioning accuracy for urban bus operations. *The Journal of Navigation*, 73(6): 1284–1305, 2020. doi: 10.1017/S0373463320000272.

[13] Valhalla. Valhalla - open source routing engine for openstreetmap. URL `github.com/valhalla`. [Date accessed: 07.18.2022].

[14] OpenStreetMap contributors. OpenStreetMap. URL `https://www.openstreetmap.org`. [Date accessed: 07.18.2022].

[15] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980. ISSN 00361429. URL `http://www.jstor.org/stable/2156610`.

[16] F. N. Fritsch and J. Butland. A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific and Statistical Computing*, 5 (2):300–304, 1984. doi: 10.1137/0905021. URL `https://doi.org/10.1137/0905021`.

[17] William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979. ISSN 01621459. URL `http://www.jstor.org/stable/2286407`.

[18] Marietta Kirchner, Patric Schubert, and Christian T Haas. Characterisation of real-world bus acceleration and deceleration signals. *Journal of Signal and Information Processing*, 2014, 2014.

[19] *Transit capacity and quality of service manual.* TCRP report, 165. Transportation Research Board, Washington, D.C, 3rd ed. edition, 2013. ISBN 9780309283441.

[20] *Travel time data collection handbook.* Office of Highway Information Management, Federal Highway Administration, U.S. Dept. of Transportation, Washington, DC, 1998 - 03.

[21] AP Akgungor and AGR Bullen. Analytical delay models for signalized intersections. *ITE journal*, 69(2):12–12, 1999. ISSN 0162-8178.

[22] Fo Vo Webster. Traffic signal settings. Technical report, 1958.

[23] City of Cambridge. Bus trips on north mass ave faster, more consistent with bus lanes, . URL `https://www.cambridgema.gov/Departments/trafficparkingandtransportation/News/2023/01/northmassavebuslanes`. [Date accessed: 04.10.2023].