May
2018

Measuring Left-Behinds on Subway

Charles D. Baker, *Governor*

Karyn E. Polito, *Lieutenant Governor*

Stephanie Pollack, *MassDOT Secretary & CEO*

U.S. Department of Transportation
Federal Highway Administration

massDOT
Massachusetts Department of Transportation

Technical Report Document Page

1. Report No. n/a	2. Government Accession No. n/a	3. Recipient's Catalog No. n/a
4. Title and Subtitle Measuring Left-Behinds on Subway		5. Report Date April 2018
		6. Performing Organization Code
7. Author(s) Eric J. Gonzales, Andronikos Keklikoglou, and Charalampos Sipetas		8. Performing Organization Report No. UMTC
9. Performing Organization Name and Address University of Massachusetts Amherst 130 Natural Resources Way, Amherst, MA 01003		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. ISA# 92312
12. Sponsoring Agency Name and Address Massachusetts Department of Transportation Office of Transportation Planning Ten Park Plaza, Suite 4150, Boston, MA 02116		13. Type of Report and Period Covered Final Report May 2018 to April 2018
		14. Sponsoring Agency Code n/a
15. Supplementary Notes n/a		

16. Abstract

The MBTA uses performance measures to monitor its service and measure improvement. This project supports the development of measures that track the customer experience instead of the performance of vehicles. Current measures are based on fare card records and assume that passengers are able to get on the first available vehicle that arrives at a stop or station. There is not currently a way to measure people left behind on subway platforms when vehicles are too full to board. This report presents the development of methods to measure or estimate the number of passengers that are left behind when vehicles are too crowded to board and the distribution of waiting times experienced by passengers, accounting for left-behind passengers. In addition to making use of existing vehicle location data, the study includes evaluation of two potential technologies for measuring passengers: automated passenger counting from surveillance video feeds, and tracking of MAC addresses from Bluetooth and Wi-Fi-enabled wireless devices. The occurrence of at least one passenger being left behind can be estimated with 90% accuracy, and the total number left-behind passengers during a rush period can be estimated within 10%. Challenges and opportunities for the future are identified.

17. Key Word left-behind, train passengers, subway platforms	18. Distribution Statement n/a		
19. Security Classif. (of this report) unclassified	20. Security Classif. (of this page) unclassified	21. No. of Pages 70	22. Price n/a

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

This page left blank intentionally.

Measuring Left-Behinds on Subway

Final Report

Prepared By:
Eric J. Gonzales, Ph.D.
Principal Investigator
(413) 545-0685; gonzales@umass.edu

Andronikos Keklikoglou
Graduate Researcher
akeklikoglou@umass.edu

Charalampos Sipetas
Graduate Researcher
csipetas@umass.edu

University of Massachusetts Amherst
130 Natural Resources Way, Amherst, MA 01003

Prepared For:
Massachusetts Department of Transportation
Office of Transportation Planning
Ten Park Plaza, Suite 4150
Boston, MA 02116

July 2018

This page left blank intentionally.

Acknowledgements

Prepared in cooperation with the Massachusetts Department of Transportation, Office of Transportation Planning, and the United States Department of Transportation, Federal Highway Administration.

The Project Team would like to acknowledge the efforts of Laurel Paget-Seekins and Arthor Prolosch, Massachusetts Bay Transportation Authority, for providing oversight and data necessary for this project. Staff from the Security & Emergency Management Department provided the surveillance video feeds for the times and locations of experimental data collection. Carpenters with the Building & Station Maintenance Department mounted the wireless data collection devices. Staff from the Office of Performance Management and Innovation provided access to the Research Database, extensive explanations of the MBTA's rail operations, and assisted in coordinating communications with the other departments within the MBTA.

The Project Team would also like to acknowledge the efforts of Gabriel Sherman, MassDOT Transportation Planner, for additional project support, as well as the University of Massachusetts Transportation Center and Melissa Paciulli, Manager of Research and Development, and Matt Mann, Research Program Coordinator.

This page left blank intentionally

Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Expansion
ACF	Aggregate Channel Feature
AFC	Automated Fare Collection
APC	Automatic Passenger Counter
AVL	Automatic Vehicle Location
FIFO	First-In-First-Out
FHWA	Federal Highway Administration
GPS	Global Positioning System
HOG	Histogram of Oriented Gradients
MAC	Media Access Control
MAE	Mean Absolute Error
MassDOT	Massachusetts Department of Transportation
MBTA	Massachusetts Bay Transportation Authority
ODX	Origin-Destination-Transfer
R-CNN	Regional Convolutional Neural Network
RMSE	Root Mean Squared Error
SDP	Service Delivery Policy
SPR	State Planning and Research Funds
SVM	Support Vector Machine
TCQSM	Transit Capacity and Quality of Service Manual
TTR	Train-Tracking Record
YOLO	You Only Look Once

This page left blank intentionally.

1 Introduction

This study of transit passengers being left behind when subway trains are too crowded to board was undertaken as part of the Massachusetts Department of Transportation (MassDOT) Research Program. This program is funded with Federal Highway Administration (FHWA) State Planning and Research (SPR) funds.

The Massachusetts Bay Transportation Authority (MBTA) uses performance measures to monitor its service and measure improvement. Measures that track the customer experience instead of the performance of the vehicles are being developed. Crowding is a challenge on buses and trains in the MBTA system, and overcrowded vehicles lead to passengers being left behind when vehicles are too full to board. Automated passenger counters provide an indication of levels of crowding on buses, but there is currently no way to measure people left behind on subway platforms. Passenger wait times for subway and light rail services are reported using estimated travel demand data from the Origin-Destination-Transfer inference model. These measures assume that passengers are able to get on the first available train, leaving left-behind passengers uncounted. Data on passengers being left behind on station platforms when trains are too crowded to board would improve both the customer-oriented performance measures and measures of capacity needed to meet demand. This project focuses on measuring passengers left-behind on the MBTA's subway system.

The primary objective of this project was the following:

> Develop a method to measure the occurrence of passengers being left behind at a station when a train is too full to board.

There are two types of approaches for addressing this problem. First, measurements can be made from direct observations of passengers. Second, the occurrence of left-behinds can be inferred by demand patterns and vehicle capacities in order to estimate the locations and times that overcrowding is likely to lead to left-behinds. The proposed research activities in this project will seek to address both, leading to two secondary objectives that support the first:

1. Identify the quality of measurements that can be made using different technologies for observing passengers. Specifically, the technologies of interest for this study were automated passenger counts using existing surveillance video feeds and detection of media access control (MAC) addresses of wireless devices in stations.

2. Identify the potential to fuse observations with existing data sources to improve estimates of the occurrence of trains leaving behind passengers, the number of passengers being left behind, and experienced delays.

The analysis and metrics developed through this study are intended to correspond with the reliability and passenger comfort standards in the MBTA Service Delivery Policy [1] to the fullest extent possible. The focus of this study is on addressing crowding and left-behinds on a rail system which regularly experiences problems with overcrowding during peak hours.

This page left blank intentionally.

2 Research Methodology

The research approach for this study consisted of three main parts: review of existing data and technologies, collection of data, and analysis to estimate the occurrence and number of left-behind passengers due to overcrowded trains.

Section 2.1 presents a review of the literature related to crowding in transit systems and the indicators that are used across transit agencies and by the MBTA to measure the passenger experience. The relevant available data that MBTA currently collects is also summarized in this section. A review of technologies that are used for counting pedestrians is then presented in Section 2.2.

Sections 2.3 through 2.6 describe the methods for collecting data. This starts with analyzing existing data sources to identify the locations and times of day that passengers are most likely to experience being left behind. Then, the details for manual data collection, automated video counts, and device detection are described.

Finally, Section 2.7 presents the modeling methods that are used to make the most accurate estimates of the occurrence of trains leaving behind passengers and the number of passengers that are left behind.

2.1 Review of Available Data and Models

This review is organized in two parts. First, a brief review of the literature and existing methods related to monitoring and managing crowding on transit is presented. Then a description is presented of the data that the MBTA has available, which can be categorized as either direct observations or inferences that are based on modelling and manipulation of the raw data.

2.1.1 Crowding on Public Transportation Systems

Crowding and the MBTA Service Delivery Policy

Crowding is a major issue for transit systems all over the world, including trains operated by the MBTA in the Boston region. The 2017 MBTA Service Delivery Policy (SDP) establishes quantifiable standards for service planning and accessibility [1]. The SDP presents passenger comfort standards as part of a specific effort to track passenger-centric measures that reflect the experience of system customers. Two types of standards are related to crowding on the system: passenger comfort is affected by the space available to each passenger within vehicles, and reliability (in terms of passenger wait time) is affected by passengers' ability to board vehicles that arrive. When vehicles are too crowded, passenger comfort is compromised and the likelihood of a passenger being left behind at a stop or station increases.

For rapid transit rail service, all lines are run on short headways, and reliability is measured as the percentage of passengers who wait for less than or equal to the scheduled headway. Quantifying reliability by this definition requires information about passenger arrivals to stations in addition to vehicle departures. Passengers expect trains to run frequently, so they tend to arrive at stations at a constant rate. If passengers are always able to board the next train, the reliability of the system can then be estimated as the percentage of total time during which elapsed time since the previous departure is less than the scheduled headway. In the case of crowding that causes some passengers to be left behind, the wait time that passengers experience may exceed the headway between trains, which would diminish reliability performance. According to Table 10 in the 2017 SDP, the reliability for rapid transit in 2016 was reported as 89%, but since it does not account for left-behind passengers, it over-estimates the true value.

Due to the limited data available for estimating crowding on rapid transit, the SDP does not specify a specific metric for rail comfort. With currently available data, the SDP refers to development of a provisional metric that compares station entries to the capacity of trains passing through each station. This would make use of data in the Rail Flow database (described in Section 2.1.3). Plans to procure vehicles with automatic passenger counters (APC) would provide direct measurement of vehicle occupancy that would allow for calculation of a passenger comfort metric similar to that of buses, but this equipment is not yet available.

Literature on Crowding

The Transit Capacity and Quality of Service Manual (TCQSM) [2] establishes some guidelines for measuring the quality of service of transit services, including availability, comfort, and convenience. Crowding affects some aspects of availability and all elements of comfort and convenience associated with the quality of service framework for fixed-route transit presented in the TCQSM:

Indicators of Availability

- **Frequency –** Crowding can cause vehicles to operate more slowly or passengers to be left behind at stations and stops, the consequence of which is an effective reduction in service frequency for users, which represents a limitation on the availability of transit service compared to an uncrowded system.
- **Service Span –** The hours of service in a day are *not typically affected by crowding*. An exception could be if crowding on the last run of the night prevents some people from being able to use the system during its hours of operation.
- **Access –** The design of physical infrastructure to allow access to all users, including passengers who may use wheeled mobility devices, is a requisite for accessible transit. Crowding can undermine this accessibility by blocking pathways or preventing a passenger from boarding a specific vehicle.

Indicators of Comfort and Convenience

- **Passenger Load –** Crowding is a direct reflection of high passenger loads, and there is a demonstrated relationship between crowding and passengers' perception of time. More crowded vehicles increase the likelihood of passengers having to stand or squeeze into the compartments with other passengers. This deteriorates the quality of the passenger experience, and evidence shows that they perceive waiting and travel times in crowded conditions to be more onerous than in uncrowded conditions.
- **Reliability –** Crowding contributes to diminished reliability in terms of on-time performance and maintaining consistent headways, because boarding and alighting are delayed when there are many people in vehicles and at stations. A further consequence for reliability is that left-behind passengers essentially experience an extra headway of waiting time if they are unable to board the first vehicle that arrives.
- **Travel Time –** In addition to increasing the perception of travel time, actual travel times are also increased by crowding for the same reasons that crowding diminishes reliability. Vehicles operate more slowly, especially because of delays associated with boarding and alighting. Furthermore, passengers that are left behind experience longer total travel times due to the additional time they must wait to board a vehicle.

There are a number of ways that transit agencies seek to monitor and mitigate the effects of crowding in public transportation systems. From a monitoring perspective, several methods have been used to estimate the level of crowding on transit systems. Most focus on measuring the density of passengers, which is an objective measure of crowding [3, 4]. However, there are also subjective measures that are associated with the level of discomfort that have to do with passenger perceptions [5-7].

A number of studies have investigated the effect that crowding has on passenger perceptions and choices. Efforts to quantify the value that users put on crowding in terms of value of time or willingness to pay to avoid it is currently a very active research area [8-12]. Other studies have sought to determine the effect that crowding has on passengers' travel decisions. For example, evidence from Seoul, South Korea, suggests that crowding has an effect on path choice in networks that are large and connected enough to offer multiple paths between origin-destination pairs [13].

Most of the literature on crowding in transit systems focuses on the problem of passengers squeezing into vehicles or onto platforms from the perspective of passenger discomfort. To a limited extent research has addressed actions that a transit agency can take in response to information about crowding. Crowding also has demonstrated effects on operating speed, waiting time, and travel time reliability which in turn have effects on transit operations and the demand patterns on bus and rail systems [14]. A recent effort has been made to link transit crowding to decisions regarding time-dependent fares, service frequency, or investment in higher-capacity vehicles by accounting for the implicit cost that users experience by spending time standing in crowded vehicles compared to sitting comfortably [15]. This research provides some guidance for making decisions regarding demand management or increasing the frequency and size of vehicles. It does not provide any insight regarding reliability of operations or the effect of uneven headways.

Although crowding is known to be a major problem in transit systems throughout the world, very little attention has been given to the problem of passengers being left behind when vehicles are too full to board. This is in part because planning methods provide guidance for designing systems so that capacity is not exceeded [16]. In many real systems, overcrowding is a result of operations that have already maximized the length of trainsets and the frequency of train service, given physical limitations of platforms, tracks, and signal systems. As a result, knowledge of crowding costs may be useful for planning purposes and cost-benefit analysis [17-19].

Recent research aimed at estimating left-behinds has developed algorithms for systems with tap-in and tap-out smartcards, which are used in systems with multiple fare zones [20]. To date, a reliable method has not been developed to estimate or detect left-behind passengers in systems with only smartcard entry data, as is the case for the MBTA, as well as transit systems in Chicago and New York City. Some related research has sought to develop methods to infer passenger origin-destination patterns from entry-only data [21-24], but these methods have been developed with the primary objective of estimating passenger travel patterns and in-vehicle crowding rather than quantifying the occurrence of passenger left-behinds.

A related body of research investigates algorithms for tracking pedestrian movements within transit stations. The primary goal of these studies is to track individuals in order to understand how passengers move through entry points, ticketing areas, fare gates, and platforms. Based on level of service measures, methods have been developed to improve the planning process for the design of rail stations [25]. Related models have been developed to make use of multiple data sources to estimate origin-destination flows of passengers within a train station [26]. A number of studies have developed image processing tools to track pedestrians in video footage [27-30]. Many of these tools are limited to camera angles that are mounted high enough to view pedestrian activity from above and that have an unobstructed view of the areas where pedestrians move [31, 32]. While some methods can detect general density of passengers, tools that can reliably track pedestrians in order to directly observe left-behinds have not yet been developed.

2.1.2 Available Raw Data

The primary focus of this study is to investigate left-behinds on MBTA's rail system. There are three main sources of raw data related to passenger and vehicle movements in the MBTA system.

Automatic Fare Collection (AFC)

Automatic fare collection data is collected from the fare collection system at station fare gates and on-board buses and light rail vehicles. The AFC records are associated with events in which Charlie Cards are used to load value, pay a fare, or validate a pass. The data is partitioned by month and year, and includes records of Charlie Card transactions from individual fare cards as well as passes. Relevant AFC data that could potentially be useful for assessing crowding are:

- Unique identifier of the device that records the AFC event
- The station location of the device (e.g., fare gate, firebox, or ticket vending machine) that recorded the event
- The timestamp of the event
- Card/ticket serial number from the AFC system
- Type of transaction (e.g., top-up, validation, or fare deduction)

From this raw data, counts of passengers entering transit stations can be tracked over time based on the transactions' times and locations. The dataset includes good coverage of passengers entering fare gate-controlled stations on the red, orange, and blue lines. However, passengers are able to board inbound green line trains without necessarily validating a ticket, so some passengers are able to enter the system and make transfers without being counted.

The MBTA's rapid transit fare system charges a single fare for entry to the system, and passengers do not tap out when they leave the system. As a result, AFC records only account for station and vehicle entry, and there are no direct observations of exits.

Automatic Passenger Counter (APC)

Automatic passenger counters (APC) are devices that count the number of passengers boarding and alighting each vehicle. APC devices are not in widespread deployment on MBTA rail vehicles, so this is not a data source that can be reliably used for assessing crowding in the system.

Train Tracking (TTR)

The train-tracking system records the position of heavy rail vehicles as they move from track circuit to track circuit through the system. The analogous data for tracking bus positions on the network are reported through the Automatic Vehicle Location (AVL) systems. Since much of the heavy rail operations are in tunnels, track circuits are used to identify train locations. There is typically one track circuit associated with each station, and a few circuits between consecutive stations. The relevant TTR data for this study are:

- The timestamp of the train-tracking record
- Numeric code for heavy rail line
- Letter code for heavy rail line
- Numeric code identifying a trainset
- Latitude associated with track circuit
- Longitude associated with track circuit
- Unique identification number for track circuit
- The direction of train traffic on the track circuit
- The location type of the track circuit
- Name of station associated with the track circuit

The AVL data provides detailed data about vehicle movements in the system that can be compared against passenger data from the AFC data. From the AVL data, it is possible to piece together the progression of an individual vehicle along a line. It is also possible to look at the headways of departures from a specific station.

2.1.3 Models and Inferred Data

The raw data collected and logged by the MBTA contains extensive (although not complete) information about passenger entrances to rail stations and boarding buses at bus stops. It also contains comprehensive records of vehicle movements. By itself, this data is sufficient to count passenger entries and track performance of transit vehicles for schedule or headway adherence. In order to assess crowding, additional processing of the data is necessary to link records and infer travel patterns.

Origin-Destination-Transfer Model (ODX)

A model to link trip records and infer origin-destination and transfer patterns in the system has been developed to populate a database of ODX records. Inference models based on farecard data have been improved over the years. The most recent advances make use of dynamic programming to minimize generalized disutility for travelers, accounting for path-specific waiting time, in-vehicle time, and transfers [33]. The model identifies records from AFC that can be linked to infer transfers or return trip patterns. For example, a passenger using a Charlie Card to enter a rail station and later board a bus near a different rail station can be assumed to have used the rail system and then transferred to the bus. Another passenger who enters one rail station in the morning and enters a different rail station in the afternoon may be completing a round-trip commute, so the destination of the morning and afternoon trips can be inferred by linking the two trips. Through this method, the model infers values for 97% of trip origins, 75% of trip destinations, and 92% of transfers.

The ODX model is structured in three levels:

1. **Ride –** One ride; boarding and alighting one vehicle
2. **Stage –** One fare card tap; this could be a single ride, boarding a bus and riding to a destination stop to alight. This could also be a station entry that is followed by a ride on a train and then a gateless transfer to another train
3. **Journey –** One trip from origin to destination; this may consist of one or more rides and stages. For example, a multi-stage journey could include a first stage consisting of a ride on a bus and then a second stage consisting of entry to a rail station. The stages are each recorded by a separate tap (on the bus and at the fare gate), but a transfer from one mode or route to another may be required to complete a trip.

The ODX records are based on the raw data from AFC and AVL, but the dataset contains information related to journeys by inferring the destination and transfer locations and times associated with each origin. The relevant data from the ODX records are:

- Serial number of card, or arbitrary assigned number for cash transactions
- Location of the stop or station where fare transaction was recorded
- Timestamp of fare transaction

- The sequence of the journey for a specific card
- Sequence of the stage within the journey
- Total number of stages within the journey
- Recorded or inferred journey origin location
- Inferred journey destination location
- Timestamp when the stage starts, based on vehicle's departure time from origin stop
- Timestamp when the stage ends, based on vehicle's arrival at destination stop
- The route of the vehicle trip or the route of the station where the fare card was tapped
- The direction of the vehicle trip
- Code indicating if the origin was inferred, or the reason it was not inferred
- Code indicating if the destination was inferred, or the reason it was not inferred
- Code indicating if a transfer was inferred, or the reason it was not inferred
- The given or inferred origin of a ride, usually a bus stop or station platform
- The time at which the vehicle departed from the ride's origin
- The inferred destination of a ride, usually a bus stop or station platform
- The time at which the vehicle arrived at this ride's destination

The ODX data provides a comprehensive and useful view of travel patterns in the MBTA system. Although it appears at a glance to provide the same information as records from a tap-in and tap-out AFC would provide, it is important to be mindful of the assumptions on which inferences are based. Notably, for this study, inferred stages are based on the assumption that passengers are always able to board the next arriving vehicle. Therefore, destination times provide an optimistic estimate, assuming that crowding did not prevent a passenger from boarding the next arriving vehicle.

Rail Flow

The Rail Flow tool provides processed and aggregated data based on the ODX records. This data includes estimates of passenger boardings and alightings at stations for 15 minute increments. In this way, the ODX model provides valuable data for estimating the level of crowding in the system. The tool shows the variability of passenger flows between stations and provides an indication of locations and times that are likely to be experiencing the greatest crowding. However, Rail Flow does not provide an indication of left-behind passengers, because the ODX data is built on the assumption that passengers are not left behind.

Perhaps a subsampling of stage data could be extracted to consider only multi-stage journeys in which the start time of the second stage can be used to work backward to estimate when the previous stage likely ended. Comparing the estimate of stage end time to the passage of vehicles may provide a rough estimate of whether or not a passenger was left behind. This would not provide a comprehensive measure of the left-behinds problem.

2.1.4 Models and Inferred Data

Stations throughout the MBTA are equipped with surveillance cameras for security purposes. The placement of cameras has been designed to provide coverage for security purposes, and the view angles are not necessarily optimized for counting passengers on platforms. Variations in station architecture (e.g., side platforms vs. island platforms, columned stations with low station ceilings vs. open vaulted ceilings) create many different contexts for video observation. A challenge is that columns and curvature in the station limit how much of the platform, where passengers may be walking or waiting, is visible in a singleframe. The extensive placement of cameras, especially in recently renovated stations, provides multiple vantage points to observe platform crowding and vehicle boarding. Figure 2.1 shows examples of two surveillance feeds from the same station: one with an unobstructed view of the platform, and the other with the platform view blocked by many columns.

Figure 2.1: Two surveillance camera views in Maverick Station (Blue Line) showing a) an unobstructed platform view, and b) a platform view obstructed by many columns.

2.2 Review of Technologies

There are a number of technologies that can be used to observe pedestrians and pedestrian movements in an area. Two broad categories of technologies are considered: video technologies that use image processing to directly observe and track people, and device detection technologies that register the unique device address associated with Bluetooth, Wi-Fi, or cellular signals. Additional technologies for simple pedestrian counting exist, but they are of limited value for assessing the problem of left-behind passengers.

2.2.1 Video Analysis Technologies

Pedestrian counting and tracking technologies have been developed outside the field of public transportation. Applications range from counting pedestrians on sidewalks and street crossings to tracking pedestrian movements in shopping areas. In recent years, there have been a number of commercial video image processing tools developed in order to analyze pedestrian trajectories in stores and shopping centers in order to study shopping psychology and improve marketing practices. Examples include Placemeter and ShopperTrak.

Several types of algorithms exist for identifying and tracking pedestrians:

- **Pedestrian Counting –** Counting the number of people in a single frame returns the number of people in a given area. This can be used as a measure of crowding but does not give any indication of where people are going, and the total count only provides an indication of how long people have been waiting in an aggregate way if counts are compared across many frames.

- **Pedestrian Trajectory Tracing –** If an identified person is tracked from one frame to the next in a video feed, the movement of the person can be used to construct a trajectory of the person's movement in the field of view. This allows for counting of directional flows and also tracing the locations where people spend time. In the context of retail applications, this has been developed to study customer behavior. For transit stations, efforts to track pedestrians have been used to monitor passenger flows through station areas, but there is potential to watch passengers on a station platform from the time they arrive to the time that they are able to board a train.

- **Pedestrian Re-Identification –** It is common for a person's trajectory to traverse the fields of view of multiple cameras in the course of passing through a station or platform area. In order to construct a complete trajectory, pedestrians must be re-identified when they leave one camera view and enter another. In the case that the fields of view overlap, algorithms have been developed to stitch together trajectories based on re-identification at a common position and time. This process relies on highly reliable trajectory tracing capabilities as described above.

- **Person Identification/Facial Recognition –** A far more complicated problem than the tracking tasks listed above is to link the identity of a person in a video feed to another person in a feed at a different time and location. For example, if a person enters a station at ground level and then later appears on a station platform without full observation of the path in between, then identifying features of the person must be recorded and re-identified. The most sophisticated image processing tools seek to identify a person's identity based on facial recognition methods. With perfect reliability, this would allow for identification of passengers not only within different parts of the same station but also at different stations in the network. The challenge is that surveillance feeds are often not at sufficient resolution for facial recognition algorithms to work effectively. Furthermore, there are complexities in the station environment due to occluded sight lines, shadows, obscure angles of view, and many situations in which passengers do not face directly toward cameras. Such a system to recognize specific individuals is unlikely to work unless an extensive network of cameras are specifically placed to view passengers' faces at key locations throughout the system.

A recent comparison of algorithms for pedestrian detection from surveillance feeds reveals that there can be very different outcomes from different methods in terms of accuracy, depending on the context of data collection. The specific study by Kurilkin and Ivanov [34] compares four algorithms for measuring people flow: Aggregate Channel Feature (ACF) Caltech, ACF INRIA, Viola-Jones, and Histogram of Oriented Gradients (HOG). The ACF methods attempt to identify a person's entire body in the frame; the Viola-Jones method focuses on identifying faces; and the HOG method identifies clusters of moving objects. A visualization of the different identified objects is shown in Figure 2.2. The accuracy of the methods varies with the ACF producing pedestrian counts within about 5%, the Viola-Jones method undercounts by about 80% because of challenges related to identifying faces, and the HOG method overcounts by as much as 80% because of multiple groupings identified among large crowds of people.

1. ACF (Caltech)

2. ACF (INRIA)

3. Viola-Jones

4. HOG

Figure 2.2 Detection results from four methods [34]

A number of studies have used image processing to track pedestrian movements in transit stations. Researchers have noted some of the unique challenges of applications in transit stations associated with obscured fields of view, awkward sight angles, and large crowds [30]. Efforts to implement video counting systems have achieved accuracy rates for counting pedestrian flows that exceed 90% in several instances [28, 35]. It is less clear how well such algorithms will work for continuing to track waiting passengers, especially with the goal of identifying which passengers are being left behind and which may be waiting for a different train or moving toward an exit.

Digital Image Processing for Object Detection

The detection of objects in surveillance videos is an invaluable tool for passenger counting and has numerous applications. For example, object detection can be used for passenger counting or tracking, crowding recognition, hazardous object recognition and safety evaluation of autonomous technologies that use object detection to avoid conflicts. Computer vision aims to deuplcate human vision in order to electronically perceive, understand, and store information regarding one or multiple images [36]. There are various techniques to detect objects in images using computers.

Recent methods for detecting objects use feature-based techniques, rather than segmentation of a moving foreground from a static background that was used in the past. Then, the detected features are extracted and subjected to a classification stage, typically using either boosted classifiers or Support Vector Machine (SVM) methods [37, 38]. SVM is one of the most popular methods used in object detection algorithms, especially passenger counting, because it offers a method to estimate a hyperplane that splits feature vectors extracted from pedestrian and negative samples [37], differentiating pedestrians from other unwanted features. Boosting aims to use a sequence of algorithms to convert weak learners to strong learners [39]. The main idea of the boosted classifiers is weighing weak classifiers and combining them to form a strong hypothesis when training the algorithm to attain an accurate detection.

Current methods for object detection take a classifier for an object and evaluate it at several locations and scales in a test image. This has been found to be time-consuming and created numerous computational instabilities at large scales [40]. The most recent methods such as R-CNN use another method to decrease the region in which the classifier runs, the SVM. First, category-independent regions are proposed to generate potential bounding boxes. Second, the classifier runs and extracts fixed-length feature vectors for each of the proposed regions. Finally, the bounding boxes are refined by the elimination of duplicate detections and rescoring the boxes based on other objects on the scene using SVMs [41]. The bounding box is a rectangular box located around the detections in order to represent their detection [42, 43]. Object detection datasets are images with tags used to classify different categories [44, 45]. The technique that is used in this project is bounding boxes prediction.

You Only Look Once (YOLO)

You Only Look Once (YOLO) software uses a different method than the above-mentioned techniques for object detection. It generates a single regression problem, straight from image pixels to bounding box coordinates and class probabilities [46]. YOLO uses a single convolutional network that simultaneously predicts multiple bounding boxes and class probabilities for these boxes [47]. The ability to train YOLO on images has the potential to directly optimize detection performance and increase bounding box probabilities [47]. Another advantage of YOLO is that, unlike other techniques such as SVMs, it sees the entire image globally instead of sections of the image. This feature enables YOLO to implicitly transform contextual information to the code about classes and their appearance and at the same time makes YOLO accurate, with less than half the number of errors compared to Fast R-CNN [47]. YOLO uses COCO which is a large-scale object detection, segmentation, and captioning dataset. The minimum bounding box restricted size is 13x13 tiles [46].

Additionally, YOLO can learn and detect generalizable representations of objects, outperforming other detection methods, including R-CNN. This makes YOLO useful for numerous applications, including real-time object detection, and it is less likely to break down when applied to new domains or fed unexpected inputs [46, 47].

2.2.2 Wireless Device Detection Technologies

An alternative to tracking pedestrians visually is to track the devices that passengers typically carry with them, most notably mobile phones. Kurkcu and Ozbay [48] present a succinct review of existing and emerging technologies for pedestrian counting. A comparison table of technologies is reproduced in Table 2.1.

Table 2.1: Pedestrian Counting Technologies [48]

	Technology	Pedestrian Detection	Cost
Permanent	Wi-Fi & Bluetooth Sensors	○	$
↑	Pressure Sensor	○	$$
	Radar Sensor	○	$-$$
How Long?	Seismic sensor	○	$$
	Video Imaging: Automated	○	$-$$
↓	Infrared Sensor (Active or Passive)	●	$-$$
	Video Imaging: Manual	○	$-$$$
Temporary/Short Term	Manual Observers	●	$$-$$$
○	Indicates what is technologically possible		
●	Indicates a common practice		
$, $$, $$	Indicates relative cost per data point		

There are a number of technologies for detecting electronic devices which can be used for passenger counting. The widespread standard is Bluetooth technology, which facilitates radio communications between smart devices. In order to be detected, a Bluetooth device must be set to discoverable, and this is reportedly between 5%–12% of potential Bluetooth devices [49]. The Bluetooth device detects the unique media access control (MAC) address for each device within range. The Bluetooth detector pings for devices over a period of time repeatedly, and a running list of detected devices is collected with time-stamps for the times that they were observed. By considering devices as a proxy for passengers, this data provides observations of the time that a passenger arrives in a station (based on the first observed timestamp) and the time that they leave a station (based on the last observed timestamp), which is critical for detecting whether that passenger was left behind. An additional benefit of tracking MAC addresses is that it does not change for the same device, so if the same MAC address is later observed at another station, it represents a direct observation of a passenger movement from one location to another. Some previous studies have sought to use Bluetooth data to estimate transit wait times [48] and origin-destination flows [50, 51].

Bluetooth scanning is based on polling, and not on passive listening. This makes Bluetooth detection slow and leaves the chance for a device to avoid detection by ignoring a polling request. Any smartphone can be configured to be visible or not by other Bluetooth devices. Setting this option as “not visible” will make the smartphone undetectable by any other Bluetooth device or sensor. This relates to the major downside of Bluetooth detection, which is that the sampling rate is very low, as stated above. When collecting data to aggregate over long periods of time, this may not be a big problem because the aggregation of a low sampling rate can still yield a large data set. Hence, estimates of how many passengers were left behind due to crowding could not be reliably made for a specific date and time. For the problem of identifying left behind passengers, it would be useful to have a much richer data set. For this purpose, there has been recent development of sensors that use both Bluetooth and Wi-Fi signals to detect devices.

Some products even use cellular signal Wi-Fi detection to make use of a communications channel that allows devices to connect to a wireless local area network. This is a common communication for smart phones, tablets, and laptop computers that passengers often carry with them. Manufacturers claim that by using both Wi-Fi and Bluetooth signals, as many as 95% of smartphone, tablets, hands free devices, and laptops are detected by their MAC address within the detection range. Detection of Wi-Fi enabled devices requires that Wi-Fi is on, and this is more likely to be the case in environments where people are used to using free Wi-Fi services. Nevertheless, more and more devices are left to scan for Wi-Fi signals at all times, so the detection rate is likely to be high, and certain to be higher than Bluetooth alone. A few challenges and complications are related to the use of wireless detection devices to identify passengers.

- **Scanning devices must be installed –** Unlike surveillance systems that have cameras already installed in stations, new devices would have to be acquired. For a long-term solution, these devices would have to be wired into communications channels in order to log records in a database.
- **Scanner range –** The range of detection systems varies greatly from outdoor to indoor settings. It is not clear what range the devices will have in underground stations, especially those that have many concrete columns and walls, which are likely to block signals. Therefore, depending on the architecture more than one device might need to be installed.
- **Electronic devices do not map one-to-one with passengers –** The essence of the technology is that it detects electronic devices that are enabled with communications, typically included in smart phones, tablets, computers, etc. Many commuters carry multiple devices, so it is likely that some passengers will be double counted. Likewise, some commuters do not carry any device at all or may not be detected at all. There is a risk that data from these sources will oversample relatively wealthier socioeconomic groups and under sample others. This raises some potential concerns for equity and sampling rates which will need to be carefully considered as part of a data collection trial.

There are a few manufacturers who produce scanners that detect Bluetooth and Wi-Fi signals:

- **Libelium[1]** – Products from Libelium include a high-powered scanner, called Meshlium, that is designed to collect the maximum number of MAC address signals using a combination of Bluetooth and Wi-Fi signals. Their applications include indoor environments where the scanner is used to count and track pedestrian movements.
- **BlueMark Innovations[2]** – BlueMark produces a modular platform to detect, track and locate smartphones based on Wi-Fi and Bluetooth (Classic, Low Energy,

---

[1] http://www.libelium.com/products/meshlium/smartphone-detection/
[2] https://bluemark.io/products/

iBeacon, Eddystone) technology. The product provides a data dashboard for viewing metrics such as counting unique visitors and detecting specific users. The manufacturer claims to have a 25-meter range in an indoor location with pillars. The platform also offers ports for 3G/4G detection.

- **SMATS[3]** – A highly portable product called TrafficTab and mountable product called the TrafficBox provide Bluetooth and Wi-Fi detection capabilities in a portable case that can be easily mounted for temporary data collection. A more permanent product called TrafficXHub connects with a constant power supply for an extended scanning range and long term data collection.

- **TrafficCast[4]** – A high-quality Bluetooth detector called BlueTOAD Spectra is designed to detect both discoverable and non-discoverable Bluetooth signals. The manufacturer claims that sampling rates are competitive with or even exceed Wi-Fi detection. The product was developed as a roadside detector, so it can be wired into a traffic control cabinet. For experimental purposes the device can be run on batteries, but it requires cellular coverage in order to communicate the data.

After comparing the strengths and weaknesses of the various products available on the market, the project team recommended the purchase of four (4) SMATS TrafficBox units for the purposes of this project. This recommendation was based on consideration of the types of signals detected, the strength of the antennas, the portability of the device, the rugged housing for outdoor and dusty environments, and the price.

2.2.3 Other Pedestrian Counters

Other forms of pedestrian counters use either radar or infrared technology to count the passage of pedestrians past a location. The simplest of such counters simply count the number of times that an *infrared* beam across a corridor is broken as an undirected count of pedestrians. This is a technology that has long been employed to count customers entering a store, for example. A slightly more sophisticated implementation with two beams closely spaced in series allows for inference of pedestrian direction, because the order in which the beams are disrupted indicates the direction of movement. Infrared technologies are limited by the fact that observations are based only on disruption of a signal, so the technology works best in low traffic environments in which people pass individually through a corridor to be counted. In crowds, large groups of people may disrupt the signal continuously leading to large undercounting errors.

Radar technologies have been developed to track individual pedestrians (as well as cyclists and motor vehicles) simultaneously within a field of view. When mounted high above the traffic stream, radar systems are capable of tracking the trajectories of individual pedestrians moving toward and away from the radar. Some devices are also able to record lateral position.

---

[3] http://www.smatstraffic.com/products/
[4] http://www.trafficcast.com/spectra/index.html

Although there are some benefits to using pedestrian counters in transit stations, the applicability for measuring crowding and left behinds are limited by the fact that specific individuals cannot typically be identified and tracked throughout their path in a station. There may be some cases where the installation of passenger counters could provide measures of station exits and transfer movements that could be used to supplement inferred destinations and transfers in the ODX model. For tracking left-behinds, these data sources provide only indirect observations that would rely on extensive data fusion and inference to be of value. For this reason, this study will focus on the potential benefits of video and device detection technologies.

2.3 Identification of Locations and Times for Evaluation

The purpose of this project is to test different methods for detecting and measuring passengers being left behind when trains are too full to board. In order to make a detailed data collection plan, it was first necessary to identify locations and times of day that would be suitable for observation.

First, the project team focused attention on the Orange and Blue lines, because the MBTA operates service on a single line without branches; see Figure 2.3. The challenge with branching lines (such as the Red line or Green line) is that it is not always clear whether a person who does not board a train is left behind due to crowding or is waiting to travel to a destination on a different branch. Furthermore, the station layout can contribute to ambiguous passenger behaviors such as the case at JFK/UMass (Red Line) in which there are two different platforms that passengers can choose for inbound trains. These complexities make it difficult to separate the problem of left behind passengers from other types of passenger movements.

Figure 2.3: MBTA Rapid Transit Network Map (Source: MBTA)

Second, the team identified stations and times of day that most commonly experience crowding conditions that cause passengers to be left behind. Specifically, passengers are most likely to be left behind on a platform when many passengers are waiting to board trains that are already fully loaded when they enter the station. Using Rail Flow data (which is an aggregation of ODX records) and train schedules, the average occupancy of trains at each station and time of day can be compared in order to identify the most likely points where passengers would have difficulty boarding. By conducting an analysis of passenger flows boarding and alighting trains at each station, an estimate of passenger loads on trains and demand for boarding was used to identify candidate stations for conducting detailed data collection in the field.

2.3.1 Crowding Analysis

A crowding analysis is a necessary step in the methodology applied to identify the times and stations where crowding is observed and left-behinds have a higher probability of occurring. The data used in this part of the analysis have been extracted from the Rail Flow database in the MBTA Research and Analytics Platform. Each 15-minute count of boardings and alightings represents an aggregated average of estimated passenger boarding and alighting counts across all days in the Winter 2017 quarter.

Cumulative counts of the numbers of passengers boarding and passengers alighting have been created with respect to stations along the direction of train travel. For a 15-minute time period, $B(n,t)$ is the count of all of the passengers that are assumed to board trains in the direction of interest at stations preceding and including station n by time t. Similarly, the cumulative number of passengers alighting, $A(n,t)$, is the count of all passengers that are assumed to have exited trains traveling in the direction of interest at stations preceding and including station n by time t.

It should always be true that $A(n,t) \leq B(n,t)$, because passengers can only alight a train after boarding it. The difference between the cumulative boardings, $B(n,t)$, and alightings, $A(n,t)$, provides an estimation of the passenger flow, $Q(n,t)$, between adjacent stations during each 15-minute time period.

$$Q(n,t) = B(n,t) - A(n,t) \tag{1}$$

This calculation is approximate, because cumulative counts are calculated for a single 15-minute time period, and real trains take more than 15 minutes to traverse the length of a line. Moreover, to calculate the crowding on trains, the passenger flow per time period should be converted to a passenger occupancy, $O(n,t)$ (passengers/train), which is calculated by multiplying the passenger flow by the scheduled headway of trains, $h(t)$ (minutes), at time t; see Table 2.2.

$$O(n,t) = Q(n)\frac{h(t)}{15} \tag{2}$$

In this equation, the headway is divided by 15 minutes to account for the fact that the passenger flow is per 15-minute time period. This measure is an approximation of the number of passengers onboard each train and is based on the assumption that real headways are uniform and that passengers are always able to board the next arriving train. In reality, variations and headways may lead to increased crowding after longer headways, increasing the likelihood that some passengers will be left behind.

The 2017 MBTA Service Delivery Policy (SDP) [1] provides guidelines for reliability and vehicle loads. In the 2010 MBTA SDP, the maximum vehicle load was explicitly defined as 245% of seating capacity in the peak hours and 143% of seating capacity in other hours. The 2017 SDP notes that accurately monitoring the passenger occupancy of heavy rail transit is not yet feasible on the MBTA system. Nevertheless, the guidelines from Table B2 in the 2017 SDP are used to identify general crowding levels, recognizing that each Blue Line and Orange Line train are six cars long as shown in Table 2.3.

Table 2.2: Scheduled Headways for Weekday Service

Time Period (approximate)	Orange Line	Blue Line
First Train	5:16am	5:16am
AM Peak (6:30am – 9:00am)	6 min	5 min
Midday (9:00am – 3:30pm)	9 min	9 min
PM Peak (3:30pm – 6:30pm)	6 min	5 min
Evening (6:30pm – 8:00pm)	10 min	9 min
Late Night (8:00pm – Close)	10 min	9 min
Last Train	12:30am	1:00am

Table 2.3: Vehicle Load Limits for MBTA Heavy Rail Trains [1]

Time Period	Orange Line		Blue Line	
	Seats	Passengers	Seats	Passengers
Early AM/AM Peak (Start – 9:00am)	348	846	210	516
Midday Base (9:00am – 1:30pm)	348	498	210	300
Midday School/PM Peak (1:30pm – 6:30pm)	348	846	210	516
Evening/Weekends (6:30pm – Close)	348	498	210	300

Tracking the average number of passengers onboard trains provides one indicator for the likelihood of passengers being left behind, because full trains leave little room for additional passengers to board. During the most crowded times of the day, it is also useful to look at the numbers of passengers boarding and alighting trains at each station. Passengers are most likely to be left behind at stations where trains arrive with high occupancy, few passengers alight, and many passengers wait to board.

2.3.2 Station Geometry and Camera Views

In addition to identifying stations with the greatest likelihood of passengers being left behind by crowded trains, the stations selected for detailed analysis should also have characteristics that are amenable to successful testing of video surveillance counting methods. There are a variety of station layouts and architectures that contribute complicating factors to the analysis of left-behind passengers, and the goal of this study is to identify the potential for different detection methods under the best possible conditions.

Ideal conditions for the proposed analysis are:

1. **Dedicated Platform for Line and Direction of Interest –** In this case, all passengers on a platform are waiting for the same train, so any passenger that does not board can be counted as being left behind. In the case of an island platform, observed passengers may be waiting for trains arriving on either track.
2. **High Quality Camera Views –** Surveillance cameras vary in age, quality, and placement throughout the MBTA system. Newer cameras have higher definition video feeds. The quality of the view is also affected by lighting conditions, especially

at above-ground stations where sunlight and shadows can affect the clarity of the images.

3. **Platform Coverage of Camera Views –** The surveillance systems are designed to provide views of the entire platform area for security purposes. In some stations, the locations of columns obfuscate the views, requiring more cameras to provide this coverage.

Surveillance camera views were considered from five stations on the Orange Line (Back Bay, Chinatown, North Station, Sullivan Square, and Wellington) and two stations on the Blue Line (State and Maverick). Ultimately, Sullivan Square and North Station were selected for more detailed analysis because these stations satisfy all three of the requirements listed above.

Sullivan Square is an above ground station that is located under I-93 and parallel to the tracks serving the Haverhill, Newburyport, and Rockport commuter rail lines. The station features heavy concrete construction and the platforms are open to the outdoors. The station serves only the Orange Line. The platform shown in Figure 2.4 serves southbound (inbound) trains on the right track. Although northbound trains use the left track and often open doors onto the same platform, the primary platform for northbound passengers is located in the left side of the view. Sullivan Square is an important transfer station with many passengers transferring to the Orange Line from bus routes serving surrounding neighborhoods.

Figure 2.4: Sullivan Square, Southbound Orange Line Platform

North Station is an underground station that serves both the Orange Line and the Green Line. Northbound Orange Line trains have a dedicated side platform, visible on the left side of Figure 2.5. The open atrium design adds some complexity to the analysis in that passengers traveling on other lines or directions are in the same space, and their devices may be detected by equipment placed to detect northbound Orange Line passengers. The surveillance views provide good dedicated coverage of the northbound platform, which is the platform of interest for evening peak travelers.

Figure 2.5: North Station, Northbound Orange Line Platform

2.4 Collection of Direct Observations

Detailed data collection was conducted for two stations on two dates: Sullivan Square for southbound Orange Line trains in the AM peak (6:30 – 9:30am) and North Station for northbound Orange Line trains in the PM peak (3:30 – 6:30pm), with the selected dates midweek days during non-holiday weeks (Wednesday, November 15, 2017, and Wednesday, January 31, 2018). In both cases, manual observations on the platform were collected to establish a ground truth against which to compare alternative methods for measuring and estimating passengers being left behind by crowded trains. In each case, three observers worked simultaneously on the station platform to record observations.

2.4.1 Train Door Opening and Closing Times

Although train-tracking records (TTR) report the times that each train enters the track circuit associated with a station, there is no automated record of the precise times that doors open and close. Since passengers can only board and alight trains while the doors are open, recording these times manually is important for identifying when passengers board trains, when they are left behind, the precise dwell time in the station, and the precise headway between trains. Each of the three observers recorded the times of doors opening and closing. The average of these observations is considered the true value.

2.4.2 Number of Passengers Left Behind

Each observer counted the number of passengers left behind on the station platforms after the train doors closed. In order to avoid double-counting, each observer was responsible for observing passengers in a two-car segment of the train (front, middle, and back). Some judgement was necessary in determining which passengers to count, because some passengers linger on the platform after alighting the train and some choose to wait for a later train even when there is clearly space available to board. The goal of the left-behind passenger count is to measure the number of left-behind passengers within ±2 passengers of the true number.

2.4.3 Number of Passengers Waiting on Platform

In addition to counting the number of passengers left behind by crowded trains, it is important to get an accurate count of the number of passengers waiting to board each arriving train. Given the large number of commuters using the heavy rail system during commuting hours, it is not possible to accurately count this total number of passengers in person.

Surveillance video feeds of escalators, stairs, and elevators used to access the platform of interest were used to manually count the number of passengers entering and exiting the platform offline. Specifically, an open-source software tool was used to track passenger movements by logging keystrokes to the video timestamp during playback. A student data logger conducted the counts by watching the surveillance video playback of each entry and exit point from the platform and logging the entry and exit of each individual passenger. The resulting data log records the time (to the nearest second) that each passenger entered and exited the platform. Since the platforms of interest serve only one train line in one direction, all entering passengers are assumed to wait to board the next train, and all exiting passengers are assumed to have alighted the previous train. Combining these counts with the direct observations of the number of passengers left behind each time the doors close provides an accurate estimate of the number of passengers that were successfully able to board each train.

2.5 Automated Video Counting

In order to evaluate the accuracy and feasibility of counting passengers on station platforms using surveillance video feeds, an automated pedestrian identification algorithm was calibrated and used to compare against the direct passenger counts in stations. The method for conducting automated video counts involved three main steps: 1) identification of camera feeds for analysis, 2) calibration of algorithm parameters, and 3) smoothing of the raw data feed into a time series of estimated passenger counts. For this analysis, the YOLO algorithm (as described in Section 2.2.1) was used. This is an open source software tool that is freely available.

2.5.1 Identification of Surveillance Camera Feeds for Analysis

The automated passenger counting algorithm uses pattern recognition to identify passengers in each frame of surveillance video. In order to maximize the accuracy of the automated video counts, camera views that provide a clear and unobstructed view of the platform were selected. The selected video views for automated passenger counts on platforms should have the following characteristics:

- Clear view of platform edge and train doors
- High enough viewing angle to see individual people
- Minimal occlusion of view from columns or other objects
- Lighting that reveals features of the station and objects
- High resolution and clear focus

The suitable video feeds for Sullivan Square and North Station were selected from samples of each surveillance feed; see

Figure 2.66 and Figure 2.77, respectively.

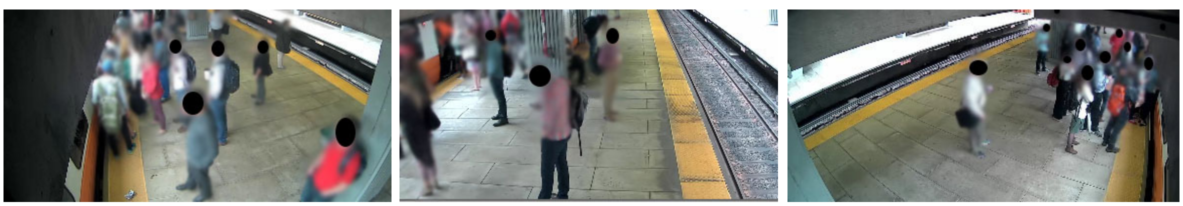


Figure 2.6: Selected surveillance camera views of Sullivan Square southbound platform

Figure 2.7: Selected surveillance camera views of North Station northbound platform

2.5.2 Calibration of Parameters

The YOLO algorithm uses pattern recognition to identify objects in an image. A threshold for certainty can be calibrated to adjust the number of identified objects in a specific frame. If the threshold is set too high, the algorithm will fail to recognize some objects that do not adequately match the training dataset. If the threshold is set too low, the algorithm will falsely identify objects that are not really present. In order to identify the optimal threshold, 14 camera views were analyzed. Each frame was run separately for threshold values ranging from 6% to 25% to determine the optimal threshold value in relation to a manual count of passengers visible in the frame. The optimal threshold across all camera views is 7%, which minimizes the mean squared error between YOLO and manual counts as shown in Table 2.4. Figure 2.8 shows the identified objects at each threshold level for the same frame from a camera in North Station.

Table 2.4: Count Errors for YOLO Thresholds

Threshold	Mean Error	Mean Squared Error
25%	6.82	62.0
20%	5.82	43.7
15%	4.59	28.5
10%	3.06	14.1
9%	2.23	8.0
8%	1.35	4.6
7%	0.12	1.2
6%	-1.53	6.8

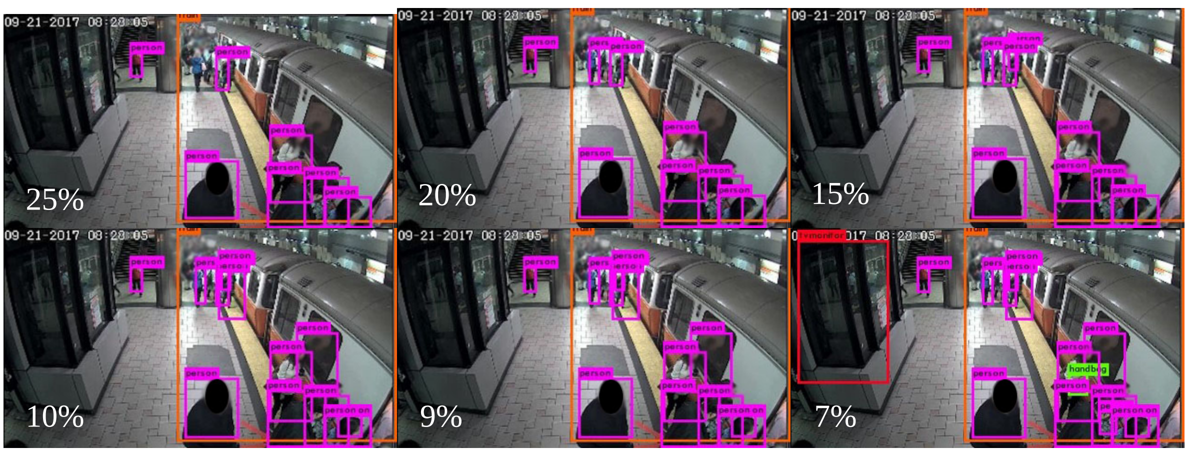
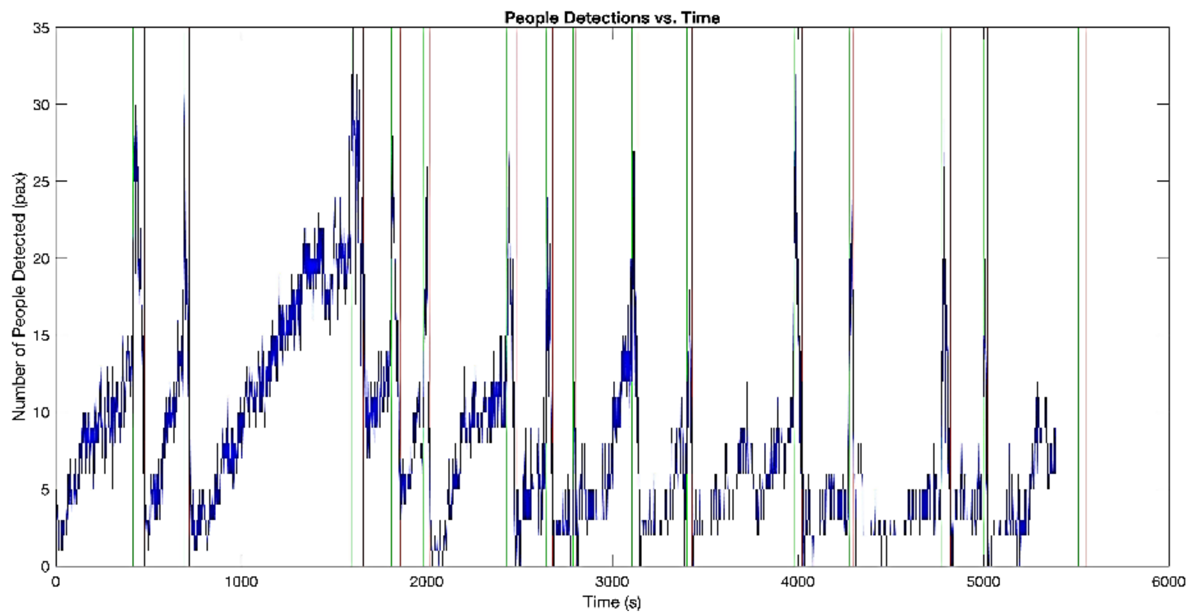


Figure 2.8: Passenger counts at various thresholds using YOLO

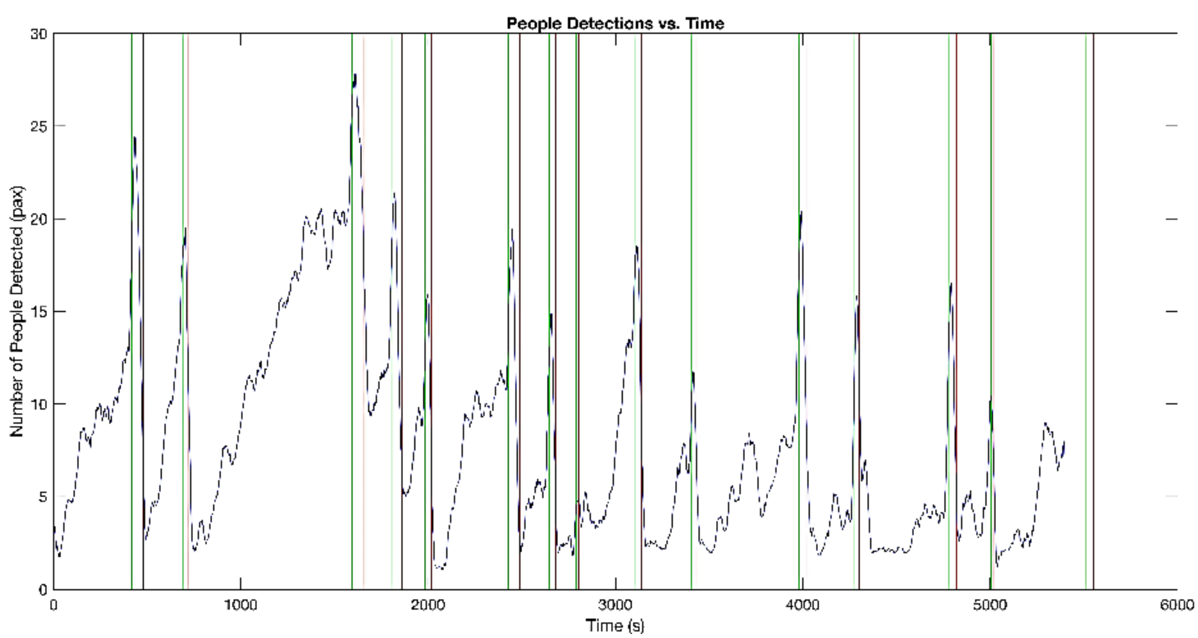
2.5.3 Processing of Raw Video Counts

The YOLO algorithm processes each frame of surveillance video as an isolated image.The raw video feeds for the time period of interest are available in .ASF format at 30 frames per second (fps). Since passengers do not move very quickly around platform areas, the raw video is downsampled to 1 fps in order to reduce file sizes and computation time, but each of these frames maintains the original image resolution. The reduced video file is also converted to .MP4 format using a Matlab script in order to be compatible with the YOLO software. YOLO processes each frame in less than one second, so this method is fast enough that it could be implemented in real-time.

The output from YOLO is a text file that lists for each frame the objects detected (e.g., person, handbag, train, etc.) and the bounding box for the object within the image. A time series count of passengers on the platform is simply the sum of the counts of the number of "person" objects identified in the corresponding frames from each sampled video feed. Figure 2.9a shows the raw passenger counts on the platform at North Station for the time period from 5:30 – 6:00pm on November 15, 2017. Although there are noisy fluctuations, there is a clear pattern of increasing passenger counts until door opening times (shown in green). A surge of passenger counts while doors are open (between green and red) represents the passengers alighting the train and exiting the platform. Passenger counts typically drop off dramatically following the door closing time (shown in red), except in cases that passengers are left behind. For example, the third train arrives after a long headway and shows roughly nine passengers left behind.



a) Raw (unsmoothed) time series



b) Smoothed time series

Figure 2.9: Raw and smoothed passenger counts from video at North Station (5:00 – 6:30pm), November 15, 2017

To facilitate analysis of the automatic passenger counts from the surveillance videos, it is useful to work with a smoothed time series of passenger counts. Using a smoothing window of ±10 seconds, the smoothed series appears as in Figure 2.9b. This smoothed time series is more suitable for a local search to identify the minimum passenger count following each door closing time. This represents the count of left-behind passengers identified through automated video counts.

2.6 Wireless Device Detection

Wireless device detection tools detect the unique MAC addresses of the wireless devices that people carry rather than counting or tracking the individual passengers themselves. The devices regularly ping Bluetooth, Wi-Fi, or cellular network signals (depending on the specific device used) to detect the presence of wireless devices such as mobile phones, tablets, and laptop computers within range of the antenna. A common application in the transportation domain is to use wireless device detectors to measure travel times between points in a network by identifying and then reidentifying devices at different locations and calculating the travel time or speed of travel between observations.

For the application of counting passengers and measuring waiting times in a transit station, the goal is to analyze the records of pinged MAC addresses to identify when devices are first and last identified within a station environment. The duration from the first to last observation is a measure of the time that device was present in the station, and the timing of the last observation relative to train departures can be used to assign each device to a departing train. If the devices are assumed to be a proxy for passengers, then the times that devices spend waiting for each departing train may be interpreted as a proxy measure for the times that passengers wait for departing trains. The feasibility of this method depends on the detection range of the tool and the reliability of detecting individual devices.

For the purposes of this project, four SMATS TrafficBox devices were procured to test the feasibility of using Bluetooth and Wi-Fi device detection to measure passenger waiting times and the occurrence of passengers being left behind by crowded trains. A comparison with other products is presented in Section 2.2.2.

Each TrafficBox contains five key components as illustrated in Figure 2.100:

1. **Control Board –** The control board connects all of the other components and records observations to an internal memory card for later download. The memory card has a 16GB capacity, suitable for 2 million MAC records. The control board can be connected with an Ethernet connection for real-time transmission of data.
2. **Battery Pack –** A 5V, 20Ah battery pack can power the unit for roughly 20-24 hours. The unit can also be plugged in for extended operation.

3. **GPS Antenna –** A GPS antenna is used to synchronize the system clock prior to data collection. Once the clock is synchronized, the GPS connection is no longer necessary, so this does not pose a problem for use in underground stations.
4. **Bluetooth Antenna –** An omnidirectional Bluetooth antenna detects Bluetooth MAC addresses in three modes: classic discovery mode, low-energy discovery mode, and classic paired mode.
5. **Wi-Fi Antenna –** An omnidirectional Wi-Fi antenna detects Wi-Fi MAC addresses.

Each TrafficBox can be mounted to a pole or conduit with mounting brackets, which enables them to be placed in many potential locations within each transit station. Each box can also be padlocked to discourage tampering.

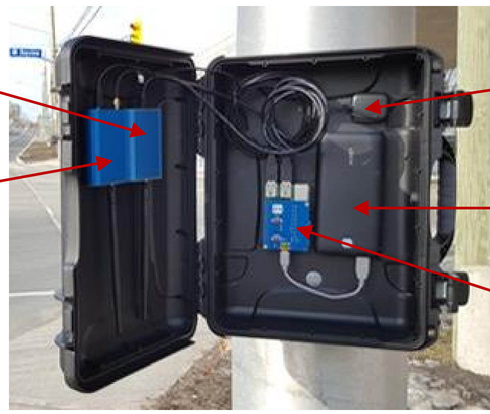


Figure 2.10: Components of the SMATS TrafficBox wireless device detector

2.6.1 Detection Range

To determine placement of the TrafficBox devices within Sullivan Square and North Station, it was first necessary to determine the detection range. The manufacturer advertises a detection range of up to 500m in an unobstructed field of view. For this study, it was necessary to measure the detection range of the devices from various locations within the stations so that the units could be placed at locations that would provide the most complete coverage for the platform of interest. The detection range was measured using the unit's range test mode to detect a known target device (an Apple iPhone 6) by MAC address.

Sullivan Square Detection Range

The detection range for the TrafficBox unit was measured from three locations within the Sullivan Square station: the south end of the platform, the middle of the platform, and the north end of the platform. The target device was detected only on the same platform within half a platform length of the unit. The target device was not detected on the other station platform or in the station lobby. The limited detection range is likely attributable to the heavy concrete construction of the station which blocks signals that are not within the line of sight. Figure 2.11 shows the selected placement of the devices at Sullivan Square. The devices were mounted near the ceiling to scan over passengers' heads.

North Station Detection Range

The detection range for the TrafficBox unit was measured from four locations within North Station: the south end of the platform, 1/3 of the platform length from the south end, 1/3 of the platform length from the north end, and at the north end of the platform. At each location, the unit was able to detect the target device anywhere within the station, including the opposite platform, the upper Green Line platform, and the ticketing mezzanine. The target device was not detected at street level, which is good because this would add an additional challenge to filter out people walking by on the street above the station. It is likely that the station's open design and many metal surfaces increase reflections of signals that allow devices to be detected from anywhere in the station. Figure 2.12 shows the selected placement of the devices at North Station. The devices were mounted on overhead light fixtures directly over the platform.

Figure 2.11: Placement of wireless device detectors at Sullivan Square

Figure 2.12: Placement of wireless device detectors at North Station

2.6.2 Processing Raw Data

The raw data recorded by each TrafficBox unit are three data fields: timestamp, MAC address, and detection type (Bluetooth or Wi-Fi). The unit records observed MAC addresses as frequently as every second, although a wireless device may not transmit a detectable signal for several seconds or even minutes at a time. Each of the units recorded hundreds of thousands of observations during the day of data collection. These observations must be processed and filtered to identify the useful observations that are most likely associated with travelers on the line and direction of interest. The data is filtered in the following steps:

1. Observations are extracted for the peak period. From the raw data log, observations are extracted for the 3-hour time period corresponding to the manual counts and the surveillance video observations (6:30 – 9:30am at Sullivan Square, 3:30 – 6:30pm at North Station).
2. Records are sorted by MAC address and then by timestamp so that the time of the first and last observation are identified. The difference between these two times is the duration of observation.
3. Records are filtered to remove transient observations (less than 5 seconds duration) and stationary devices (greater than 960 seconds duration). The first category removes devices that are not observed for long enough to plausibly be a waiting passenger; e.g., passing traffic. The second category removes devices that linger longer than any passenger; e.g., Wi-Fi routers or devices belonging to transit employees.
4. Records with the last observations from 120 seconds before doors closing to the time of doors closing are associated with a particular train departure. Since the average device is pinged about once per minute (and sometimes longer) this time range associates records with the most likely train departure.

From this process, a set of MAC addresses serving as a proxy measure for passengers is identified for each departing train. The timestamp of first observation provides an indication of when the device arrived to the station platform, and how long it waited. If a device is first observed before the prior train's doors close, it appears to have been left behind.

2.7 Modeling Left-Behind Passengers

Direct observations from automated video counts and filtered wireless device detection provide quantitative measures of the number of passengers (or devices) observed on a station platform after a train's doors close. However, the automated measurement processes are known to be associated with noise and are likely to undercount true values. If the automated measurements from video and wireless device detection provide an indication of left-behinds, it would be appropriate to develop a model to estimate relevant metrics based on the observed data.

The focus of this study is to measure passengers that are left behind due to crowded trains. A model that predicts the probability that a passenger waiting to board a train is left behind on the platform and the total number of left-behind passengers would provide a useful measure for tracking system performance. A logistic regression modeling approach for making these estimates is presented in Section 2.7.1.

A second measure of interest is the service reliability, as defined in the 2017 MBTA Service Delivery Policy: the percentage of passengers waiting less than a published headway to board a train. This measure is currently estimated based on reported headways, assuming that each passenger is able to board the next arriving train. A method to account for the additional waiting time experienced by left-behind passengers is presented in Section 2.7.2.

2.7.1 Model for Likelihood of Passenger Being Left Behind

Each passenger waiting on a platform to board a train is either able to board the train or is left behind. These two discrete outcomes allow the system to be formulated as a binary logistic model in which the probability that a passenger is left behind is estimated as a function of observed explanatory variables. Mathematically, this is the same as a logit model, commonly used to model passenger mode choice in the transportation domain. In the context of left-behind passengers, a logistic regression uses maximum likelihood estimation to fit parameters of a model to predict the probability that a passenger is left behind based on some combination of the observed factors.

From the manual observations, the number of passengers waiting on the platform for each departing train is known, as well as the number of these passengers that are left behind. For estimation of the logistic regression, each passenger is represented as a separate observation, and all passengers waiting for the same departing train are associated with the same set of explanatory variables. Over the course of a 3-hour rush period, there are typically about 30 trains serving Sullivan Square and North Station, serving 1,500 to 3,000 passengers per period, and leaving behind well over 100 passengers. Logistic regression models are generally expected to give stable estimates when the data set for fitting includes at least 10 observations for each outcome. Structuring the model in terms of individual passengers is consistent with these guidelines.

The logistic function defines the probability that a passenger is left behind by

$$P(\boldsymbol{x}) = \frac{1}{1+e^{-(\beta_0+\boldsymbol{\beta x})}} \tag{3}$$

where $\boldsymbol{x}$ is a vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of estimated coefficients for the explanatory variables, and β_0 is an estimated constant. The estimation of the model may be understood as identifying the values of β_0 and $\boldsymbol{\beta}$ that best fit

$$y = \begin{cases} 1 & \beta_0 + \boldsymbol{\beta x} + \varepsilon > 0 \\ 0 & \text{else} \end{cases} \tag{4}$$

where $y = 1$ corresponds to a passenger being left behind and $y = 0$ corresponds to a passenger successfully boarding. The underlying assumption in this formulation is that the likelihood of being left behind can be expressed in terms of a linear combination of explanatory variables and a random error term, ε, which is logistically distributed.

The explanatory variables that are considered in this study are as follows:

1. Train-tracking records (TTR)
    - Dwell Time
    - Headway
2. Video count of passengers on platform following doors closing

3. Wireless device detection estimate of devices on platform following doors closing

These explanatory variables can all be monitored automatically, without manual observations. In the case of dwell time, an extra modeling step is needed to convert TTR reported times that trains enter and exit track circuits. This is done by fitting a simple regression model to estimate dwell time based on the duration of time that the train occupies the station track circuit.

Formulations of the logistic regression with various combinations of the explanatory variables are compared to identify which variables have statistically significant explanatory power and which model provides the best fit to the data. The model estimation process for train-tracking and video data is conducted with observations from November 2017, and these models are validated with observations from January 2018. Observations from wireless device detection are only available from January 2018, so the models are estimated to identify the potential benefit of that data source.

2.7.2 Distribution of Experienced Waiting Times

A second measure of interest is the distribution of experienced waiting times that passengers experience for a specific line and direction of travel. From the direct manual counts, a cumulative count of passengers arriving onto the platform and of passengers boarding trains provides a time series count of the number of passengers on the platform. If passengers are assumed to board trains in the same order that they enter the platform, the system follows a first-in-first-out (FIFO) queue discipline. Although it certainly is not true that passengers follow FIFO order in all cases, this assumption allows the cumulative count curves to be converted to estimated waiting times for each individual passenger. The FIFO assumption yields the minimum possible experienced waiting times that are consistent with the numbers of passengers left behind by each train. This distribution implies a percentage of passengers that wait less than the published headway for a train departure.

The models presented in Section 2.7.1 provide the estimated probability that a passenger is left behind each time train doors close. If a constant arrival rate is assumed over the course of the rush period, the door closing times and probability of passengers being left behind can be used to estimate the cumulative boardings onto trains over time. Under the same FIFO assumptions described above, the distribution of experienced waiting times can be estimated based on train-tracking and video counts.

Finally, data from wireless device detection that is filtered as described in Section 2.6.2 provides observations of durations of time that devices are present in the station area. No further modeling is necessary to characterize the distribution of these times, if they are considered to be representative proxies of passenger presence.

3 Results

3.1 Estimated Crowding by Location and Time

The following subsections show inferred vehicle crowding, boarding, and alighting counts for the Blue and Orange lines in order to identify the specific locations for more detailed data collection and investigation. The results for the Orange Line are presented here. The results for the Blue Line are presented in Appendix A.

The average number of passengers boarding and alighting at each station during each 15-minute period during Winter 2017 are used to calculate average passenger loads as described in equations (1) and (2). The resulting passenger occupancy estimates, $O(n,t)$, can be plotted across the stations for an average weekday to show where trains are consistently the most crowded. The red areas in Figure 3.1 show the locations and times that Orange Line trains are consistently the most crowded. Not surprisingly, these are inbound trains during the AM peak and outbound trains during the PM peak.

The most severe crowding on the Orange Line appears to be from 8:15 – 8:30am for southbound trains and 5:15 – 5:30pm for northbound trains. The values of $B(n,t)$, $A(n,t)$, and $O(n,t)$ are shown in Figure 3.2 for southbound trains in the morning. Inbound trains experience large boarding loads at Oak Grove, Malden Center, Wellington, and Sullivan Square, before entering the center of Boston. Sullivan Square appears to be particularly prone to crowding conditions that lead to left-behind passengers in that trains arrive to the station with high passenger loads, there are very few passengers alighting (so additional space is not opening up), and there are many passengers wishing to board. Figure 3.3 shows similar data for northbound trains in the evening. Outbound trains fill with passengers at Back Bay, Downtown Crossing, and State Street on their way north. North Station is the last stop within central Boston and there is a relatively high volume of passengers seeking to board fully loaded trains.

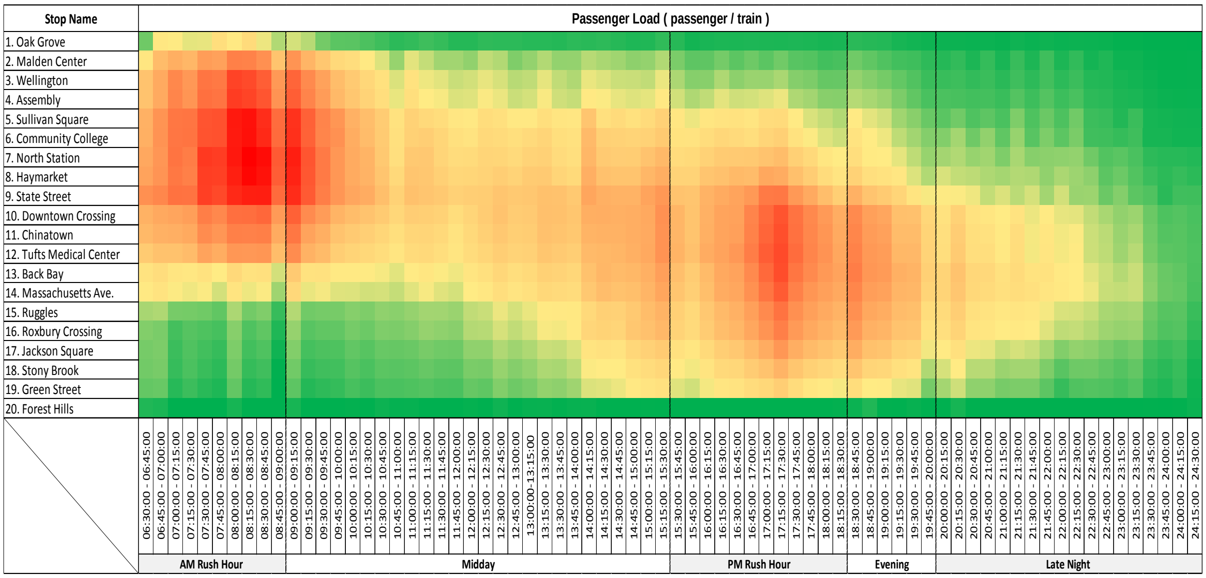
3.2 Data Collection

Detailed data collection was conducted at Sullivan Square and at North Station for the Orange Line on two weekdays. Table 3.1 summarizes the data collection schedule.

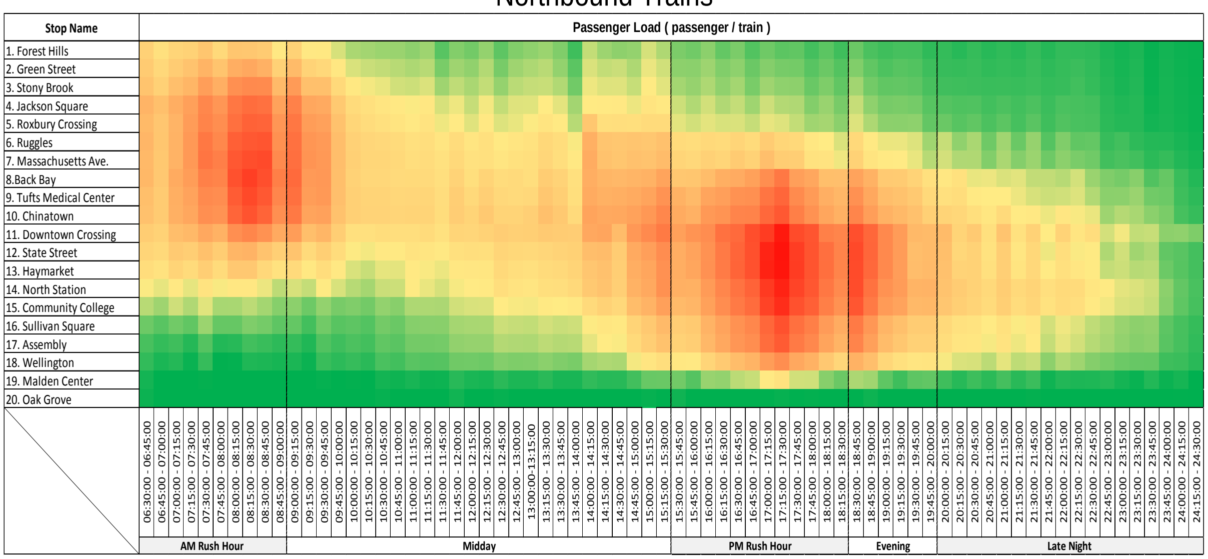
Table 3.1: Data Collection Schedule for Orange Line Trains

	Wednesday, November 15, 2017	Wednesday, January 31, 2018
Locations and Times	Sullivan Square (south): 6:30 – 9:30am North Station (north): 3:30 – 6:30pm	Sullivan Square (south): 6:30 – 9:30am North Station (north): 3:30 – 6:30pm
Manual Counts	●	●
Surveillance Video	●	●
Wireless Device Detection		●

Southbound Trains



Northbound Trains



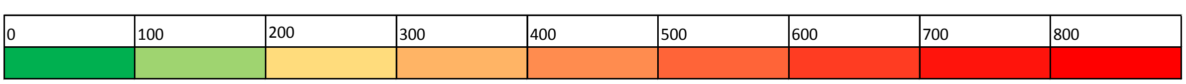
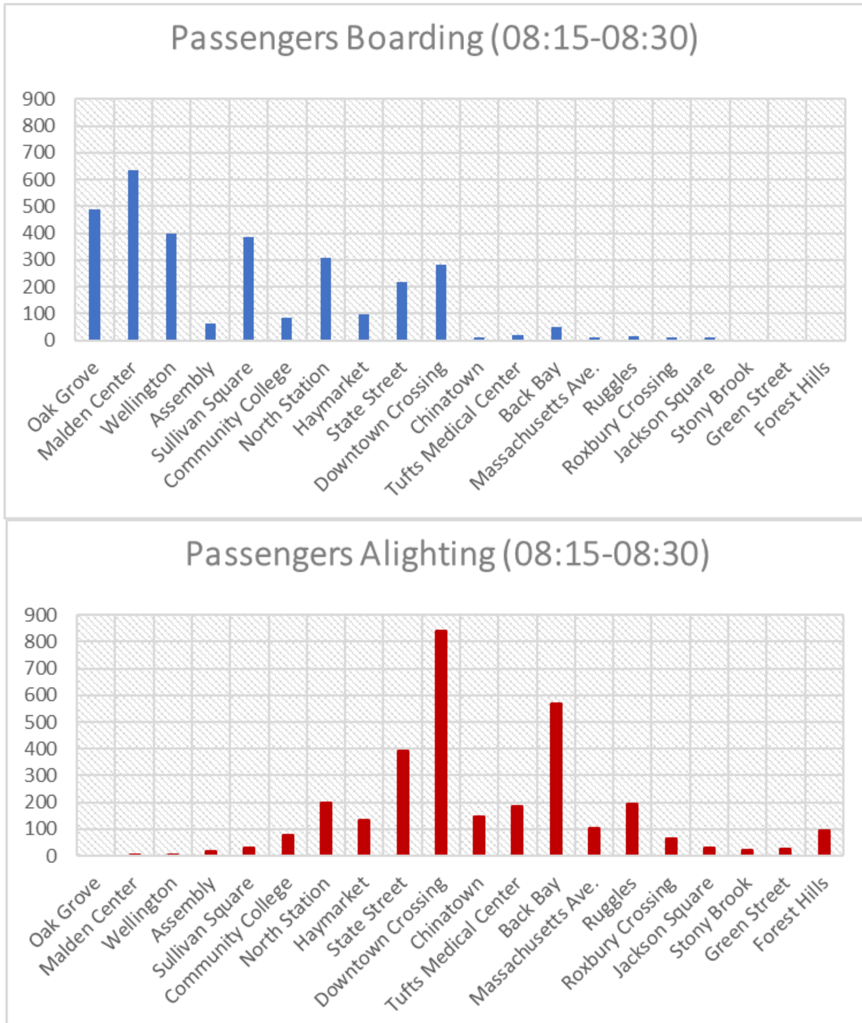


Figure 3.1: Inferred Passenger Occupancy for Orange Line Trains, Winter 2017 (Source: ODX Data from MBTA Research Database)



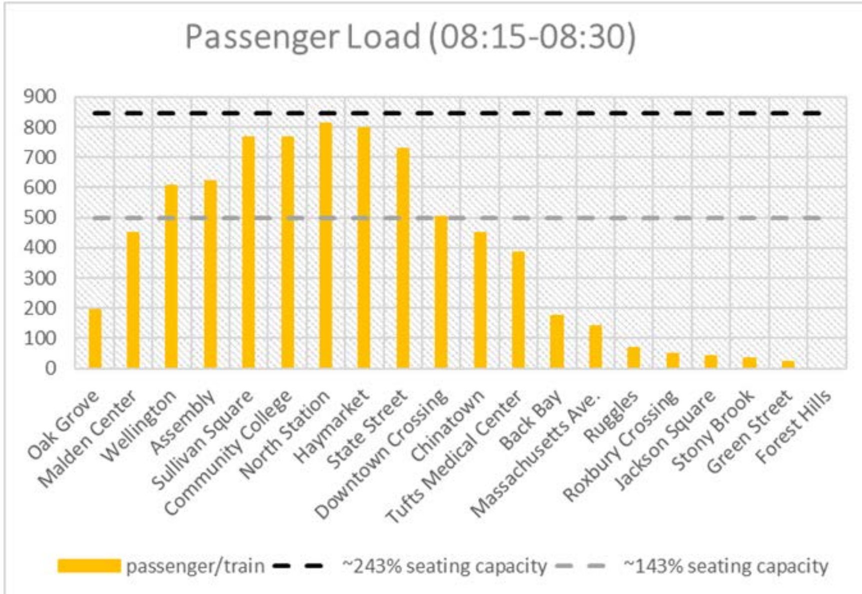
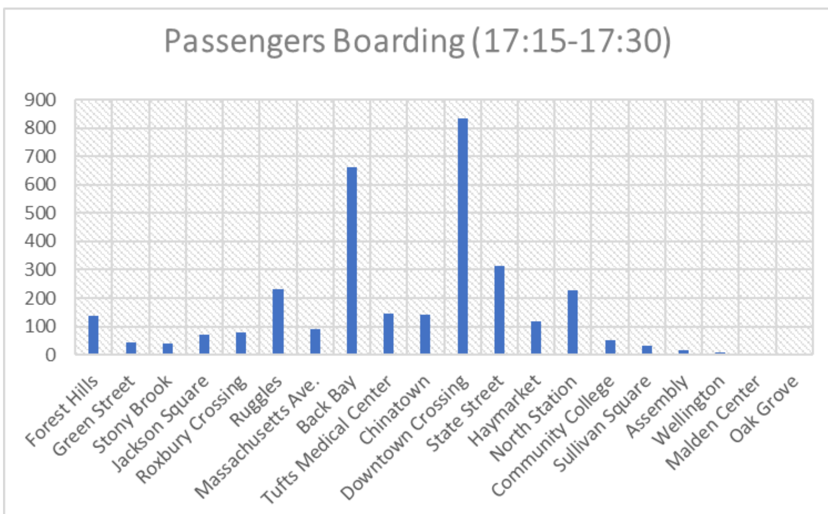
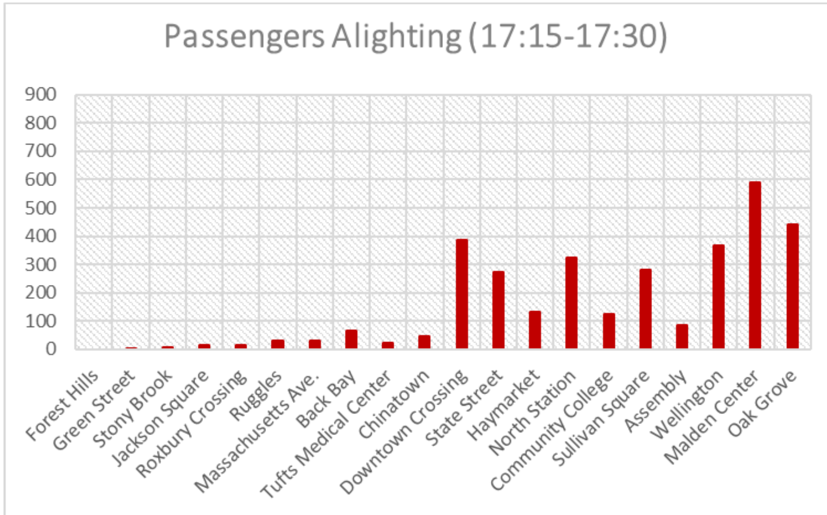


Figure 3.2: Passenger movements for southbound Orange Line, 8:15 – 8:30am (Source: ODX Data from MBTA Research Database)





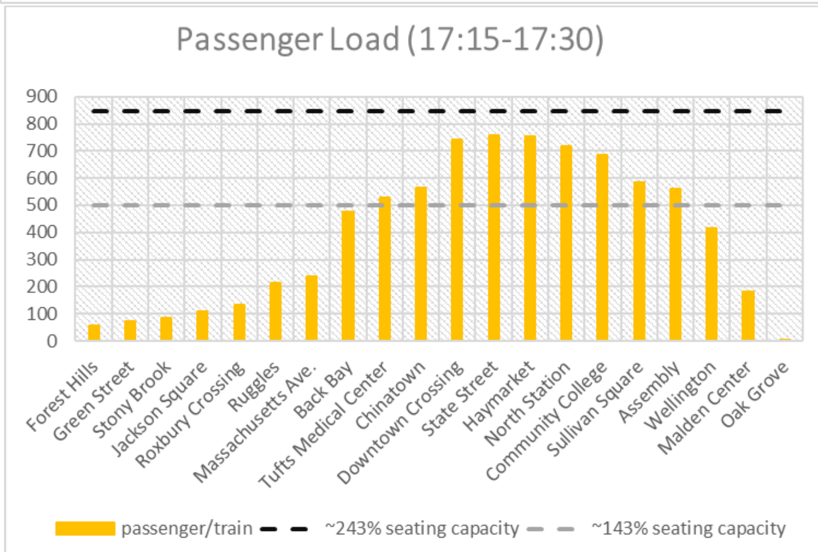


Figure 3.3: Passenger movements for northbound Orange Line, 5:15 – 5:30pm (Source: ODX Data from MBTA Research Database)

3.2.1 Direct Observations

Direct observations were collected of doors opening and closing, passengers left behind on the platform, and the passengers entering and exiting the platform. Table 6.1 through Table 6.4 in Appendix B present the raw counts for the four direct observation events: Sullivan Square and North Station in November 2017 and January 2018. Trains are numbered in order of arrival to the station during the time period of interest. The door opening and closing times are based on the directly observed times. Passengers waiting on the platform is the sum the passengers left behind by the previous train and the number of passengers that entered the platform since the previous train's doors closed. The number of left-behind passengers are reported for each of the three observed locations and the platform total.

Time series of the number of passengers on the platform are shown in Figure 3.4 through Figure 3.7. The sawtooth pattern in each figure shows the growing number of passengers on the platform as time elapses since the previous train. The sudden drops correspond to the times when doors close. When no passengers are left behind, the platform count drops to zero. When passengers are left behind, the time series drops to that number of left-behind passengers (shown in red in the figures).

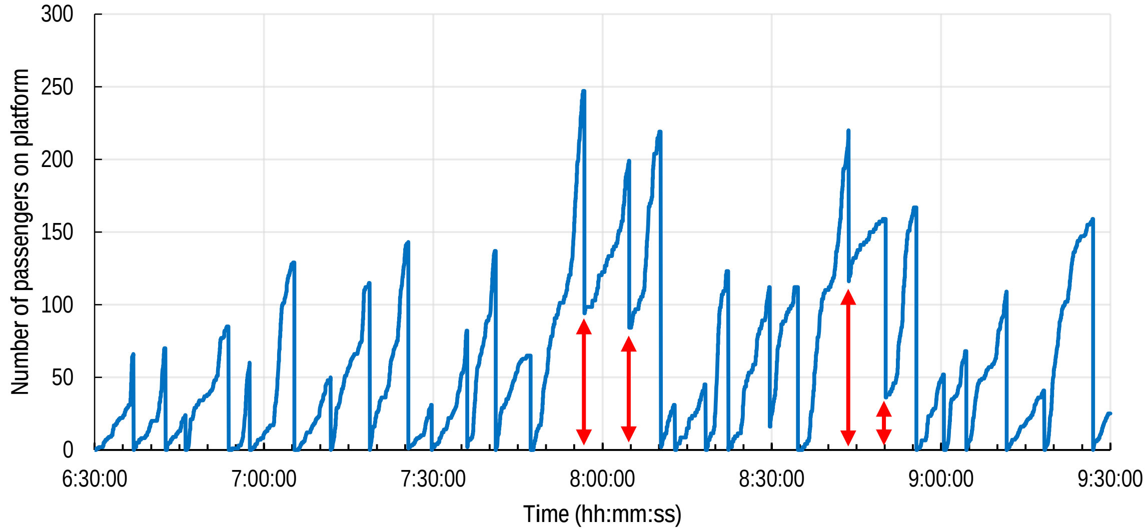


Figure 3.4: Time series of passengers on platform from manual counts, Sullivan Square, November 15, 2017

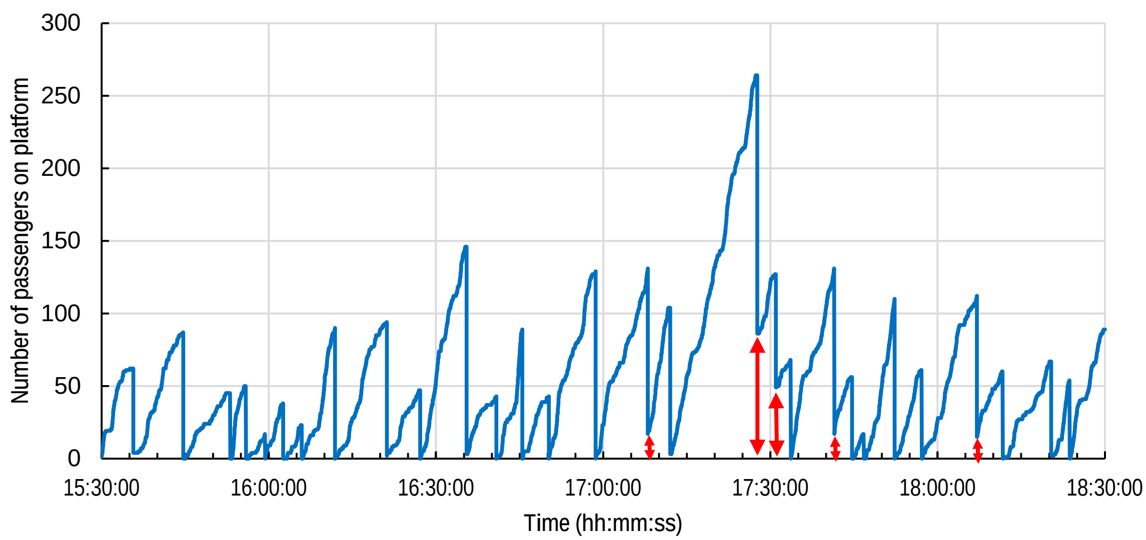


Figure 3.5: Time series of passengers on platform from manual counts, North Station, November 15, 2017

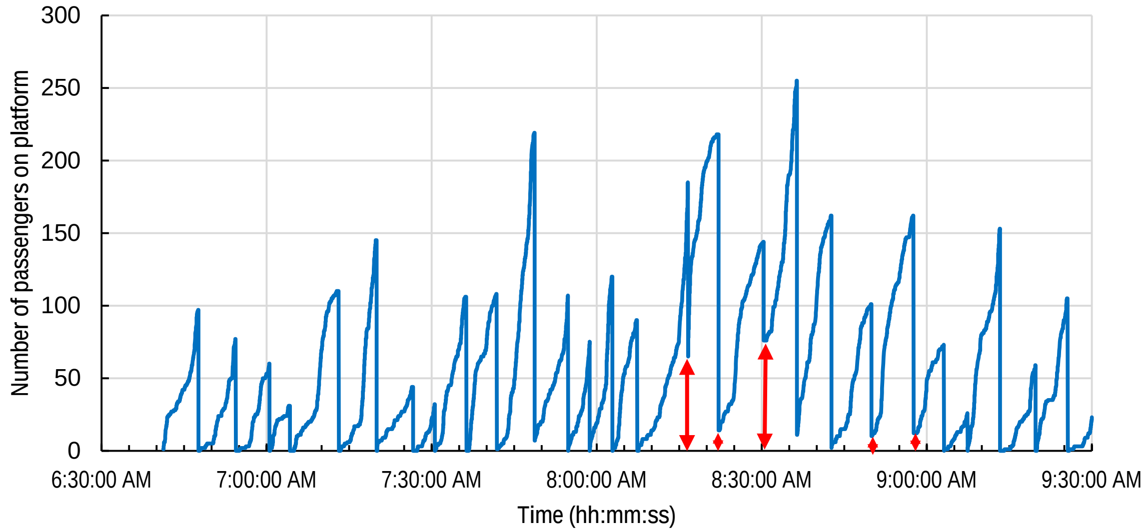


Figure 3.6: Time series of passengers on platform from manual counts, Sullivan Square, January 31, 2018

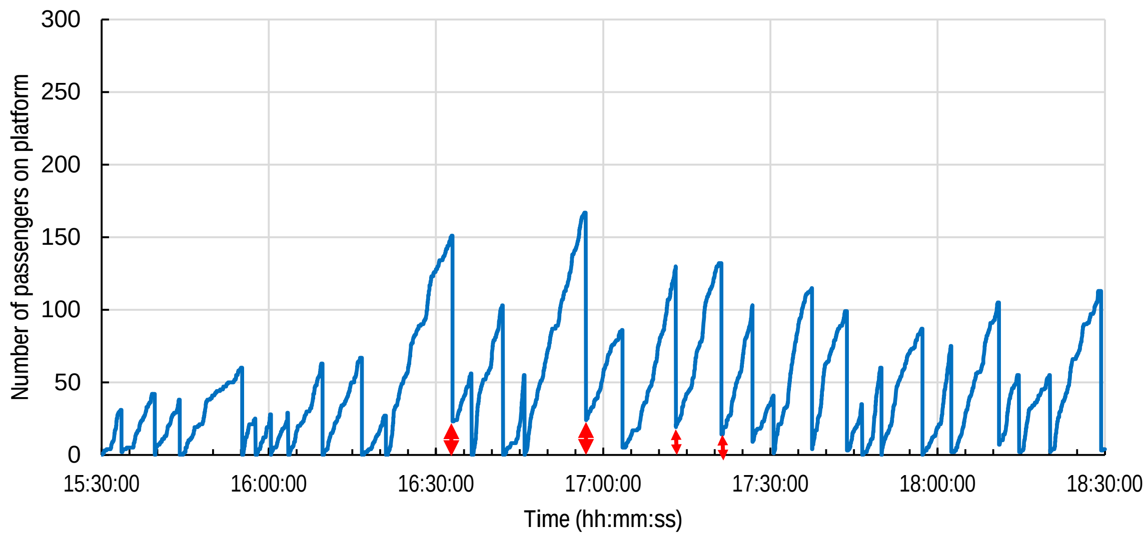


Figure 3.7: Time series of passengers on platform from manual counts, North Station, January 31, 2018

The pattern of passenger arrivals to the platform appear to be independent of the departure times of trains. This makes sense since the Orange Line operates on a short headway without a specific schedule of departures. At Sullivan Square, the arrivals tend to be more clustered, because many passengers are transferring from buses. In all cases, there is a clear pattern that longer headways correspond to greater accumulations of waiting passengers and increased likelihood of passengers being left behind.

An analysis of the time that each passenger waits on the platform is performed based on the assumption that passengers board trains in the order that they enter the platform (see Section 2.7.2). Figure 3.8 through Figure 3.11 show the cumulative distributions of waiting times in each of the four observation periods (blue curve). This is compared to the distribution of waiting times that would be experienced if no passenger were left behind (red curve). In each case, the reliability standard specified in the 2017 MBTA Service Delivery Policy is for 90% of passengers to wait less than the 6 minute (360 second) published headway. When left-behind passengers are accounted for, the waiting times that they actually experience are longer than if every passenger is able to board the next departing train. As a result, the reliability measure is lower when passengers are recognized as being left behind. It should be noted that the reliability standard is a systemwide measure across locations and times of day, so low reliability at Sullivan Square and North Station during peak hours does not imply that the whole system is underperforming to the same extent.

The direct observations of passenger counts, left-behind passengers, and waiting times is summarized in Table 3.2. In all cases, passengers were observed being left behind by some crowded trains, and accounting for these left-behind passengers has implications for the measures of reliability (i.e., percent of passengers waiting less than a scheduled headway) and the average experienced waiting time.

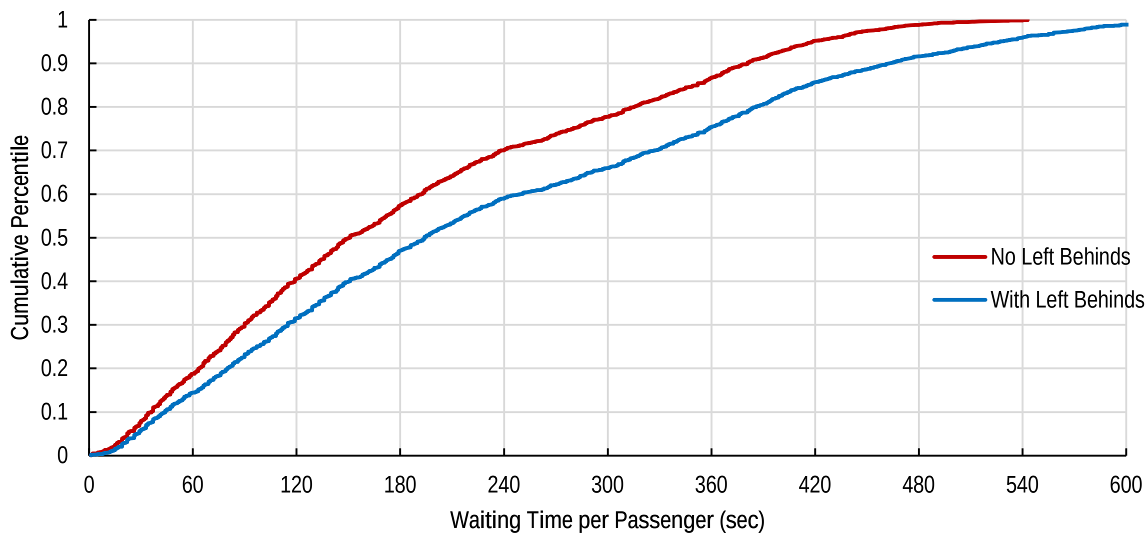


Figure 3.8: Cumulative distribution of passenger waiting times at Sullivan Square, November 15, 2017

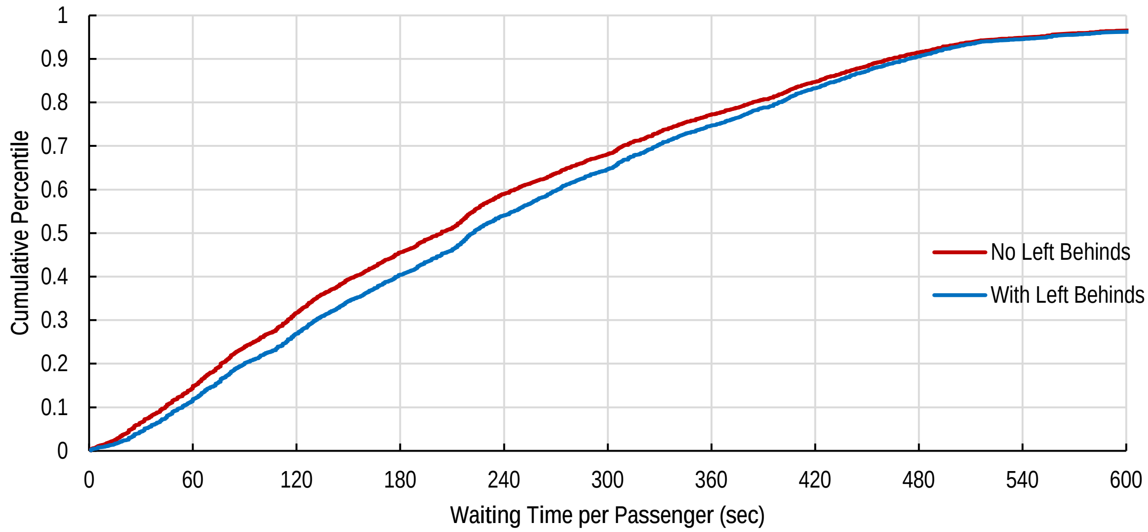


Figure 3.9: Cumulative distribution of passenger waiting times at North Station, November 15, 2017

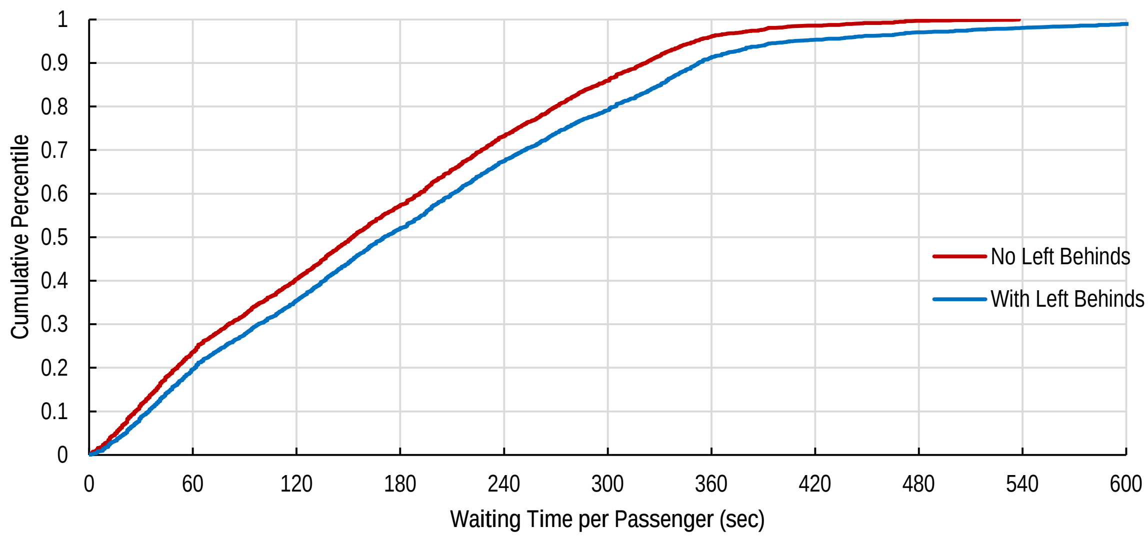


Figure 3.10: Cumulative distribution of passenger waiting times at Sullivan Square, January 31, 2018

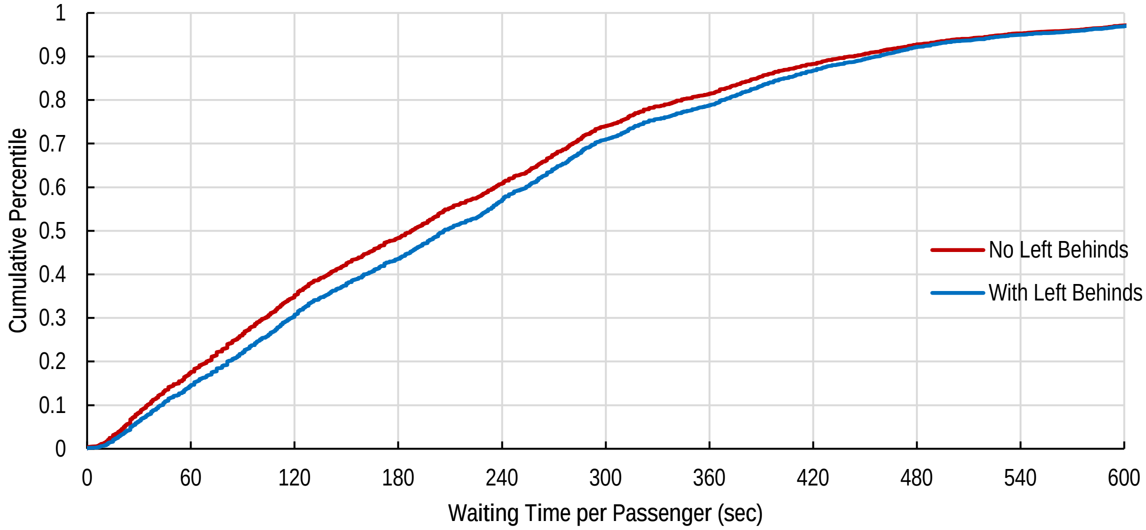


Figure 3.11: Cumulative distribution of passenger waiting times at North Station, January 31, 2018

Table 3.2: Summary of Left-Behind Passengers and Waiting Times Observed

	Wednesday, November 15, 2017		Wednesday, January 31, 2018	
	Sullivan Square	North Station	Sullivan Square	North Station
# of Trains	29	29	27	30
# of Trains Leaving Passengers	8	9	9	11
# of Passengers	2,681	1,503	3,064	2,202
# of Passengers Left Behind	351	198	198	120
Reliability (not including LB)	86.6%	77.2%	96.0%	78.7%
Reliability (including LB)	75.3%	74.6%	91.2%	81.3%
Average Wait Time, sec (not including LB)	183	237	164	216
Average Wait Time, sec (including LB)	232	253	189	231

According to the MBTA Performance Dashboard, November 15, 2017, and January 31, 2018 had good system performance. The overall subway reliability is reported as 88% and 87% on these dates, respectively. It is not surprising that rush hour conditions at especially crowded stations appear to perform worse than the systemwide average.

One point that is worth noting is that January 31, 2018, was an especially cold day with morning temperatures well below freezing. Since the station platform at Sullivan Square is above ground and exposed to wind, passengers have an incentive to wait in the lobby area until the train is arriving before descending the stairs to board the train. Since counts of passengers entering the platform were conducted based on stairwell flows, it is possible that some of the waiting time at Sullivan Square was not accounted for on January 31. This would explain the higher reliability and lower average waiting times compared to November 15. This effect would not impact the counts of left-behind passengers, because passengers expect to be able to board a train once they are on the platform. Nevertheless, the cold conditions may have had some effect on changing passengers' behaviors and encouraging more people to attempt crowding onto full trains rather than waiting.

3.2.2 Automated Video Counts

The automated video processing algorithm was used to analyze the relevant surveillance video feeds with views of the platforms, as described in Section 2.5.3. Figure 3.12 shows the time series of passenger counts at North Station from November 15, 2017. The blue line represents manual counts of passengers on the platform, which is considered the ground truth. The green line is the smoothed time series from the three surveillance feeds used to monitor the northbound Orange Line platform at North Station.

The automated passenger counting algorithm clearly undercounts the total number of passengers on the platform. The reason for this large discrepancy is that the algorithm can only identify people in the foreground of the images, when each person is large enough for the algorithm to identify. Therefore, the three camera views do not actually provide complete coverage of the platform for automated counting purposes. Furthermore, when conditions get very crowded, it becomes more difficult to identify separate bodies within the large mass of people. Even for a human, it is difficult to accurately count passengers in the crowded distance parts of the platform, as in Figure 3.13.

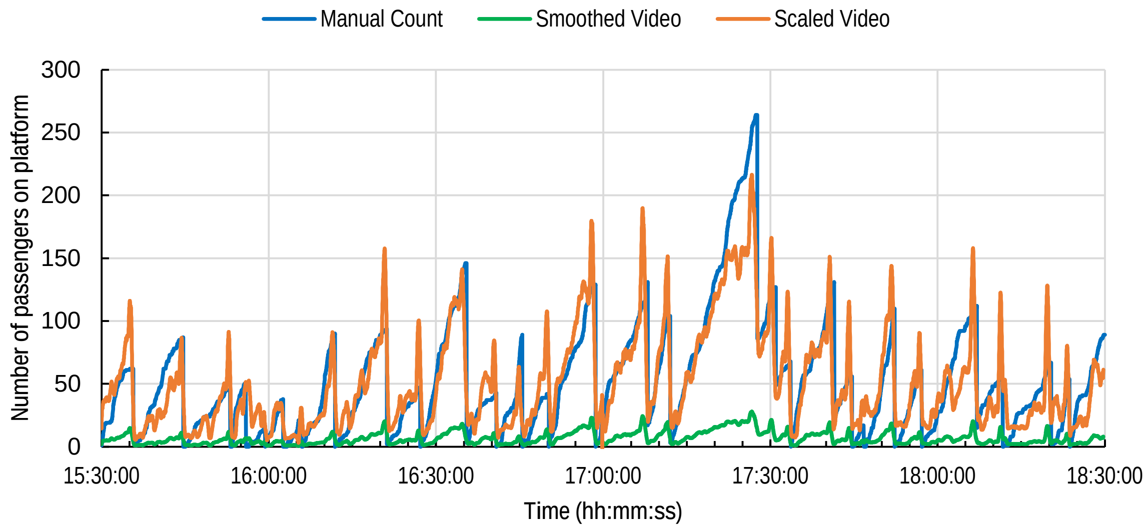


Figure 3.12: Automated passenger counts from surveillance video at North Station, November 15, 2017

Figure 3.13: Example camera frame showing identified passengers at North Station

The problem of undercounting aside, it is clear that the automated counts generate a pattern that is representative of the total number of passengers on the platform. Using regression, the smoothed time series can be linearly transformed into a scaled time series (shown in orange), which minimizes the squared error compared with the manually counted time series. For the case of North Station, the following regression model with $R^2 = 0.74$ converts the smoothed datapoints into a scaled estimate of the number of passengers on the platform:

$$Scaled(t) = -1.07 + 7.81Smoothed(t) \tag{5}$$

Using this scaling method, the data from November 15, 2017, was used to compare estimated counts of left-behind passengers in the peak periods with the directly observed values. This provides a measure of the accuracy of automated video counts. The total number of left-behind passengers at Sullivan Square and North Station are presented in Table 3.3, where the mean absolute error (MAE) and root mean squared error (RMSE) are calculated by comparing the number of passengers left behind each time the train doors close.

Table 3.3: Accuracy of Video Counts of Left-Behind Passengers, November 15, 2017

	Sullivan Square, 6:30 – 9:30am			North Station, 3:30 – 6:30pm		
	Total Left-Behinds	MAE	RMSE	Total Left-Behinds	MAE	RMSE
Manual Observation	351			198		
Unscaled (Smoothed) Video Count	103	11.8	28.2	73	6.8	16.7
Scaled Video Count	582	12.7	15.0	336	9.2	11.9

The automated video counts do not accurately reflect the true number left-behind passengers. The unscaled video count undercounts passengers for the same reason that the smoothed time series lies well below the manual count of passengers on the platform; i.e., the countable area in the surveillance videos does not cover the entire platform. The scaled videos, on the other hand, substantially overcount left-behinds, because the scaling factor that matches the manually counted time series most closely tends to over-inflate the counts when there are few passengers on the platform. As a direct measurement method, automated video counting, at least as implemented with YOLO, is not satisfactory. However, Figure 3.12 shows a clear relationship between the video counts and the number of passengers being left behind on station platforms. Therefore, there is a potential to use the video feed as an explanatory variable in a logistic regression model.

3.2.3 Wireless Device Detection

The wireless device detection units were mounted in Sullivan Square and North Station after regular rail operations were completed shortly after midnight on Tuesday, January 30, 2018. The placement was as described in Section 2.6. The devices were left running to log all MAC addresses in the station at the highest temporal resolution, as often as 1 Hz. When the devices were later taken down and returned to the lab, the records showed that MAC addresses were logged for about 20 hours, which is shorter than expected (perhaps due to the effect of cold temperatures on battery life) but long enough to cover both the morning and evening peak periods. In total, over 1.4 million observations were collectively logged by the four units. A summary of the observations is presented in Table 3.4.

Table 3.4: Combined Bluetooth and Wi-Fi Observations, January 31, 2018

	Sullivan Square, 6:30 – 9:30am		North Station, 3:30 – 6:30pm	
Data	Box 3	Box 4	Box 1	Box 2
Total Number of Observations (~20 hours)	187,732	439,294	306,156	553,673
Observations During Peak (3 hours)	55,628	115,719	79,239	128,425
Unique MAC Addresses in Peak	16,396		12,431	
Filtered MAC Addresses	3,963		8,406	

From the initial data log, the filtering process to keep records for unique MAC addresses resulted in 3,963 records for Sullivan Square and 8,406 records for North Station. These are the records associated with observed durations of greater than 5 seconds, less than 960 seconds, and within 120 seconds of a departing train. The greater number of observations at

North Station is attributable to the greater number of total passengers using the Green and Orange lines that service the station.

By extracting the observations associated with trains traveling in the direction of interest, the first observation times are used to create a cumulative arrival count to the platform, and the last observation times are used to create a cumulative boarding count onto trains. From these cumulative count curves, the time series of the number of devices on the platform of interest are tracked. The time series for North Station is shown in Figure 3.14, compared with the manual count of passengers on the platform.

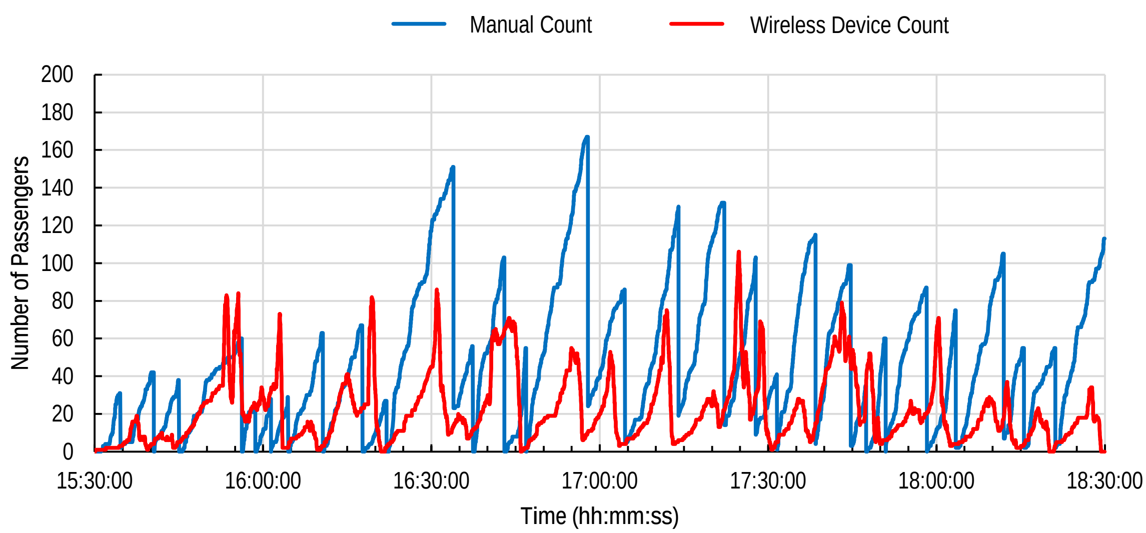


Figure 3.14: Wireless device counts from North Station, January 31, 2018

The time series of wireless device counts exhibits a sawtooth pattern like that of the manual and video counts as a consequence of the filtering process. The difference is that there is no statistically significant linear transformation that can rescale the wireless count time series to predict manually counted values. In fact, the implied number of left-behind devices does not appear to be correlated with the observed number of left-behind passengers. This makes wireless device detection unsuitable as a tool for directly counting the number of passengers left behind by crowded vehicles.

An alternative way to view the wireless device detection data is to plot the cumulative distribution of the observed durations and compare it against the cumulative distribution of waiting times estimated from the manual counts, as shown in Figure 3.15. If wireless devices are considered to be a proxy for the passengers who carry them, they make a poor indicator of experienced waiting time. This discrepancy is likely due in large part to the fact that devices are not logged every second that they are in the station, even though this is theoretically possible. The data shows that devices are pinged on random intervals averaging about once per minute, but ranging from a few seconds to more than two minutes. This variability leads to latency in which a device may be in the station for one or two minutes before it is first detected and it may linger for one or two minutes after it is last detected. The steep slope associated with wireless devices for low waiting times is associated with a relatively large number of observations. Since peak hour headways on the Orange Line are 6 minutes on average, losing a couple of minutes at the beginning and end of observation can contribute substantial errors and lead to undercounting the waiting time.

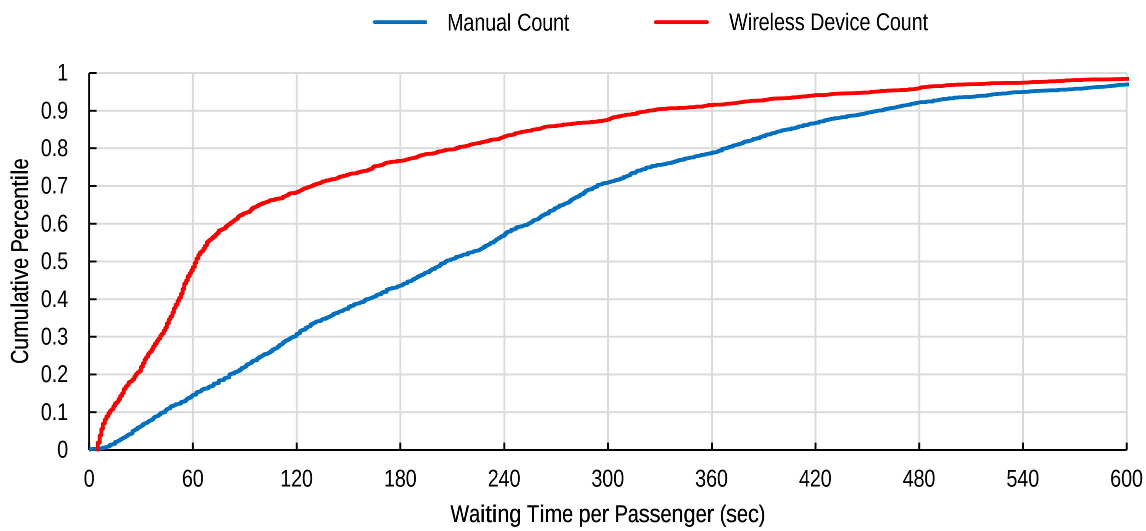


Figure 3.15: Comparison of distribution of waiting times measured by wireless devices and manual observation, North Station, January 31, 2018

The data suggest that latency associated with wireless device detection is too large to use Bluetooth and Wi-Fi detection to track passengers during their relatively short waits within a transit station. The technology is likely better suited to origin-destination matching and travel time measures, which is beyond the scope of this study.

3.3 Estimation of Models

The comparison of manual observations of left-behind passengers and automated video counts revealed inaccuracies in the magnitudes of direct counts. However, the patterns revealed through automated counts of surveillance videos shows promising potential to predict the occurrence of passengers being left behind with the use of modeling techniques. First, a simple model is fitted to estimate train door opening and closing times from train-tracking records. Then these train-tracking data are used along with video counts to fit logistic regression models in order to estimate the probability of passengers being left behind by crowded trains.

3.3.1 Door Opening and Closing Times

The directly observed door opening and closing times were also compared with the reported TTR times in order to determine how accurately the door times can be estimated without manual observations.

First, a comparison of the precise door closing times and the reported TTR times are shown for Sullivan Square and North Station in Table 3.5. The door closing time of train i is denoted $DC(i)$ and the track circuit departure time is denoted $TTR_d(i)$. This relatively stable relationship allows the TTR departure times to be used to estimate the precise time that doors close and the headway between trains. Headway is a potentially important predictor of left-behind passengers during peak periods.

Table 3.5: Comparison of Door Closing Times and TTR Departure Times, November 15, 2017

	Sullivan Square, 6:30 – 9:30am	North Station 3:30 – 6:30pm
Number of Observations	29	29
Average Duration $\boldsymbol{TTR_d(i) - DC(i)}$ (sec)	17.5	13.3
St. Dev. of $\boldsymbol{TTR_d(i) - DC(i)}$ (sec)	2.2	3.6

Second, dwell time between doors opening and closing may be estimated from the difference between TTR arrival and departure times using linear regression. The regression results for Sullivan Square and North Station are shown graphically in Figure 3.16 and Figure 3.17, respectively. The dwell time is also a potentially important predictor of left-behind passengers, because longer dwell times are associated with greater crowding on trains.

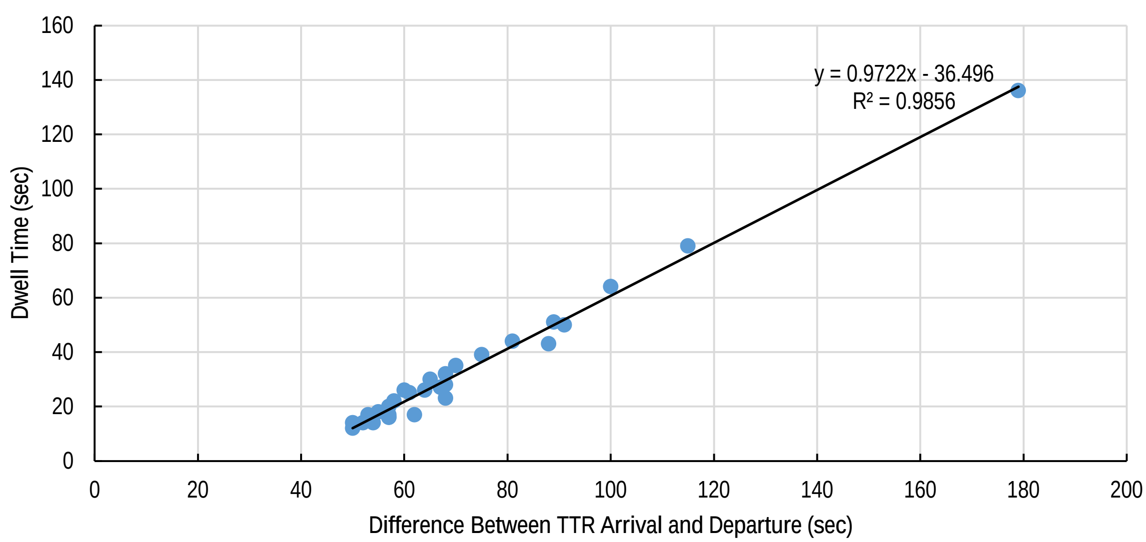


Figure 3.16: Regression of Dwell Time and Track Circuit Data for Sullivan Square, November 15, 2017

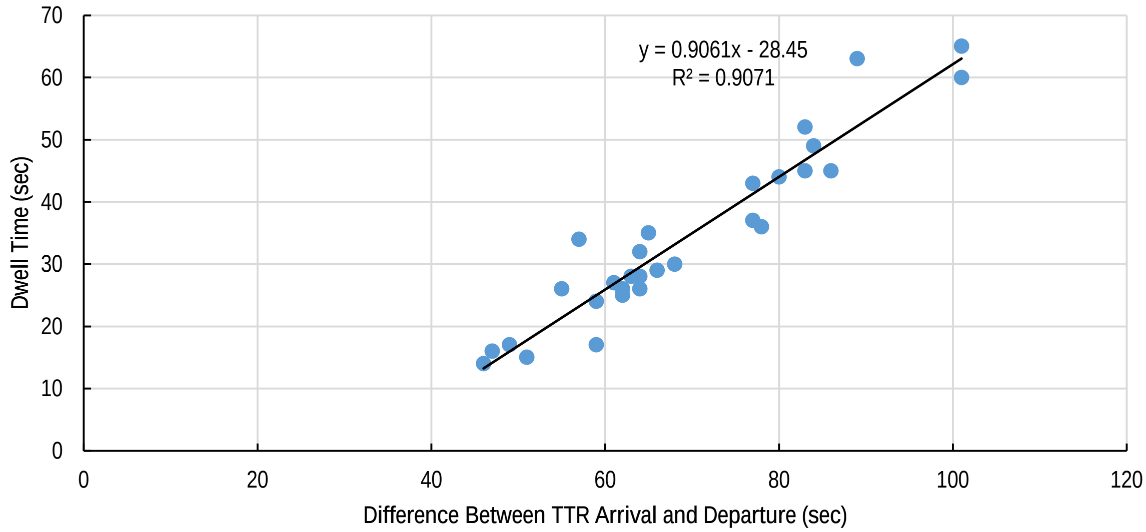


Figure 3.17: Regression of Dwell Time and Track Circuit Data for North Station, November 15, 2017

These robust relationships allow headways and dwell times to be estimated from TTR data which are automatically tracked rather than requiring manual observations which are costly to conduct over long periods of time.

3.3.2 Models of Likelihood of Passenger Left-Behinds

Following the logistic regression modeling structure presented in Section 2.7.1, the observations from November 15, 2017 were used to estimate the parameters of logistic regression models with various combinations of explanatory variables. The goal of the model is to support accurate prediction of the number of passengers that are left behind by crowded trains using data sources that can be collected and processed automatically without requiring additional manual counts.

Initially, three models were estimated, making use of only TTR data (Model 1), only video counts (Model 2), and then both TTR and video together (Model 3). A summary of the estimated model coefficients and fit statistics is presented in

Table 3.6 for Sullivan Square and Table 3.7 for North Station. The negative constant indicates that, all else being equal, passengers are less likely to be left behind than to board a train. Negative coefficients imply that increasing the explanatory variable's value decreases the likelihood of being left behind, whereas a positive coefficient implies increased likelihood of a passenger being left behind.

The log likelihood is a measure of how well the estimated probability of a passenger being left behind matches the observations. The null log likelihood is associated with no model at all (every data point is assigned a 50% chance of being left behind), and values closer to zero indicate a better fit. The ρ^2 value is a related measure of model fit, with values closer to 1 indicating a better model.

Table 3.6: Logistic Regression Model Parameters for Sullivan Square, November 2017

Parameter	Model 1		Model 2		Model 3	
	Value	p-stat	Value	p-stat	Value	p-stat
Constant for Left-Behind	-9.5	0.00	-2.7	0.00	-10.4	0.00
Dwell Time (sec)	-0.015	0.00			-0.016	0.00
Headway (sec)	0.019	0.00			0.022	0.00
Video Count			0.118	0.00	-0.057	0.01
Null Log Likelihood, $\boldsymbol{LL_0}$	-2197		-2197		-2197	
Model Log Likelihood, $\boldsymbol{LL}$	-803		-1058		-800	
$\boldsymbol{\rho^2}$	0.634		0.518		0.636	

Table 3.7: Logistic Regression Model Parameters for North Station, November 2017

Parameter	Model 1		Model 2		Model 3	
	Value	p-stat	Value	p-stat	Value	p-stat
Constant for Left-Behind	-10.5	0.00	-4.2	0.00	-10.0	0.00
Dwell Time (sec)	0.100	0.00			0.091	0.00
Headway (sec)	-0.001	0.08			-0.003	0.00
Video Count			0.370	0.00	0.291	0.00
Null Log Likelihood, $\boldsymbol{LL_0}$	-1639		-1639		-1639	
Model Log Likelihood, $\boldsymbol{LL}$	-550		-533		-494	
$\boldsymbol{\rho^2}$	0.662		0.675		0.699	

In order to compare two models, a likelihood ratio statistic is used to determine whether the improvement of one model is statistically significant compared to another. The likelihood ratio test statistic is calculated by comparing the log likelihood of the restricted model (with fewer explanatory variables) to the log likelihood of the unrestricted model (with more explanatory variables):

$$D = 2(LL_{unrestricted} - LL_{restricted}) \tag{6}$$

Comparing Model 1 (restricted) to Model 3 (unrestricted), there is one additional variable in Model 3, indicating 1 degree of freedom. To reject the null hypothesis at the 0.05 significance level with 1 degree of freedom, $D > 3.84$. For both Sullivan Square ($D = 7.2$) and North Station ($D = 113.7$), Model 3 is a statistically significant improvement over Model 1, meaning that that video counts add explanatory power to the model. There is a greater improvement from the video at North Station. The model parameters at Sullivan Square raise some question about correlation among the explanatory variables, because it does not make intuitive sense for the video count to have negative coefficient. This implies that greater video counts lead to a lower probability of passengers being counted as left behind.

The data collection on November 15, 2017, did not include wireless device detection, so a full comparison of model specifications could not be completed with the initial dataset. In order to evaluate the potential explanatory power of the wireless device counts, Models 1 and 3 were estimated again with the January 31, 2018 data and compared with two additional model specifications: using TTR and wireless device data (Model 4), and using all data sources (Model 5). The results are compared for North Station in Table 3.8. The coefficient for wireless device counts is not statistically significant at the 0.05 level.

Table 3.8: Logistic Regression Model Parameters for North Station, January 2018

Parameter	Model 1		Model 3		Model 4		Model 5	
	Value	p-stat	Value	p-stat	Value	p-stat	Value	p-stat
Constant for Left-Behind	-9.5	0.00	-9.4	0.00	-9.4	0.00	-9.4	0.00
Dwell Time (sec)	0.014	0.00	0.016	0.00	0.081	0.00	0.016	0.00
Headway (sec)	0.005	0.59	0.001	0.77	0.000	0.90	0.000	0.98
Video Count			0.043	0.68			-0.018	0.71
Wireless Device Count					-0.092	0.06	-0.099	0.06
Null Log Likelihood, LL_0	-1549		-1549		-1549		-1549	
Model Log Likelihood, LL	-406		-406		-404		-403	
ρ^2	0.738		0.738		0.739		0.739	

3.4 Validation of Models of Left-Behind Passengers

The validation test for the proposed models is used to estimate parameters from observations in November 2017 to predict the number of left-behind passengers from automated data sources in January 2018. The validation procedure is to make estimates using only dwell times and headways estimated from TTR data and video counts with scaling factors and regression results fitted with the November data.

3.4.1 Estimation of the Number of Passengers Left Behind

The logistic regression provides an estimate of the probability that passengers are left behind each time a train closes its doors. In order to translate this probability into a passenger count, an estimate of the number of passengers waiting on the platform is needed. For this purpose, the scaled video count of passengers on the platform is used as an estimate of the number of passengers waiting to board. Table 3.9 shows the validation results when the models were applied to the January 2018 data for North Station.

As shown in Table 3.3, video counts do not provide accurate estimates of the total numbers of passengers left behind without some additional modeling. The unscaled video counts underestimate the total, while the scaled video counts overestimate the total. The logistic regression provides much better results. Although there are some discrepancies for specific train departures, the estimated numbers of passengers left behind are not significantly biased and the total number of passengers left behind during the rush hour is similar to the manually counted total.

An important note about the two logistic regressions is the probability of a passenger being left behind is calculated using only the explanatory variables listed in Table 3.7. However, the estimated number of left-behind passengers is calculated by multiplying the probability by the scaled video count of passengers on the platform at the time the doors opened as estimated from the TTR data. Therefore, the estimated number of passengers left behind with

Model 1 and Model 3 rely only on TTR data that is currently being logged and supplemented by automated counts of passengers from existing surveillance video feeds.

Table 3.9: Validation of Estimated Left-Behind Passengers, North Station

Train	Manual Count		Unscaled Video	Scaled Video	Model 1		Model 3	
	Probability	Number	Number	Number	Probability	Number	Probability	Number
1	6.5%	2	1	7	2.4%	2	0.6%	1
2	0.0%	0	2	15	0.8%	0	0.7%	0
3	0.0%	0	2	15	0.3%	0	0.3%	0
4	0.0%	0	1	7	0.5%	0	0.2%	0
5	0.0%	0	0	0	0.3%	0	0.2%	0
6	0.0%	0	0	0	0.7%	0	0.5%	0
7	0.0%	0	0	0	0.6%	0	0.4%	0
8	0.0%	0	2	15	1.0%	1	0.8%	0
9	0.0%	0	2	15	1.8%	1	1.2%	1
10	0.0%	0	1	7	0.6%	0	0.4%	0
11	15.2%	23	2	15	16.5%	29	5.2%	9
12	0.0%	0	1	7	1.1%	1	1.0%	1
13	0.0%	0	1	7	1.9%	1	1.1%	1
14	0.0%	0	2	15	0.7%	0	0.7%	0
15	14.4%	24	3	22	14.4%	21	6.7%	10
16	5.8%	5	4	30	7.0%	8	7.7%	8
17	14.6%	19	3	22	19.5%	29	10.9%	16
18	10.6%	14	10	77	18.9%	25	52.2%	69
19	8.7%	9	3	22	6.6%	4	6.3%	3
20	2.4%	1	4	30	0.8%	1	1.6%	1
21	3.5%	4	3	22	3.6%	3	3.0%	3
22	3.0%	3	2	15	1.5%	2	1.1%	1
23	0.0%	0	1	7	0.7%	0	0.7%	0
24	0.0%	0	3	22	0.7%	0	1.1%	1
25	0.0%	0	3	22	0.9%	1	0.8%	1
26	2.7%	2	2	15	2.6%	1	2.1%	1
27	6.7%	7	2	15	2.7%	3	1.4%	1
28	3.6%	2	3	22	0.5%	0	0.7%	1
29	3.6%	2	1	7	0.6%	0	0.4%	0
30	2.7%	3	2	15	2.0%	2	1.0%	1
Total		120	66	490		134		130
MAE			3.5	12.9		1.9		3.9
RMSE			6.5	17.6		3.3		10.8

3.4.2 Estimation of the Occurrence of Left-Behinds

Another way to evaluate the performance of the methods is to consider whether or not trains that leave behind passengers can be distinguished from trains that allow all passengers to board. Through the course of data collection and analysis, it appeared that passengers being left behind because of overcrowding can only be reliably observed within approximately ±2 passengers. The reason being that sometimes people choose not to board a train for reasons other than crowding and one or two passengers left on the platform did not appear to be consistent with problematic crowding conditions.

If a train is defined to be leaving behind passengers when more than 2 passengers are observed or estimated to be left behind, the results presented in Table 3.9 can be reinterpreted to evaluate each method by three measures:

- Correct Identification Rate – The percent of trains that are correctly classified as leaving behind passengers or not leaving behind passengers, as compared to the manual count. This value should be as close to 1 as possible.
- Detection Rate – The percent of departing trains that were observed to leave behind passengers that are also flagged as such by the estimation method. This value should be as close to 1 as possible.
- False Alarm Rate – The percent of departing trains that are estimated to leave behind passengers but have not, according to manual observations. This value should be as close to 0 as possible.

A comparison of the rates of detection are compared in Table 3.10 for the 30 trains that departed North Station between 3:30pm and 6:30pm on January 31, 2018. Unscaled and scaled video counts are poor estimators for the occurrence of left-behind passengers because the unscaled counts are consistently too low for detection and the scaled counts are high enough to trigger too many false alarms. The modeled estimates both perform well, never falsely identifying a train as leaving behind passengers when it did not, and correctly detecting most occurrences of passengers being left behind. Like the count estimates above, both Model 1 and Model 3 rely on the scaled video counts to estimate the number of passengers waiting on the platform when the train doors open, so a fusion of TTR records and automated video counts appears to provide the most reliable measure.

Table 3.10: Validation of Estimated Occurrence of Left-Behinds, North Station

	Manual Count	Unscaled Video	Scaled Video	Model 1	Model 3
Total Departing Trains	30	30	30	30	30
Trains Leaving Behind Passengers	8	3	27	6	5
Correct Identification Rate		0.77	0.37	0.93	0.90
Detection Rate		0.25	1.00	0.75	0.62
False Alarm Rate		0.33	0.70	0.00	0.00

3.4.3 Distribution of Experienced Waiting Times

A third application of the model results is to consider the distribution of waiting times implied by the modeled probabilities that passengers are being left behind by each departing train. As shown in Sections 2.7.2 and 3.2.1, the distribution of waiting times can be estimated by assuming that passengers board trains in the order that they arrive onto the platform. The simplest arrival assumption is to suppose that passengers arrive at a constant rate over the rush period. Then the probability of passengers being left behind, as estimated by the logistic regression, can be attributed to each departing train to estimate the durations of time that each passenger waits on the platform. By this process a cumulative distribution of waiting times can be estimated using probabilities from Model 1 and compared to the observed distribution as shown in Figure 3.18.

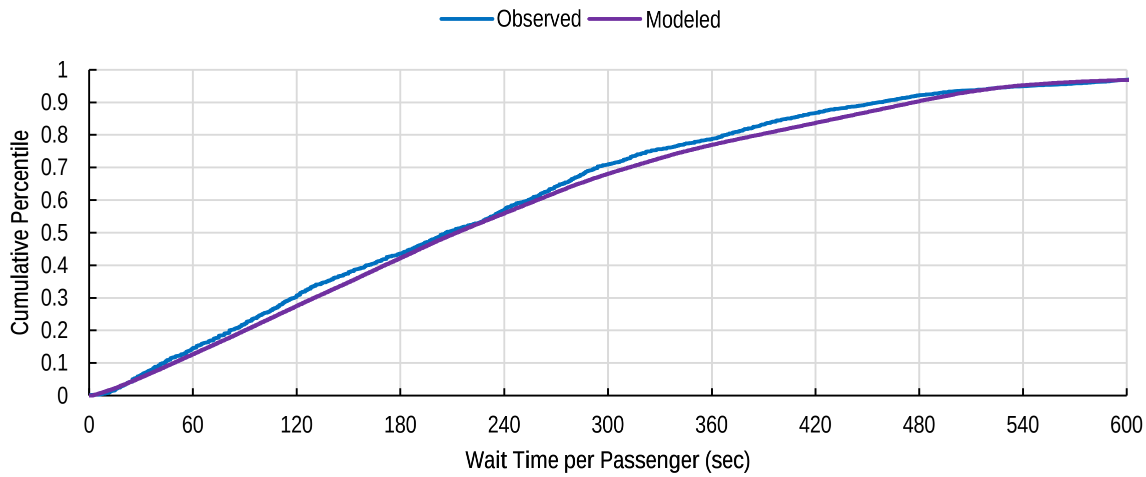


Figure 3.18: Modeled and Observed Distribution of Passenger Wait Time, North Station, January 31, 2018

The modeled distribution of waiting times closely approximates the observed distribution. This suggests that the estimated probabilities of passengers being left behind by each departing train are consistent with the overall passenger experience. From Figure 3.11, roughly 82% of passengers would be assumed to experience less than a published headway if left-behind passengers were not considered. The observed value appears to be 79%, and the modeled value is 77%. There is not a large difference in this case, because January 31, 2018, was characterized by generally smooth operations without many passengers being left behind.

This page left blank intentionally.

4 Conclusions

4.1 Tracking Left-Behind Passengers and Long Waits

This study has investigated the potential for measuring the occurrence of passengers being left behind when rapid transit trains are too crowded to board. Following a preliminary study of crowding conditions on the MBTA's Orange Line and Blue Line, the data collection and analysis activities focused specifically on southbound Orange Line trains at Sullivan Square in the morning peak and northbound Orange Line trains at North Station in the afternoon peak.

Data was collected on typical weekday peak periods and showed that overcrowding is a common problem, even on days without unusual disruptions to service. In four detailed data collection experiments (two mornings at Sullivan Square and two evenings at North Station), over 100 passengers were left behind in each case. This tendency for overcrowding to lead to passengers being left behind on station platforms is a sign that the system is operating very near capacity. It is apparent from the manual counts of passengers on station platforms that even small fluctuations in headways lead to overcrowded trains that result in left-behind passengers.

Counting the number of left-behind passengers and identifying when trains are leaving behind passengers are important for maintaining accurate performance metrics that reflect the actual customer experience. For example, the MBTA's Service Delivery Policy sets reliability goals for the percentage of customers that experience waiting times less than a published headway. Without accounting for left-behind passengers, the reliability metric is based on the false assumption that all passengers are able to board the next arriving train. Based on observed platform counts in this study, accounting for left-behind passengers would lower the reliability metric by 3 to 11 percentage points (Section 3.2.1). The passengers who are left behind are almost always able to board the subsequent train, with the exception of passengers traveling with strollers or suitcases, who were sometimes observed to wait for multiple trains before finding space to board.

This study specifically investigated the potential for measuring the number of passengers being left behind on station platforms using existing logged data sources and two potential data collection methods: automated passenger counting using surveillance video feeds, and wireless device detection. Although none of the data sources provided adequate predictive capability in isolation, the development of models to fuse data sources demonstrated good results for predicting the number of passengers left behind, identifying which specific trains leave passengers behind, and the distribution of experienced waiting times. Specifically, models that fused train-tracking records (TTR) with automated passenger counts from surveillance videos offered promising performance.

4.2 Strengths and Challenges of Using Surveillance Video

The analysis of automated passenger counting in surveillance video feeds was based on an implementation of a fast, open-source algorithm called You Only Look Once (YOLO). This algorithm was implemented with existing training sets that identify people as well as other objects. The performance is fast enough that frames from surveillance video feeds could potentially be analyzed in real time. Therefore, the results of the analysis based on YOLO provide a proof of concept that could be further improved.

The direct observations of passenger counts from the video feeds are associated with errors. The challenges of automated video detection most notably include:

1. Detection range is a small part of the whole platform – Since the algorithm uses pattern recognition methods to identify people, passengers are only identified in the foreground of surveillance video feeds where the whole body appears clearly and distinctly. Although the surveillance cameras are position to provide security views of the whole platform, only passengers in a subset of the area can be monitored.
2. Crowds make individuals difficult to distinguish – Clusters of people are difficult for the algorithm to break apart into individuals, because bodies are obscured from recognition. This causes undercounting of passengers in especially crowded conditions or in distant parts of the field of view.
3. Many camera angles are blocked or obscured – The positioning of several cameras leave parts of the platform area blocked from view by columns. Camera angles that are too low or too high also restrict the area over which the algorithm can reliably detect passengers.

These challenges result in systematic undercounting of the true number of passengers on the platform at any time. If the parts of the platform area that are observed are assumed to be a representative sample of all passengers waiting, then the automated counts still have some useful explanatory power. Although a simple rescaling of automated counts did not yield accurate estimates of the numbers of passengers left behind by crowded vehicles, the time series showed a clear relationship to the observed number of passengers on the platform (Section 3.2.2).

The accuracy of left-behind predictions was substantially improved by fusing video count data with train-tracking records in a logistic regression model to estimate the probability and number of passengers left behind. The logistic regression models fitted to November 2017 data and used to predict January 2018 left-behind counts exhibited good prediction capability:

- The total number of passengers estimated to be left behind during a rush period was estimated within 10% of the observed value.
- Departing trains were correctly identified as having left behind or not having left behind passengers more than 90% of the time.

- The reliability measure (percent of passengers waiting less than a published headway) was estimated within 2 percentage points of the observed value.

4.3 Strengths and Challenges of Using Wireless Device Detection

The use of wireless device detection did not perform as well as hoped for the application of measuring passenger waiting times in stations and estimating the number of passengers left behind. A challenge with the wireless detection method is that the data are noisy and not as reliable as video counts. There are several reasons for these challenges:

- Wireless device detection logs MAC addresses from devices rather than people – An inherent assumption in using this technology is that devices are a proxy for the passengers that carry them, but many wireless devices are built into the infrastructure and can be observed independently of the number of passengers using a station. Some passengers carry no device, while others may carry multiple wireless devices.
- Devices cannot be specifically located – Any MAC address that is observed within the antenna's range is logged as being present. In theory, the signal strength could be used to estimate the distance of a device from the antenna, but this requires specific knowledge about the type of device being observed and is further complicated by reflected signals in the station environment. In practice there is no way to know if a device is right beside the detection unit or on the other side of the station.
- Only discoverable devices can be observed – The antenna can only detect signals from devices that are set to discoverable Bluetooth mode or that are actively searching for a Wi-Fi connection. Many people leave their phones with one or both of these settings working, but not every wireless device will be detected.

Some of the challenges associated with the wireless device data can be addressed through the filtering process described in Section 2.6.2. Specifically, devices that appear to be permanent fixtures or passing by very briefly can easily be filtered out for having too long or too short a duration of observation.

The primary challenge with wireless device detection is that the sampling rate is random, highly variable, and relatively low on average. The average MAC address was observed approximately once per minute, even though the device detection unit was capable of logging MAC addresses every second. Furthermore, most devices are not logged at a regular 60 second interval but could be observed several times in the course of a few seconds and then not again until a couple of minutes later. These characteristics contribute to latency in which devices may be present in the station but not observed.

The problem in this case is that latency contributes significant errors to the time between a device entering the station platform and is first observed, as well as the time a device is last observed and when it leaves the platform. Since these errors are likely on the order to 1 to 2 minutes, and the average passenger is waiting approximately 4 minutes for a train, the observed duration of MAC addresses provides a poor measure of the waiting times that passengers experience.

The sampling rates were high in the sense that the units were able to detect thousands of unique MAC addresses within each peak period, so it is likely that MAC addresses at different locations could be matched to sample origin-destination travel times, which would also be less affected by latency.

4.4 Future Directions

There are a number of ways that the methods and finding of this study could be improved and extended. In the area of image processing to count passengers, the YOLO algorithm that was employed in this study provided a simple implementation to test the concept. A number of steps could be taken to improve the accuracy of video counts and extend the feasibility to more challenging station environments:

1. Compare the algorithm with other fast and accurate video detection algorithms – There are other image processing algorithms that can quickly analyze video footage to identify patterns. This is a quickly evolving field, so new tools, including open-source algorithms, have become available since the video analysis was conducted for this project. Alternative algorithms would likely reduce the number of false negatives in passenger counts.
2. Add tracking to link observations in consecutive frames – The algorithm implemented in this study analyzes each video frame independently, which introduces some noise in the raw data feed. Since passengers do not move very fast, a person identified in one frame provides information to the algorithm about where to look in the subsequent frame to reidentify the same person. This would be expected to improve the accuracy of the counts by reducing false positives. This functionality would also allow some tracking of movements, for example to confirm which train a specific individual boards. This would allow the model to be extended to branching lines in which passengers may be intentionally waiting to board a train that is headed for a specific branch.
3. Train the algorithm to detect heads rather than whole bodies – The current algorithm uses pattern recognition to identify a person by their whole body. In crowds of people, the whole body is not visible, so counting heads may provide a more accurate object for identification. It is not clear how well this will work, but it is likely to improve the accuracy of counts in crowded conditions.

A number of other opportunities exist to estimate system performance and crowding measures through data fusion. The inferred origin-destination-transfer (ODX) model has some known drawbacks given existing limitations of the fare card system, but ODX records may provide a useful data source as a modeling input. Future investments in MBTA infrastructure, including automated passengers counters on new trainsets and Automated Fare Collection 2.0, will provide insights about where and when passengers travel and where crowding is occurring. Although there are limitations to any single data source, the potential for improving performance metrics through data fusion and modeling continue to grow.

5 References

1. Massachusetts Bay Transportation Authority (2017). *Service Delivery Policy*. Accessed April 10, 2018. https://cdn.mbta.com/sites/default/files/2017-09/About%20the%20T/MBTA%20Service%20Delivery%20Policy%202017%20FINAL(1).pdf

2. Kittelson & Associates, Inc., Parsons Brinckerhoff, KFH Group, Inc., Texas A&M Transportation Institute, Arup (2013). *Transit Service and Quality of Service Manual, TCRP Report 165*. Transportation Research Board of the National Academies, Washington, D.C.

3. Mahudin, N.D.M., Cox, T., Griffiths, A. (2012). Measuring rail passenger crowding: Scale development and psychometric properties. *Transportation Research Part F*, 15(1):38–51.

4. Li, Z., Hensher, D.A. (2013). Crowding in public transport: A review of objective and subjective measures. *Journal of Public Transportation*, 16(2):107–134.

5. Turner, S., Corbett, E., O’Hara, R., White, J. (2004). Health and safety effects of rail crowding: Hazard identification. HSL Report RAS/04/12.

6. Cox, T., Houdmont, J., Griffiths, A. (2006). Rail passenger crowding, stress, health and safety in Britain. *Transportation Research Part A*, 40(3):244–258.

7. Evans, G.W., Wener, R.E. (2007). Crowding and personal space invasion on the train: Please don’t make me sit in the middle. *Journal of Environmental Psychology*, 27(1):90–94.

8. Li, Z., Hensher, D.A. (2011). Crowding and public transport: A review of willingness to pay evidence and its relevance to project appraisal. *Transport Policy*, 18:880–887.

9. Haywood, L., Koning, M. (2015). The distribution of crowding costs in public transport: New evidence from Paris. *Transportation Research Part A*, 77:182–201.

10. Haywood, L., Koning, M., Monchambert, G. (2017). Crowding in public transport: Who cares and why? *Transportation Research Part A*, 100:215–227.

11. Hörcher, D., Graham, D.J., Anderson, R.J. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B*, 95:105–125.

12. Tirachini, A., Hurtubia, Dekker, T., Daziano, R.A. (2017). Estimation of crowding discomfort in public transport: Results from Santiago de Chile. *Transportation Research Part A*, 103:311–326.

13. Kim, K.M., Hong, S.-P., Ko, S.-J., Kim, D. (2015). Does crowding affect the path choice of metro passengers. *Transportation Research Part A*, 77:292–304.

14. Tirachini, A., Hensher, D.A., Rose, J.M. (2013). Crowding in public transport systems: Effect on users, operation and implication for the estimation of demand. *Transportation Research Part A*, 53:36–52.

15. De Palma, A., Lindsey, R., Monchambert, G. (2017). The economics of crowding in rail transit. *Journal of Urban Economics*, (forthcoming) doi:10.1016/j.jue.2017.06.003.

16. Lee, Y.-J., Vuchic, V.R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, 131(1):1–10.

17. Lam, W.H.K., Cheung, C.-Y., Lam, C.F. (1999). A study of crowding effects at the Hong Kong light rail transit stations. *Transportation Research Part A*, 33(5):401–415.

18. Batarce, M., Muñoz, J.C., de Dios Ortúzar (2016). Valuing crowding in public transport: Implications for cost-benefit analysis. *Transportation Research Part A*, 91:358–378.

19. Moccia, L., Laporte, G. (2016). Improved models for technology choice in a corridor with fixed demand. *Transportation Research Part B*, 83:245–270.

20. Zhu, Y., Koutsopoulos, H.N., Wilson, N.H.M. (2017). Inferring left behind passengers in congested metro systems from automated data. *Transportation Research Procedia: 22nd International Symposium on Transportation and Traffic Theory*, 23:362–379.

21. Zhao, J., Rahbee, A., Wilson, N.H.M. (2007). Estimating a rail passenger origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, 22(5):376–387.

22. Barry, J., Freimer, R., Slavin, H. (2009). Use of entry-only automatic fare collection data to estimate linked transit trips in New York City. *Transportation Research Record*, (2112):53–61.

23. Paul, E.C. (2010). Estimating train passenger load from automated data systems: Application to London Underground. Master's Thesis. Massachusetts Institute of Technology.

24. Southwick, C.W. (2016). Understanding bus passenger crowding through origin-destination inference. Master's Thesis. Massachusetts Institute of Technology.

25. Hänseler, F.S., Bierlaire, M., Scarinci, R. (2016). Assessing the usage and level-of-service of pedestrian facilities in train stations: A Swiss case study. *Transportation Research Part A*, 89:106–123.

26. Hänseler, F.S., Molyneaux, N.A., Bierlaire, M. (2017). Estimation of pedestrian origin-destination demand in train stations. *Transportation Science*, (forthcoming) doi:10.1287/trsc.2016.0723.

27. Mukherjee, S., Saha, B., Jamal, I., Leclerc, R., Ray, N. (2011). A novel framework for automatic passenger counting. *18th IEEE International Conference on Image Processing*, 2969–2972.

28. Li, B., Zhang, J., Zhang, Z., Xu, Y. (2014). A people counting method based on head detection and tracking. *Smart Computing, 2014 International Conference*, 136–141.

29. Ozer, B., Wolf, M. (2014). A train station surveillance system: Challenges and solutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 638–643.

30. Yan-Yan, C., Ning, C., Yu-Yang, Z., Ke-han, W., & Wei-wei, Z. (2014). Pedestrian detection and tracking for counting applications in metro station. *Discrete dynamics in nature and society*, 2014, Article ID 712041.

31. Lo, B.P.L., Velastin, S.A. (2001). Automatic congestion detection system for underground platforms.

32. Dewei, L., Lin, Z., Yuxing, S. (2013). Crowd estimation using image processing based gray scale edge histogram in subway station of Beijing. *Measuring Technology and Mechatronics Automation, Fifth International Conference*, 415–419.

33. Sánchez-Martínez, G.E. (2017). Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. *Transportation Research Record*, 2652:1-7.

34. Kutilkin, A.V., Ivanov, S.V. (2016). A comparison of methods to detect people flow using video processing. *Procedia Computer Science*, 101:125–134.

35. Hou, Y.L., Pang, G.K. (2011). People counting and human detection in a challenging situation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 41(1):24–33.

36. Sonka M., Hlavac V., Boyle R. (2014). *Image Processing, Analysis, and Machine Vision.* Cengage Learning.

37. Viola P. (1993). *Feature-Based Recognition of Objects.* MIT AI Lab, AAAI Technical Report FS-93-04.

38. Cheng D., Wang J., Wei X., Gong Y. (2015). Training mixture of weighted SVM for object detection using EM algorithm. *Neurocomputing*, 149:473-482.

39. Z., Zhi-Hua. (2012). *Ensemble Methods: Foundations and Algorithms.* Chapman and Hall/CRC Press.

40. Deng J., Berg A. C., Li K., Fei-Fei L. (2010). What does classifying more than 10,000 image categories tell us? *European Conference on Computer Vision*, Heidelberg, Germany: Springer, 71-84.

41. Girshick R., Donahue J., Darrell T., Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580-587.

42. Lézoray O., Grady L. (2012). *Image processing and analysis with graphs: theory and practice.* Digital Imaging and Computer Vision, CRC Press / Taylor and Francis.

43. Coniglio C., Meurie C., Lézoray O., Berbineau M. (2017). People silhouette extraction from people detection bounding boxes in images. *Pattern Recognition Letters*, 93:182-191.

44. Deng J., Dong W., Socher R. Li L.-J., Li K. Fei-Fei L. (2009). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.

45. Everingham M., Van Gool L., Williams C. K., Winn J., Zisserman A. (2010). The Pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303-338.

46. Redmon, J. (2016). *Darknet: Open source neural networks in C.* Accessed April 10, 2018. http://pjreddie.com/darknet/.

47. Redmon J., Divvala S., Girshick R., Farhadi A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779-788.

48. Kurkcu, A., Ozbay, K. (2017). Estimating pedestrian densities, wait times, and flows using wi-fi and Bluetooth sensors. Paper Number 17-02095. *Transportation Research Board 96th Annual Meeting*, Washington, D.C.

49. Brennan, T.M., Ernst, J.M., Day, C.M., Bullock, D.M., Krogmeier, J.V., Martchouk, M. (2010). Influence of vertical sensor placement on data collection efficiency from Bluetooth MAC address collection devices. *Journal of Transportation Engineering*, 136(12):1104–1109.

50. Kostakos, V. (2008). Using Bluetooth to capture passenger trips on public transport buses. *arXiv preprint arXiv:0806.0874*

51. Dunlap, M., Li, Z., Henrickson, K., Wang, Y. (2016). Estimation of origin and destination information from Bluetooth and Wi-Fi sensing for transit. *Transportation Research Record*, 2595:11–17.

6 Appendices

6.1 Appendix A: Blue Line Crowding Analysis

A crowding analysis for the Blue Line was conducted in a similar manner as for the Orange Line. Figure 6.1 shows the resulting passenger occupancy estimates, $O(n, t)$, plotted across the stations for an average weekday to show where trains are consistently the most crowded. Trains grow increasingly crowded heading into central Boston in the AM peak, and the pattern reverses in the evening.

The most severe crowding on the Blue Line appears to be from 8:30 – 8:45am for southbound trains and 5:15 – 5:30pm for northbound trains. The values of $B(n, t)$, $A(n, t)$, and $O(n, t)$ are shown in Figure 6.2 for southbound trains in the morning. Inbound trains experience large boarding loads at Wonderland and Maverick. The combined circumstances at Maverick of fully loaded trains, many passengers boarding, and few alighting all contribute to crowding conditions that could lead to passengers being left behind. Figure 6.3 shows similar circumstances at State Street in the afternoon peak.

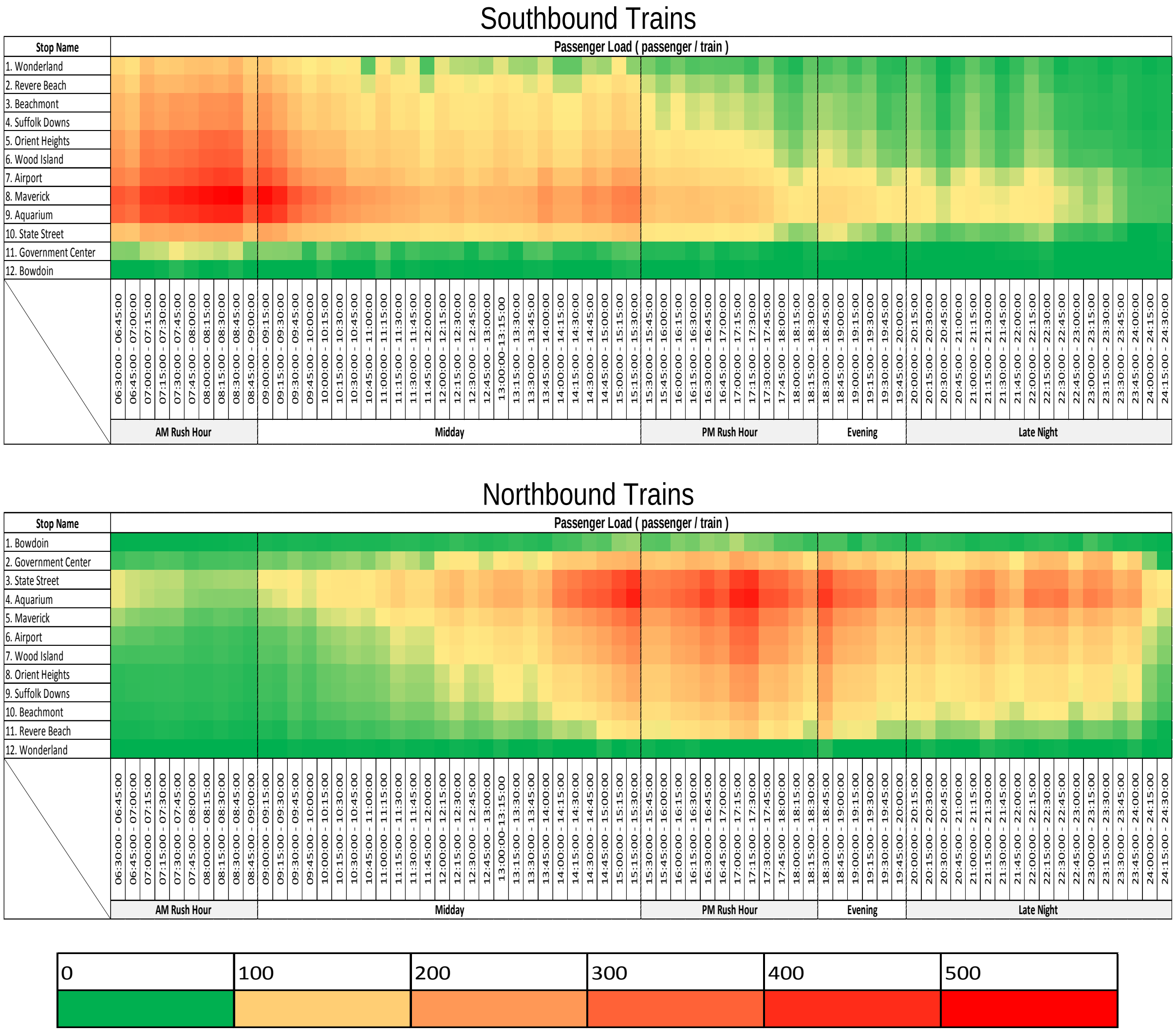


Figure 6.1: Inferred Passenger Occupancy for Blue Line Trains, Winter 2017 (Source: ODX Data from MBTA Research Database)

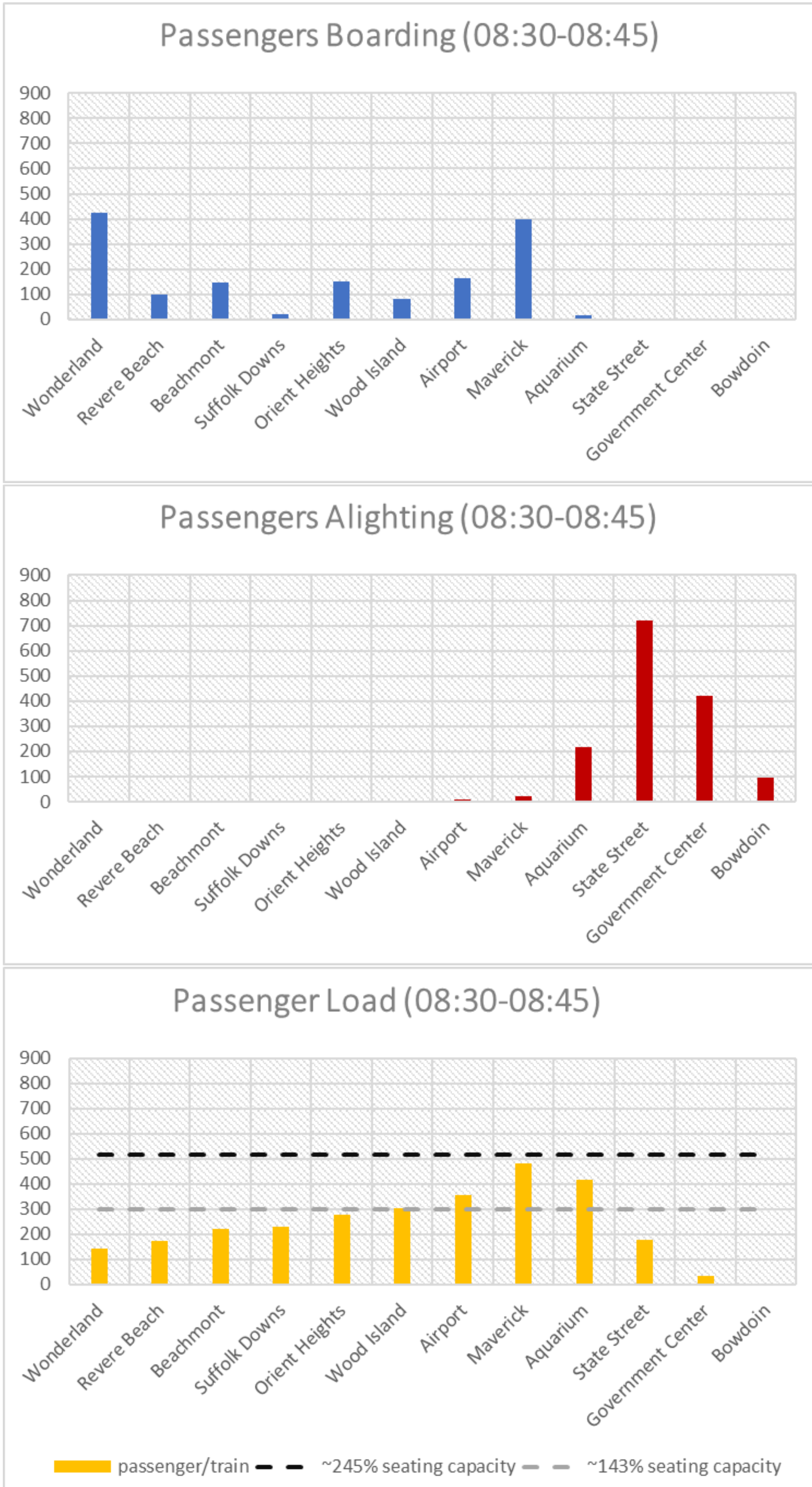


Figure 6.2: Passenger movements for southbound Blue Line, 8:30 – 8:45am (Source: ODX Data from MBTA Research Database)

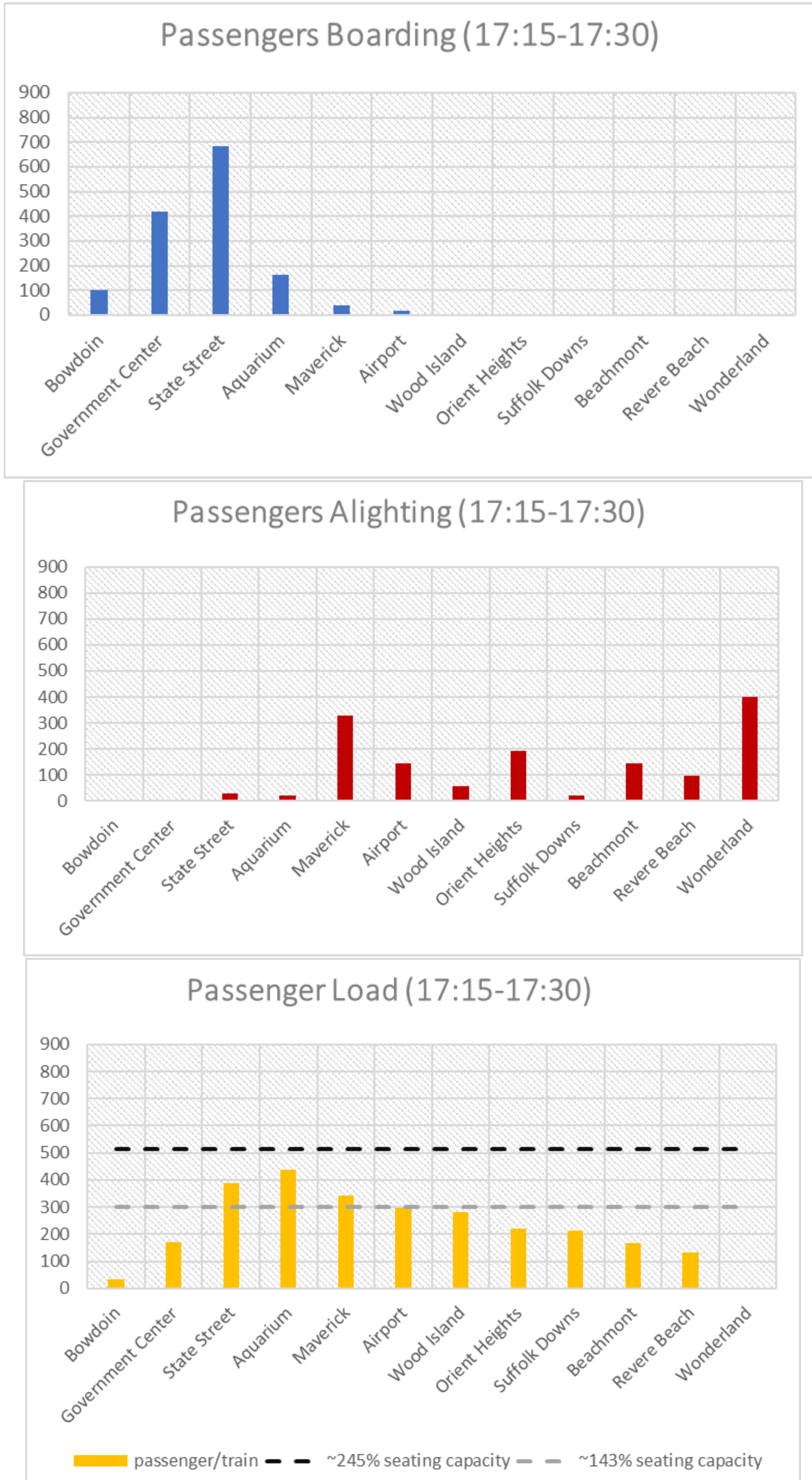


Figure 6.3: Passenger movements for northbound Blue Line, 5:15 – 5:30pm (Source: ODX Data from MBTA Research Database)

6.2 Appendix B: Observations from Manual Counts

Table 6.1: Direct Observations for Sullivan Square (SB): November 15, 2017

Train	Doors Open	Doors Close	Dwell Time (sec)	Passengers Waiting	Passengers Left Behind by Platform Location			
					Front	Middle	Back	Total
1	6:36:33	6:36:53	20	25	0	0	0	0
2	6:42:19	6:42:33	14	62	0	0	0	0
3	6:45:50	6:46:07	17	23	0	0	0	0
4	6:53:28	6:53:44	16	70	0	0	0	0
5	6:57:11	6:57:29	18	40	0	0	0	0
6	7:05:01	7:05:23	22	109	0	0	0	0
7	7:11:22	7:11:50	28	36	0	0	0	0
8	7:18:20	7:18:45	25	91	0	0	0	0
9	7:25:09	7:25:36	27	101	0	0	1	1
10	7:29:31	7:29:45	14	23	0	0	0	0
11	7:35:37	7:36:03	26	51	1	1	0	2
12	7:40:41	7:41:04	23	124	0	0	0	0
13	7:45:57	7:47:16	79	61	0	0	0	0
14	7:55:58	7:56:49	51	313	36	28	30	94
15	8:04:02	8:04:45	43	190	39	33	12	84
16	8:09:16	8:10:20	64	124	0	2	0	2
17	8:12:36	8:12:50	14	27	0	0	0	0
18	8:18:00	8:18:17	17	43	0	0	0	0
19	8:21:43	8:22:18	35	109	0	0	0	0
20	8:28:54	8:29:38	44	113	1	4	11	16
21	8:34:09	8:34:39	30	97	0	0	0	0
22	8:42:48	8:43:38	50	295	46	42	28	116
23	8:49:39	8:50:11	32	94	11	13	12	36
24	8:55:12	8:55:38	26	123	0	0	0	0
25	9:00:15	9:00:31	16	49	0	0	0	0
26	9:04:13	9:04:30	17	56	0	0	0	0
27	9:10:59	9:11:38	39	73	0	0	0	0
28	9:18:06	9:18:18	12	37	0	0	0	0
29	9:24:42	9:26:58	136	122	0	0	0	0

Table 6.2: Direct Observations for North Station (NB): November 15, 2017

Train	Doors Open	Doors Close	Dwell Time (sec)	Passengers Waiting	Passengers Left Behind by Platform Location			
					Front	Middle	Back	Total
1	15:35:04	15:35:41	37	36	4	0	0	4
2	15:44:11	15:44:39	28	58	0	0	0	0
3	15:52:39	15:53:05	26	31	0	0	0	0
4	15:55:38	15:55:54	17	9	0	0	0	0
5	15:58:56	15:59:20	24	4	0	0	0	0
6	16:02:21	16:02:36	15	20	0	0	0	0
7	16:05:42	16:05:56	14	11	0	0	0	0
8	16:11:27	16:11:53	26	57	0	0	0	0
9	16:20:35	16:21:11	36	60	2	0	0	2
10	16:26:42	16:27:10	28	26	0	0	0	0
11	16:34:46	16:35:30	44	77	3	0	0	3
12	16:40:21	16:40:50	29	34	0	0	0	0
13	16:44:47	16:45:30	43	27	0	0	0	0
14	16:49:48	16:50:15	27	41	0	0	0	0
15	16:57:50	16:58:39	49	71	0	0	0	0
16	17:06:57	17:08:00	63	78	12	4	1	17
17	17:11:29	17:12:01	32	101	3	0	0	3
18	17:26:33	17:27:38	65	250	38	23	25	86
19	17:30:08	17:31:00	52	84	23	12	14	49
20	17:33:04	17:33:39	35	10	0	0	0	0
21	17:40:28	17:41:28	60	92	12	4	1	17
22	17:44:06	17:44:40	34	26	0	0	0	0
23	17:46:29	17:46:45	16	9	0	0	0	0
24	17:51:47	17:52:17	30	48	0	0	0	0
25	17:56:43	17:57:09	26	35	0	0	0	0
26	18:06:19	18:07:04	45	89	7	7	1	15
27	18:11:16	18:11:41	25	25	0	0	0	0
28	18:19:37	18:20:22	45	70	1	1	0	2
29	18:23:24	18:23:41	17	24	0	0	0	0

Table 6.3: Direct Observations for Sullivan Square (SB): January 31, 2018

Train	Doors Open	Doors Close	Dwell Time (sec)	Passengers Waiting	Passengers Left Behind by Platform Location			
					Front	Middle	Back	Total
1	6:47:09	6:47:37	28	97	0	0	0	0
2	6:54:08	6:54:24	16	77	0	0	0	0
3	7:00:05	7:00:31	26	60	0	0	0	0
4	7:04:03	7:04:15	12	31	0	0	0	0
5	7:12:46	7:13:07	21	110	0	0	0	0
6	7:19:32	7:20:01	29	145	0	0	0	0
7	7:26:25	7:26:39	14	44	0	0	0	0
8	7:30:20	7:30:35	15	32	0	0	0	0
9	7:36:01	7:36:19	18	106	0	0	0	0
10	7:41:31	7:41:51	20	108	0	1	0	1
11	7:48:12	7:48:45	33	219	0	3	4	7
12	7:54:03	7:54:48	45	107	0	0	0	0
13	7:58:23	7:58:45	22	75	0	0	0	0
14	8:02:29	8:02:51	22	120	0	0	0	0
15	8:07:03	8:07:22	19	90	0	0	0	0
16	8:14:52	8:16:36	104	185	22	22	21	65
17	8:21:30	8:22:09	39	218	0	2	12	14
18	8:29:38	8:30:21	43	144	24	23	29	76
19	8:34:10	8:36:25	135	255	3	3	5	11
20	8:41:19	8:42:41	82	162	1	0	1	2
21	8:48:30	8:49:58	88	101	7	3	0	10
22	8:56:49	8:57:36	47	162	0	7	5	12
23	9:02:51	9:03:07	16	73	0	0	0	0
24	9:07:08	9:07:25	17	26	0	0	0	0
25	9:12:57	9:13:20	23	153	0	0	0	0
26	9:19:32	9:19:47	15	59	0	0	0	0
27	9:25:21	9:25:37	16	105	0	0	0	0

Table 6.4: Direct Observations for North Station (NB): January 31, 2018

Train	Doors Open	Doors Close	Dwell Time (sec)	Passengers Waiting	Passengers Left Behind by Platform Location			
					Front	Middle	Back	Total
1	15:32:57	15:33:35	38	31	2	0	0	2
2	15:39:07	15:39:32	25	42	0	0	0	0
3	15:43:47	15:44:01	14	38	0	0	0	0
4	15:54:50	15:55:13	23	60	0	0	0	0
5	15:57:24	15:57:37	13	25	0	0	0	0
6	16:00:04	16:00:25	21	28	0	0	0	0
7	16:03:08	16:03:28	20	29	0	0	0	0
8	16:09:12	16:09:39	27	63	0	0	0	0
9	16:16:09	16:16:42	33	67	0	0	0	0
10	16:20:41	16:21:01	20	27	0	0	0	0
11	16:31:58	16:32:57	59	151	18	1	4	23
12	16:35:56	16:36:23	27	56	0	0	0	0
13	16:41:28	16:42:01	33	103	0	0	0	0
14	16:45:32	16:45:54	22	55	0	0	0	0
15	16:55:57	16:56:54	57	167	21	3	0	24
16	17:02:42	17:03:29	47	86	3	2	0	5
17	17:12:02	17:13:02	60	130	9	10	0	19
18	17:20:13	17:21:12	59	132	3	6	5	14
19	17:26:02	17:26:48	46	103	5	4	0	9
20	17:30:11	17:30:35	24	41	1	0	0	1
21	17:36:48	17:37:28	40	115	4	0	0	4
22	17:43:13	17:43:44	31	99	3	0	0	3
23	17:46:06	17:46:28	22	35	0	0	0	0
24	17:49:34	17:49:56	22	60	0	0	0	0
25	17:56:51	17:57:17	26	87	0	0	0	0
26	18:01:51	18:02:27	36	75	0	2	0	2
27	18:10:23	18:11:01	38	105	6	1	0	7
28	18:14:18	18:14:36	18	55	0	2	0	2
29	18:19:49	18:20:11	22	55	0	2	0	2
30	18:28:46	18:29:21	35	113	3	0	0	3