Automatic Data for Applied Railway Management: Passenger Demand, Service Quality Measurement, and Tactical Planning on the London Overground Network

by

Michael S Frumin

B.S., Computer Science, Stanford University (2000)

Submitted to the Department of Civil and Environmental Engineering
and the Operations Research Center
in partial fulfillment of the requirements for the degrees of

Master of Science in Transportation

and

Master of Science in Operations Research

at the

Massachusetts Institute of Technology

June 2010

Author . . . .
Department of Civil and Environmental Engineering
Operations Research Center
May 18, 2010

Certified by . . . .
Nigel H.M. Wilson
Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by . . . .
Jinhua Zhao
Research Scientist of Civil and Environmental Engineering
Thesis Supervisor

Accepted by . . . .
Dimitris J. Bertsimas
Boeing Professor of Operations Research
Codirector, Operations Research Center

Accepted by . . . .
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Automatic Data for Applied Railway Management: Passenger Demand, Service Quality Measurement, and Tactical Planning on the London Overground Network

by

Michael S Frumin

Submitted to the Department of Civil and Environmental Engineering
and the Operations Research Center
on May 18, 2010, in partial fulfillment of the requirements for the degrees of

Master of Science in Transportation
and
Master of Science in Operations Research

Abstract

The broad goal of this thesis is to demonstrate the potential positive impacts of applying automatic data to the management and tactical planning of a modern urban railway. Tactical planning is taken here to mean the set of transport-specific analyses and decisions required to manage and improve a railway with time horizons measured in weeks, months, or up to a year and little or no capital investment requirements.

This thesis develops and tests methods to (i) estimate on-train loads from automatic measurements of train payload weight, (ii) estimate origin-destination matrices by combining multiple types of automatic data, (iii) study passenger incidence (station arrival) behavior relative to the published timetable, (iv) characterize service quality in terms of the difference between automatically measured passenger journey times and journey times implied by the published timetable. It does so using (i) disaggregate journey records from an entry- and exit-controlled automatic fare collection system, (ii) payload weight measurements from "loadweigh" sensors in train suspension systems, and (iii) aggregate passenger volumes from electronic station gatelines. The methods developed to analyze passenger incidence behavior and service quality using these data sources include new methodologies that facilitate such analysis under a wide variety of service conditions and passenger behaviors.

The above methods and data are used to characterize passenger demand and service quality on the rapidly growing, largely circumferential London Overground network in London, England. A case study documents how a tactical planning intervention on the Overground network was influenced by the application of these methods, and evaluates the outcomes of this intervention. The proposed analytical methods are judged to be successful in that they estimate the desired quantities with sufficient accuracy and are found to make a positive contribution to the Overground's tactical planning process.

It is concluded that relative measures of service quality such as the one developed here can be used in cross-sectional analysis to inform tactical planning activity. However, such measures are of less utility for longitudinal evaluation of tactical planning interventions when the basis against which service quality is judged (in this case the timetable) is changed. Under such circumstances, absolute measures, such as total observed passenger journey times, should be used as well.

Thesis Supervisor: Nigel H.M. Wilson
Title: Professor of Civil and Environmental Engineering

Thesis Supervisor: Jinhua Zhao
Title: Research Scientist of Civil and Environmental Engineering

Acknowledgments

The research presented in this thesis, indeed the entire course of graduate study which culminates in this thesis, was made possible by a number of parties whom the author wishes to thank and acknowledge.

- The Massachusetts Institute of Technology, its Department of Civil and Environmental Engineering, and the faculty and staff of the Master of Science in Transportation program for making a home for students interested in the science and practice of transportation.
- Professor Nigel Wilson and the other leaders of the Transit Research Group for the opportunity to conduct research into practical problems facing the public transport providers who make our cities possible.
- Transport for London, and its many staff members in the departments of London Rail, London Underground, and Fares & Ticketing, for financial support, feedback, knowledge, and two summers of internship employment.
- Certain railway managers in and around London and New York City, including Oliver Bratton and Herbert Lambert, for helping ground this thesis in the contemporary needs and practices of the managers of modern urban railways.
- Dr. Jinhua Zhao, along with Professor Wilson, for supervising the writing of this thesis.
- The MIT Operations Research Center for allowing the author to dig deeper into many of the technical and analytical methods applied in the field of transportation and in this thesis.
- The family, friends, colleagues, and classmates without whom it is difficult to imagine how the author would be anywhere near where he is today.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis develops and applies methods for using automatic data to characterize urban railway passenger demand and service quality, primarily for the purposes of supporting railway managers in the process of tactical planning. In the context of this thesis, *tactical planning*[1] is taken to mean the set of transport-specific analyses and decisions required to manage and improve a railway with time horizons measured in weeks, months, or up to a year and little or no capital investment requirements.

This thesis proceeds in three principle endeavors. First, it develops and applies methods to characterize passenger demand spatially, temporally, and behaviorally. Second, it develops and applies a method to characterize service quality in terms of the difference between actual passenger journey times and journey times implied by the published timetable. Finally, it documents how a tactical planning intervention was influenced by the application of these methods, and evaluates the outcomes of this intervention. The broad goal of this thesis is to demonstrate the potential positive impacts of applying automatic data to the management and tactical planning of a modern urban railway.

The automatic data central to this thesis are (i) disaggregate passenger journey transactions from an entry- and exit-controlled automatic fare collection (AFC) system, (ii) payload weight measurements from sensors in train suspension systems, and (iii) aggregate passenger volumes from electronic station gatelines. In some cases, a single type of data is sufficient to derive the desired results. More often than not, it is necessary to combine multiple types of automatic data with each other and with reference data such as network representations and public timetables.

The research presented in this thesis has been developed commensurate with the analytical needs of the managers of the rapidly growing, largely circumferential London Overground network in London, England. The methods developed in this thesis are each applied to data from this railway, both as a test of the method and for the sake of analyzing the railway and its passengers *per se*. This thesis includes a case study which demonstrates the use of these methods to support and evaluate a tactical planning effort on the core portion of the Overground network. While the work in this thesis is motivated primarily by problems facing the managers and passengers of the Overground, the methods it develops should generalize well to other contexts where similar automatic data are available.

[1] What is referred to here as "tactical planning" is also commonly referred to as "service planning."

Section 1.1 of this chapter describes selected aspects of railway management, including tactical planning, and the relevance of automatic data thereto. Section 1.2 describes the motivations for this thesis, and Section 1.3 describes its specific objectives. Section 1.4 gives an overview of the approach for the research described in this thesis. Section 1.5 describes the organization of the balance of this document.

1.1 Railway Management, Tactical Planning, and Automatic Data

Railways can be complex and multi-faceted organizations, with certain management functions similar to any large enterprise. The aspect of railway management with which this thesis is primarily concerned is tactical planning, as defined in the introduction to this chapter. The decision processes that make up tactical planning include the following, as per Wilson (2008), and described in detail by Ceder (2007) and Vuchic (2005). In this type of planning, demand levels are often considered to be constant.

- *Network Design* – route design over a fixed infrastructure network.
- *Frequency Setting* – determination of service frequencies for each route, for different service periods (*e.g.* by time of day and day of week).
- *Timetable Development* – creation of a specific timetable, including running times, to provide a certain frequency of service on a set of routes.
- *Vehicle Scheduling* – scheduling of vehicles to cover all trips in the timetable.
- *Crew Scheduling* – scheduling of crews to staff all vehicles.

These decisions are often but not always described and implemented hierarchically, as shown in Figure 1-1.

At the top of the hierarchy, starting with network design, decisions tend to be infrequent, dominated by service considerations, and driven by judgment and manual analysis. Moving down the hierarchy, towards crew scheduling, decisions are more frequent, dominated by cost concerns, and can be computer driven by automatic tools such as optimization-based scheduling software. Tactical planning thus encompasses numerous types of analyses and decisions, only a few of which are treated explicitly in this work, more of which can benefit from the methods developed here, and many of which could benefit from the use of automatic data in general.

The decisions made on a day-to-day real-time basis to implement a given tactical plan can be described as *service control*. Studied by Carrel (2009) for high frequency urban railway services, these decisions include the assignment of physical vehicles to vehicle schedules, the assignment of drivers to vehicles, and various types of modifications to these schedules and to the timetable to account for disruptions and delays. Froloff et al. (1989) study and describe this topic for urban bus, rather than railway, services.

Decisions made with much longer time horizons or demanding substantial capital investment can be referred to as *strategic planning*. This type of planning is more akin to overall

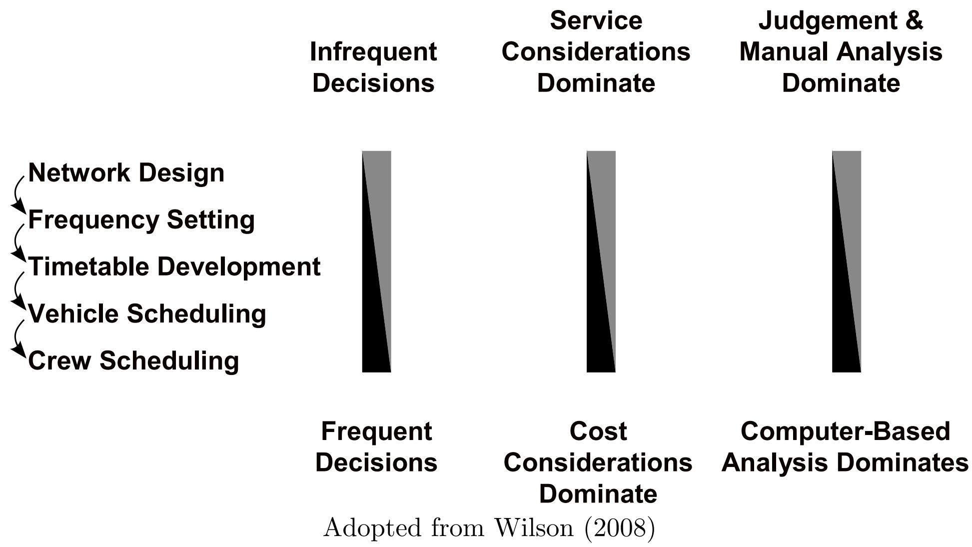


Figure 1-1: Conceptual diagram of tactical planning hierarchy.

transport planning, as described by Meyer and Miller (2001). It typically considers substantial changes to land use, economic development, and overall levels of demand as well as shifts of demand between different transport modes (*e.g.* public vs. private transport).

Information Technologies and Automatic Data

At a major urban railway by which the author has been employed, staff from the lowest to the highest levels refer to the operation of the railway as "making service." This choice of words attests to the fundamental nature of railway operation as an industrial production process. Like other service and manufacturing production processes, railways are heavily dependent on technology. For example, the capacity of a railway is largely determined by the design of its propulsion and signaling systems (Kittelson & Associates, Inc et al., 2003*b*).

In railways, as in other technology-driven industries, information technology (IT) has played an increasingly larger role over time. Traditionally, railways have invested heavily in IT systems for core production and real-time management functions such as signaling, dispatching, vehicle operation, and service control (Vuchic, 2007), as well as key auxiliary functions such as fare collection (Multisystems, Inc et al., 2003) and passenger information systems (Multisystems, 2003). By their very nature, these IT systems serve to automatically create, transmit, manipulate, display, and store electronic data. This thesis is *not* about the impact of these kinds of systems *per se* on railway operations. Rather, this thesis is about how the data produced and stored by these various IT systems can and should be used in analytical management functions, specifically tactical planning. It is hoped that the methods developed in this thesis find application in other contexts as well.

1.2 Motivation

Most broadly, this thesis is motivated by the hypothesis that the managers of railways and other public transport services have much to gain from the use of automatic data sources for tactical planning. These data sources represent a significant and often unused analytical asset that can be highly leveraged to help improve service and decrease costs in many different contexts.

On a more narrow level, this work is motivated by a need in the railway industry for methods to make more effective analytical use of available automatic data to understand passenger demand and service quality. In that sense, it is motivated by needs of managers of railways in London and around the world facing similar tactical planning challenges as the managers of the London Overground. One such need is to better understand and measure passenger behavior and service quality, and to understand how specific tactical planning interventions affect certain aspects of the passenger experience.

On the most narrow level, the work in this thesis is motivated by the immediate and ongoing analytical needs of the mangers of the Overground. Foremost among their needs are the ability to measure passenger loads on individual scheduled services, to estimate total passenger boardings at individual stations, and to understand passenger travel patterns at the level of origin-destination (OD) flows. Currently, these needs are met primarily through manual data collection efforts, and so are impossible to satisfy cost-effectively at desired frequencies or sample sizes. An additional contemporaneous need of the Overground's managers is to understand the overall impact on passengers of the recent tactical planning exercise on the core portion of the Overground network.

1.3 Objectives

The most broad-based objective of this thesis is to investigate the use of automatic data in the tactical planning of urban railways. This should be accomplished by demonstrating that the use of automatic data through application of the methods developed here in fact contributed to measurable improvement in the service quality of the London Overground.

A more specific objective is to provide to the Overground's managers insights into passenger demand and service quality that are relevant to tactical planning but are inaccessible without automatic data. In other words, to develop methods that take advantage of the richness of available automatic data to achieve the following previously unavailable analytical results.

- Understand the relationship between the published timetable and passenger arrival behavior at stations.

- Measure railway service quality as experienced by passengers in terms of the difference between actual journey times and journey times implied by the published timetable.

- Document the recent tactical planning exercise on the core portion of the Overground network, including the application of the above results to this process.

- Evaluate the outcomes of the resulting tactical planning intervention. This evaluation should be primarily from the perspective of passengers, but should also consider the experience and process of operators and managers.

An additional objective of this thesis is to develop a prototype of a data collection scheme with associated methods to meet the immediate and ongoing analytical and tactical planning needs of the mangers of the Overground. This scheme should maximize the use of automatic data and minimize cost, primarily through minimizing the need for manually collected data. It should improve the quality, breadth, and timeliness of its outputs, with the goal of supporting continued improvement in Overground services through improved tactical planning. The specific objectives are as follows.

- Understand the error associated with train payload weighing systems, and calibrate them for use in directly measuring passenger loads on trains.

- Develop and test a method to estimate station-to-station origin destination flows on the Overground at the time-period level, and use this result to estimate total boardings at the station level.

1.4 Research Approach

The overall approach to this research is pragmatic and applied. Since the goal is to address certain issues faced by the managers of urban railways, this thesis draws on models and techniques from existing literature whenever possible and proposes new methods only when necessary. Mathematical models and statistical techniques, when necessary, are selected to be as simple as possible, but no simpler.

To achieve the stated objectives, the research in this thesis will do the following.

- Calibrate loadweigh systems using a linear regression of a sample of loadweigh measurements from the London Overground against paired manual on-board passenger counts. Use the results of this regression to understand the error associated with loadweigh data.

- Estimate station-to-station passenger demand for the Overground in the AM Peak period from multiple sources of data using network and mathematical models. The data sources used are aggregate Oyster journey data, automatic gateline entry counts, and a complete set of manual on-board passenger counts (standing in for loadweigh data). Test the sensitivity of this estimate to randomness in the on-board counts when estimated from loadweigh data.

- Assign individual Overground passenger journeys with specific scheduled services in the published timetable. This assignment depends on network models and algorithms applied to each Oyster journey and its associated origin, destination, and time of entry time into the system.

- Using that assignment, analyze passenger arrival behavior at stations by comparing the entry times of passenger journeys with respective scheduled departure times.

- Also using that assignment, measure service quality in terms of the difference between actual passenger journey times estimated from disaggregate Oyster data and scheduled passenger journey times from the published timetable. Aggregate these values along a variety of dimensions to assess service quality on the Overground.

- Document the recent tactical planning exercise on the North London Line of the Overground network, including its use of these and other results.
- Evaluate the outcomes of implementing the new tactical plan resulting from this exercise, primarily in terms of its effects on passengers.
- Use this case study as means to assess the applicability of the methods developed in this thesis for improved tactical planning using automatic data.

1.5 Thesis Organization

This thesis is organized into 10 chapters, including this one. Chapter 2 provides background information on the London Overground and the broader network of public transport services in London, England. Chapter 3 describes the various data sources available to the Overground, automatic or otherwise, how they are currently used by the Overground, and how they could be used. Chapter 4 develops and applies a method for calibrating train payload weighing systems to measure passenger loads. Chapter 5 develops and applies a method, tailored to the circumstances of the Overground, to estimate time period level origin-destination matrices from multiple data sources. Chapter 6 develops and applies a method to analyze passenger arrival behavior at stations. Chapter 7 develops and justifies a unified method to use actual passenger journey times to measure service quality relative to published timetables. Chapter 8 applies that method to the Overground, validates it, and compares it to existing performance measurements. Chapter 9 presents a case study on the use of these methods to inform and evaluate a recent tactical planning exercise on the Overground. Chapter 10 offers some final remarks, including conclusions, recommendations, and areas for future research.

This thesis covers a range of topics but is intended, to a certain degree, to be consumable in a piecemeal fashion. To that end, it groups literature review, methodologies, and results topically into single chapters or consecutive sets of chapters. This applies most directly to Chapter 4, Chapter 5, Chapter 6, and the combination of Chapters 7 through 9.

Chapter 2

Public Transport in London

This chapter provides a broad background of the city of London and its public transport network, including the London Overground. Section 2.1 gives a brief introduction to the city of London. Section 2.2 describes some of the key elements of London's public transport network. Section 2.3 describes the fare structure and ticketing systems of that network. Section 2.4 describes the Overground network from a transport perspective. Section 2.5 describes the institutional structure of the Overground. Finally, Section 2.6 describes some of the key elements of the Overground's investment program.

2.1 London and the Greater London Authority

London, a city of approximately 7.5 million inhabitants covering 1,572 square kilometers, is located in the southeast of the United Kingdom, of which it is the capital (Government Offices, 2010). The Greater London Authority (GLA), created by a 1999 act of the British parliament, governs London at a regional and strategic level. The primary executive of the GLA is the popularly elected Mayor of London (Greater London Authority, 2010*a*). The Mayor has wide powers over the city's transportation agency, Transport for London (TfL), including setting its strategy and budget and appointing its board. TfL manages most facets of the transport system in London, including roads, the congestion charge, and local public transport (Greater London Authority, 2010*b*). It has an ambitious investment program of over £35 billion from 2009 through 2018 (Transport for London, 2009*c*).

2.2 The Public Transport Network

London has a world class public transport system, serving an estimated 12 million passenger "journey stages" on an average in 2007, representing a growth of almost 60% since 1991 (Transport for London, 2009*f*). Table 2-1 shows the size (in stops or stations) and annual ridership of the largest components of this system, described in further detail in the following paragraphs.

London's more than 8,000 local buses ply a network serving almost 19,000 stops on over 700 routes, carrying an estimated 2 billion yearly passengers (Transport for London, 2009*d*). The London Underground (LU), a world-famous metro system with routes going back 150

Network	Number of Stops or Stations	Approximate Annual Ridership (millions)
London Buses	19,000	2,000
London Underground (LU)	402	1,000
Docklands Light Railway (DLR)	40	70
London Overground (LO)	56	33
National Rail (NR) in London	318	883

Table 2-1: Size and patronage of the public transport networks in London

years, serves an estimated billion passengers annually on a 402 kilometer network of 270 stations on 11 lines (Transport for London, 2010*c*). The Docklands Light Railway (DLR), which opened in 1987, serves an estimated 70 million yearly passengers on a 40-station network in parts of East London and the newer Canary Wharf financial district (Transport for London, 2009*b*). The London Overground, an above-ground urban railway re-christened in 2007, serves an estimated 33.4 million passengers annually over a largely circumferential 107 kilometer network of 56 stations on four lines (Smales, 2009). It is the focus of the research in this thesis and will be described in further detail in the following sections.

The United Kingdom's system of regional and inter-city railways, referred to as National Rail (NR), serves London's commuters and visitors at 318 stations within Greater London (Office of Rail Regulation, 2009) and connects the capital with the rest of the country. The National Rail network serves an estimated 833 million passengers annually in London and the southeast of England, and an estimated 1.2 billion passengers across the entire country (Office of Rail Regulation, 2008). National Rail services are operated by twenty-nine regional Train Operating Companies (TOCs) across the entire country, most of which serve London either in a commuter or long distance capacity (National Rail, 2010). Each TOC operates according to a competitively bid franchise agreement, let by the national Department for Transport (DfT). This is of particular relevance to this thesis because of the Overground's unique relationship, described in the following sections, to both TfL and the National Rail system.

Broadly speaking, London's public transport is very well integrated. Interchanges between the different rail services are available at more or less all possible opportunities (Transport for London, 2010*e*), and most if not all bus routes connect to the rail network at one or more points.

TfL is responsible for all the services described in this section other than those operated by National Rail TOCs. Only the London Underground is operated directly by TfL employees. All other bus and railway services are provided by competitively bid operational concessions let by TfL. In all of these concessions, TfL holds all of the revenue risk in some concessions and almost all of the revenue risk in others. TfL itself is organized into modal units and a central corporate finance and planning group. The largest modal units are eponymous London Underground, Surface, managing roads and Buses, and London Rail, which manages TfL's other rail services (including the Overground and the DLR) as well as acting as TfL's liaison to the National Rail network and the TOCs. Figure 2-1 provides a high level organization chart illustrating these relationships for providing public transport service (this chart does not describe the ownership and responsibility for infrastructure).

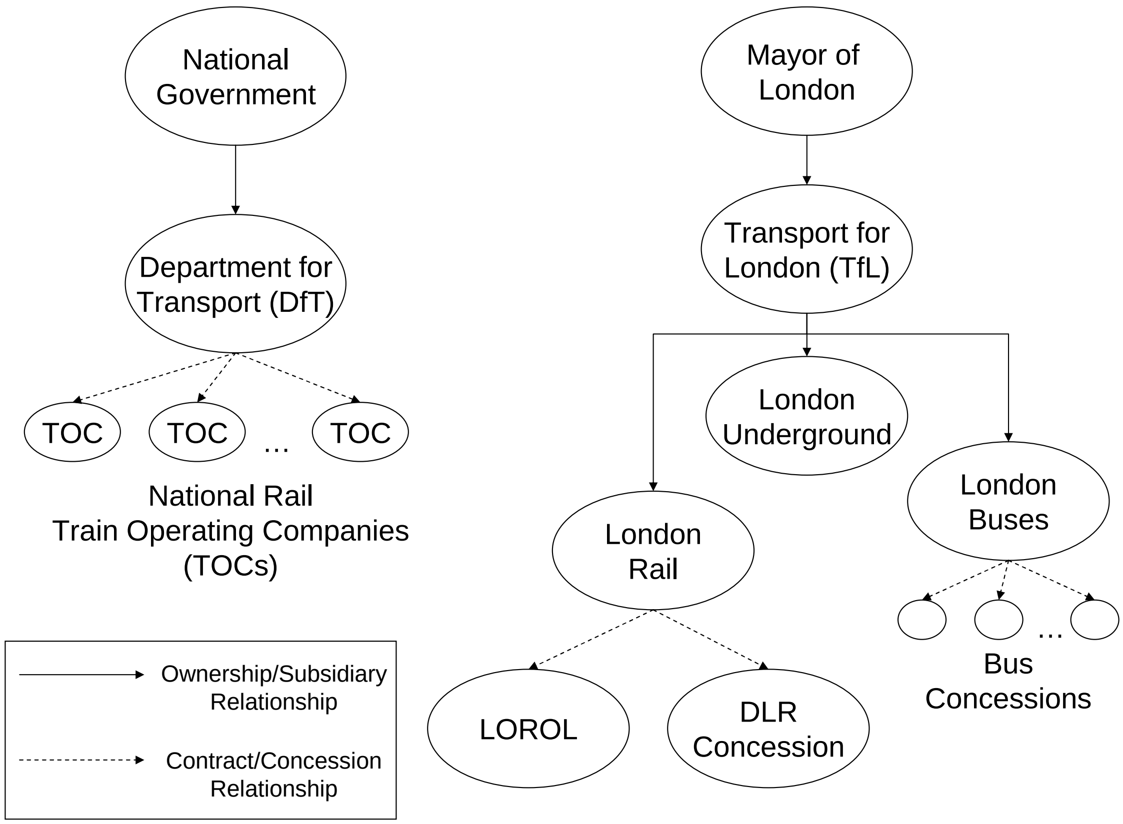


Figure 2-1: High level institutional structure for public transport service provision in London

2.3 Public Transport Fares and Ticketing

The fare structure for rail services in London is integrated, with the primary exception that neither transfers between buses nor between bus and rail are free for passengers lacking unlimited-use passes. Rail travel in London, be it by TfL or National Rail services, is generally priced on a zonal basis, where the zones are oriented concentrically around Central London (see the map in Appendix C). The price of a journey depends not only on the starting and ending zones, but also on which zones that journey has passed through – trips through lower zones (*i.e.* through more central parts of London) cost more. Rail fares also have a temporal component, with a premium charged for travel during peak commuting hours. Interchanges between rail services are free in some (*e.g.* between London Underground lines), but not all cases (*e.g.* between some National Rail and Underground services). Bus travel is priced on a per-boarding basis, with no discount for interchanging between buses or between bus and rail (Transport for London, 2010*b*).

Unlimited-use passes, called *travelcards* or *season tickets*, are also available for different zonal combinations (*e.g.* zones 1-2, 1-3, 2-4, etc) and different lengths of time (*e.g.* one day, seven days, 30 days, one year). Travelcards cover all public transport travel – all TfL and National Rail services – within the selected zones. The price of a travelcard of course depends on the zonal and temporal coverage, with the potential discount offered to the purchaser increasing with the temporal span of the ticket (Transport for London, 2010*a*). London's public transport fare structure includes a host of other discount schemes, including bus-only passes, concessions for students and the elderly, and point-to-point season tickets on National Rail.

Public transport fares in London are clearly complex, but the application of ticketing technologies has progressively simplified the details that passengers must understand and choices they must make. In particular, the introduction and evolution of the Oyster smart-card ticketing system has made certain aspects of public transport ticketing in London much more efficient for passengers and for operators. A passenger can add monetary value to his or her Oyster card in bulk and then simply "Pay As You Go" (PAYG) by validating the card when entering and exiting the transport network (Transport for London, 2009*e*). The Oyster system deducts the correct amount of money from the passenger's card for each journey, including in such complicated cases as "Out-of-Station Interchanges" (OSIs) when the journey requires exiting and re-entering the transport network to transfer between services at certain nearby but unconnected stations. This saves passengers without travelcards the effort of having to purchase individual tickets for each journey, and saves operators the effort of having to sell them. To incentivize the use of Oyster cards, TfL has imposed a significant price penalty for the purchase of single tickets (Hong, 2002).

Oyster cards also support the purchase and use of travelcards, which TfL no longer offers on traditional magnetic-stripe media. TfL no longer offers single-day travelcards either, instead offering Oyster PAYG users daily "price capping" or "best value." Under this scheme, the Oyster system calculates the price of the single-day travelcard or pass that *would* have been necessary to accommodate all of the user's rail and bus travel on that day, and stops deducting from their Oyster card's balance once that amount has been reached (Transport for London, 2009*e*).

Traditionally, neither PAYG nor daily capping were available for most of the National Rail network. In January, 2010, this changed with the negotiation and implementation of the Oyster eXtension to National Rail (OXNR) project, which extended the Oyster system to almost all National Rail stations within Greater London (Transport for London, 2010*b*). However, the various National Rail TOCs do not generally retail Oyster products (neither cards nor PAYG value nor travelcards) at their stations, so many National Rail passengers in London still use magnetic-stripe tickets, chiefly travelcards, to pay for their journeys (Chan, 2007).

Since its introduction in 2003, the Oyster system has grown to become the dominant fare media for TfL services, processing over 10 million transactions daily. Over 6 million cards are in regular use, and over 80% of all bus and London Underground journeys were made using Oyster in 2009 (Transport for London, 2009*e*). That said, there are circumstances where Oyster has significantly less penetration on the TfL network. This is most often the case at places and times where large volumes of National Rail commuters or visitors interchange to or from TfL services, for example at large intermodal facilities (*e.g.* Victoria station) during peak commuting hours (Chan, 2007). This must be considered when using the Oyster system as a source of data on passenger journeys.

2.4 The London Overground Network

Figure 2-2 shows a map of the London Overground network (with other railways) as of January, 2009. A schematic map, with detailed interchange information is shown in Appendix C. The Overground network is for the most part circumferential, primarily orbiting London to the North and West, with only a single station (Euston) in fare zone 1 and the majority of stations in zones 2 and 3. The Overground is very much part of the overall integrated network of TfL and National Rail services, with 19 of its stations offering interchanges to London Underground or DLR services and 13 of its stations offering interchanges to National Rail (24 stations offer at least one interchange). Key Overground stations, such as Stratford, Clapham Junction, and Euston, are major intermodal terminals or interchange points. In 2010, the Overground is running 407 scheduled weekday trips with 27 units of rolling stock (*i.e.* trains) (Brimbacombe, 2010).

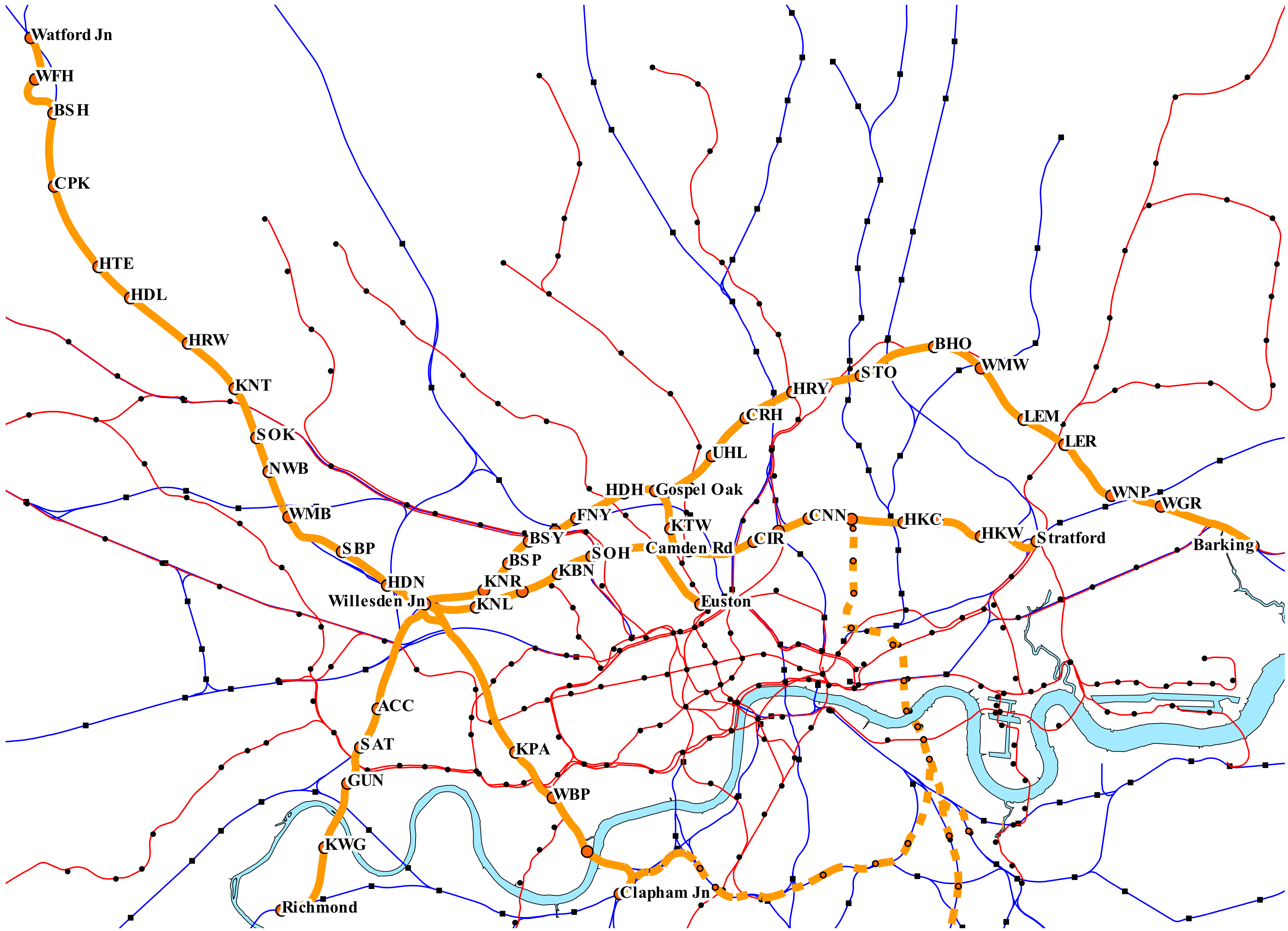
Services on the Overground are for the most part divided into four different lines (*i.e.* service patterns) as described in Table 2-2. The core of this network is the North London Line (NLL) which runs 28 kilometers between Stratford in the northeast of London and Richmond in the southwest, connecting to every other Overground service and numerous other TfL and National Rail services along the way. It is by far the busiest Overground line, with the most frequent service and an estimated 58% of all Overground boardings (Smales, 2010). The NLL runs four (end-to-end) trains per hour (tph) over most of the day, with some segments receiving six tph during the peak periods. Of the twenty-three stations on the NLL, seventeen are served only by the Overground.

The other Overground lines run at lower frequencies of three tph during the peak periods and two to three tph during other periods. In the case of the Gospel Oak to Barking (GOB) line, which is the only service available at nine of its twelve stations, these lower frequencies are the result of relatively low passenger demand. The other two Overground lines run at these low frequencies because of competition from other services. The five-station West London Line (WLL) shares all but one station (Willesden Junction, its northern Terminus) with another National Rail TOC. The much longer nineteen-station Watford DC Line runs interavailably[1] for ten consecutive stations in the middle of its route with the London Underground's Bakerloo line, and National Rail express services run much more quickly from some of its northernmost stations to its southern Central London terminus (Euston). During peak periods, select Overground services run special patterns described in detail in Chapter 9.

Service Pattern (Line)	Code	Primary Terminals	Frequency (Peak tph)
North London Line	NLL	Stratford ⇔ Richmond	4-6
Gospel Oak to Barking Line	GOB	Gospel Oak ⇔ Barking	3
Watford DC Line	WAT	Watford Junction ⇔ London Euston	3
West London Line	WLL	Clapham Junction ⇔ Willesden Junction	2-3

Table 2-2: Primary London Overground service patterns

[1] "Interavailability" describes the situation where two (or more) services are available on the same platform and travel to some (or all) of the same stations down the line.



Overground lines are denoted in orange (dashes indicate lines in planning or under construction), Underground (and DLR) in red, and National Rail in blue

Figure 2-2: Map of the London Overground network, with other rail services

On November 11, 2007, overall management and revenue responsibility for this set of services was transferred to TfL from the Silverlink TOC which held the prior National Rail franchise. At that time, the network was re-branded as the Overground and became fully Oyster-enabled. TfL's goals in absorbing the Overground were to improve service standards (including station safety and staffing) and to deliver the investments and expansions described in Section 2.6 (Transport for London, 2007*a*). TfL contracted the operation of the Overground to a private concessionaire, creating the fairly unique institutional structure described in the following section.

2.5 Institutional Structures

London Rail, also referred to legally as Rail for London (RfL), delivers London Overground services through a concession contract with the private operator London Overground Rail Operations Limited (LOROL). London Rail plays other non-operational roles in the management of the Overground network, including

- planning and specifying service levels, including frequencies and train lengths,
- planning and funding major investments (*i.e.* strategic planning),
- delivering those investments through contracts with infrastructure and rolling stock providers,
- monitoring and forecasting revenues and demand,
- working with LOROL to respond to changing conditions on the network,
- and communicating with the riding public and other stakeholders.

In this sense, tactical planning for the Overground is a shared responsibility between London Rail and LOROL.

LOROL is a joint venture between two world-class private (but largely state-owned) railway operators – MTR Corporation of Hong Kong and Deutsche Bahn AG of Germany (LOROL, 2010). Among LOROL's contracted responsibilities are to

- develop public and working timetables to meet certain service level commitments,
- hire and train station staff, train crews, and service controllers,
- make and manage train and station service on a daily basis,
- conduct light maintenance on stations and vehicles,
- gather information on passenger demand,
- cooperate with infrastructure upgrades and expansions

(Rail for London Limited and MTR Laing Metro Limited, 2007).

Note that LOROL is not responsible for maintaining and operating the infrastructure (*i.e.* tracks, switches, and signals) on which it runs. This is because the Overground operates on infrastructure owned, maintained, and operated by Network Rail, the public benefit corporation responsible for the National Rail network infrastructure. LOROL operates within a well defined institutional, financial, and political framework that structures the relationships between the infrastructure owner (*i.e.* Network Rail), operators (*i.e.* TOCs), and sponsors (*i.e.* DfT and TfL).

An in-depth discussion of this framework is beyond the scope of this thesis. It is sufficient to say that the institutional structure described here places the Overground in a unique context in London. On the one hand it is the only National Rail franchise let by other than DfT and with essentially no revenue stake for the franchisee. On the other hand, it is the only TfL service that is run by a private operator on infrastructure owned neither by TfL nor by the operator. One implication of this arrangement, further discussed in Chapter 8, is that the performance and incentive regime for the Overground is rooted in the regime used across the National Rail network.

2.6 Investment and Expansion

The London Overground network is currently the subject of over £1.5 billion of investment in extensions, infrastructure upgrades, and new rolling stock. As a result, it is set to grow significantly over the next several years in terms of network size, service density, and passenger demand. A longer-term goal of these investments is to support access to Stratford, the site of the 2012 Summer Olympic Games (Transport for London, 2007*b*, 2009*c*). Among the most significant investments in the Overground network are:

East London Line Extension (ELLX) – Opening in stages starting in the spring of 2010, this project will rehabilitate and extend what was the London Underground East London Line and add it to the Overground network. Shown in Figure 2-2, the old East London Line is being extended to the north and to the south, and will enter service with 12 tph on the trunk portion. North of Shoreditch High Street station, new infrastructure will bring rail service for the first time to some parts of inner East London. South of New Cross Gate station, a connection to the National Rail network will improve access from South London to the TfL network and to parts of East London. Some service on this line will eventually terminate at Highbury & Islington and connect directly to the North London Line. This project is projected to serve an estimated 33.2 million passengers in 2011 (AECOM, 2006), effectively doubling Overground ridership.

North London Railway Infrastructure Package (NLRIP) – To be completed in advance of the 2012 Olympics, this project will upgrade track, switches, and signals, primarily on the core portion of the North London Line between Stratford and Willesden Junction. These are set to support frequencies of up to 12 tph on the North London Line.

East London Line Phase 2b – Through a number of small connections and reconfigurations, this project will connect the Overground between Crystal Palace (on the ELL) to Clapham Junction (on the WLL) through South London. Set to open in 2012, it will complete the Overground's orbital route structure.

New Rolling Stock – By 2012, the entire planned Overground fleet of 216 rail cars, in three-car and four-car trains, will be new Class 378 vehicles from Bombardier Transportation. This fleet has wider doors and longitudinal seating to improve the capacity and performance under heavy demand, and passenger amenities such as air conditioning and improved passenger information systems. An important feature of this fleet in the context of this thesis is that its computerized braking systems will measure and report train payload weights for the sake of estimating passenger loads.

This Overground investment program is just one part of the much larger and very ambitious TfL investment program, which includes major upgrades to the London Underground and DLR as well as the new £16 billion Crossrail project (Transport for London, 2009*c*).

Chapter 3

Data Needs and Sources

This chapter proposes a new approach to data collection and demand estimation for the London Overground network. The new approach is designed to minimize cost and maximize the use of automatic data sources while satisfying the various reporting and analytical requirements of the network's public owner and private operator. It is designed to improve the quality, breadth, and timeliness of its outputs, with the goal of supporting continued improvement in Overground services through improved tactical planning. Broadly speaking, the goal of such a strategy is to change data collection from an infrequent, expensive, manual process to a frequent, cost-effective, automatic one.

Section 3.1 describes the immediate data needs of the managers and planners of the Overground and how those needs are currently met. Section 3.2 outlines an approach to meet these specific needs using a range of automatic data sources. Section 3.3 describes in detail each data source, with respective limitations, that will be used to support the proposed approach and other methods and analyses developed later in this thesis.

3.1 Data Needs

While TfL's and LOROL's interests are generally aligned, they have different analytical data needs as a consequence of their different responsibilities. The primary quantities of interest to TfL and/or LOROL analysts and tactical planners are the following (Smales, 2009; Bratton, 2009).

- Average loads on individual scheduled services (*i.e.* timetabled trips) between all consecutive stations.
- Average total boardings, by station, at the time period (*e.g.* AM Peak) level.
- Average origin-destination flows between all pairs of Overground stations at the time period level.

TfL's concession contract with the Overground's private operator requires twice-yearly passenger counts (Rail for London Limited and MTR Laing Metro Limited, 2007, Schedule 1.5). To date, these counts have been designed to directly measure boardings, alightings, and loads at each station *once* (per counting period) for each scheduled service. They indicate utilization rates and capacity issues on Overground services, of interest to both TfL and LOROL. However, because of the small sample size, these counts are likely not statistically sound.

Passenger counts do not directly indicate demand at the individual origin-destination (OD) flow level, nor do they describe how passengers use the Overground in conjunction with other modes (*e.g.* the London Underground). Such information is primarily of interest to TfL, and can be derived from OD flow matrices. Currently, the only OD matrix available to Overground analysts and tactical planners comes from the RailPlan strategic model, which in turns depends on the results of LTDS, London's regional transport model, which uses household surveys and census data as inputs . It is available only for the AM Peak, is out of date and is based on very small sample surveys (Warner, 2010). As a result, it may not accurately reproduce current demand patterns at the OD flow or station (*i.e.* boarding or alighting) level.

Broadly speaking, Overground management (Bratton, 2009; Smales, 2009) have expressed the need for a strategy that provides them with the same analytical quantities with which they are used to working, but does so in a cheaper and more timely manner than is currently available. The next section proposes a means to accomplish this, primarily through the adoption and combination of automatic data sources.

3.2 A Synthetic Approach for the London Overground

In general, data from automatic systems are cheap and plentiful but do not capture complete information from every passenger. Manual counts and surveys can fill some of the gaps left by automatic systems, but are expensive to gather and therefore often insufficiently sampled. Any source, taken alone, has some inherent ambiguity. This chapter, along with Chapters 4 and 5, seek to illuminate the limitations of each source and design the most effective strategy to use them all together. This following section proposes such a strategy that uses automatic data to meet the London Overground's analytical requirements.

Several sources of automatically collected data on passenger quantities and behavior, each with its own limitations, are readily available but as of yet largely unused in analysis and management of the Overground. Electronic transactions from the Oyster smartcard ticketing system describe individual journeys on the TfL rail network, but only a some fraction of all journeys. Ticket gatelines, where present, automatically provide aggregate counts of station entries and exits, but do not distinguish between passengers of different services at a given station. With the delivery of new rolling stock, each Overground train will automatically weigh and electronically report its passenger payload over the length of each trip. It is hoped that these "loadweigh" systems on new rolling stock can provide a cheaper and more statistically sound alternative to the manual on-board counts, but they require calibration and will not indicate station boardings and alightings nor passenger origins and destinations.

None of these electronic data sources, taken alone, tells the complete story of how passengers use the Overground. Used in conjunction, what they may lack in completeness they may make up for in quantity, variety, and cost. One objective of this thesis is to develop ways to *combine* the various automatic data sources, and to target manual data collection resources for maximum cost-effectiveness, to meet the needs of Overground managers.

The following data sources, are available for the development of the proposed approach.

- 100% samples of **Oyster journey data** for selected blocks of time. As a function of London's fare policy for rail journeys, transactions from its smartcard ticketing system record the stations and times of entry *and* exit for each journey. As mentioned, this data source does not cover journeys made using other fare payment methods.

- Aggregate **gateline counts** of entries and exits at stations with Overground services. Gatelines automatically record the total entries and exits over each fifteen-minute time interval, including passengers using non-Oyster fare media.

- **Loadweigh** measurements from new rolling stock, which will automatically sense and report the weight of the payload of each rail car. These weight measurements can be transformed into passenger counts through calibrated models relating passenger counts and weights.

- A complete set of **manual passenger counts** conducted in the Spring of 2009, covering all weekday services on the entire Overground network.

- A **network representation** of the London Underground (LU), London Overground (LO), Docklands Light Railway (DLR), and selected National Rail (NR) services. The particular network model to be used is developed by London Underground as part of their Rolling Origin and Destination Survey (RODS).

This thesis proposes to use the above automatic data sources to meet the immediate needs of the Overground's managers through the following.

- Estimate passenger loads on trains directly from loadweigh data systems. This requires calibration of the loadweigh systems to understand the associated measurement error, as discussed in Chapter 4.

- Use a mathematical process to estimate origin-destination flows by combining those passenger loads with automatic gateline entry/exit counts, representative Oyster journey data, and selected strategic manual counts. This requires a significant modeling effort, as discussed in Chapter 5.

- Using the models developed for OD estimation, assign the estimated OD matrix to the Overground network to determine the total number of boardings at each station. This is also discussed in Chapter 5.

- Estimate the total number of trips on the Overground (where a trip can include multiple boardings) as the sum of the OD matrix. Note that the current strategy does not provide estimates of this quantity.

Table 3-1 summarizes the current and proposed strategies.

Analytical Quantity	Current			Proposed		
	Data Source	Temporal Aggregation	Statistical Basis	Data Source	Termporal Aggregation	Statistical Basis
Train Loads	Manual Counts	Train Trip	1 Day	Loadweigh	Train Trip	Weeks/ Months
OD Matrix	LTDS/ Railplan	AM Peak Only	1 Day	Oyster & Loadweigh & Gatelines	Time Period	Weeks/ Months
Boardings & Alightings	Manual Counts	Train Trip	1 Day	OD Matrix	Time Period	Weeks/ Months
Total Ridership	N/A	N/A	N/A	OD Matrix	Time Period	Weeks/ Months

Table 3-1: Current and proposed data collection strategies for the London Overground

3.3 Data Sources

The balance of this chapter describes the various data sources upon which the proposed approach and other aspects of this thesis depend, including known issues for each source that require investigation.

3.3.1 The Oyster Smartcard Ticketing System

As discussed in Chapter 2, the structure of London's fare policy and technologies requires most Oyster users to validate their cards upon all entries and exits to the system. The centralized computer systems that support the Oyster system record and archive these entry and exit transactions in an easily accessible modern database. As a result, disaggregate Oyster journey data are cheap to gather in large volumes, and provide a prime source of data on individual passenger journeys and aggregate OD flows.

Oyster transactions are stored in the Oyster "Central System" across a collection of database tables – one for rail entries, one for rail exits, one for bus boardings, etc. A specialized query has been designed to extract the necessary data from these tables to support research purposes such as that described in this thesis. This query links data from these tables with each other and with reference tables to produce a single table describing all journeys recorded in the Oyster system (Gaitskell, 2008). Some fields of this table are populated differently for bus and for rail journeys. For rail journeys, the information provided includes:

- The Oyster card identifier ("card ID"), uniquely identifying each Oyster card in the database, anonymized to protect passenger privacy. It is typically assumed in analyzing Oyster data that each card ID represents a unique passenger.
- The station and time of first entry into the system.
- The station and time of last exit from the system.
- The date of the journey (as determined by the date of the entry transaction).

- The fare type of the journey (*i.e.* single fare, unlimited use, or a mix between the two).
- The fare paid.
- The innermost and outermost fare zones for which the journey was charged.

These data provide an extremely rich corpus with which to study many aspects of passenger demand, behavior, and experience on London's public transport network. However, Oyster data do not provide a complete picture for a number of reasons.

- Not everyone uses Oyster. The penetration rate across all TfL services is estimated to be approximately 80% (Transport for London, 2009*e*), but varies in space and in time across the TfL network (Chan, 2007).
- Some stations, including many of those served by the Overground, are ungated. Passengers using these stations in unlimited-use fare categories (*e.g.* weekly and monthly travelcard users) are not required to validate their Oyster cards at ungated stations.
- Oyster data describe only the first and last station used on a given trip. TfL's rail network is integrated and complex, with many journeys involving free interchanges, many stations served by numerous services, and multiple possible routings between station pairs. Many Oyster records are inherently ambiguous with regard to whether the Overground was in fact used at all, depending on the available routes between the origin and destination stations.
- The timestamps of all Oyster transactions are stored in the Central System in a truncated form – they indicate the time of day in minutes but not in seconds. Consequently, the times of passenger entry and exit available for research purposes are less precise than would be desired.

Nevertheless, as discussed in Chapter 5, Oyster data will play an important role in estimation of OD matrices for the Overground. The methods for estimating OD matrices that will be considered generally depend on some prior estimate of the OD matrix (also called a "seed matrix") to produce good results. The Oyster system will provide that estimate. Oyster data will also be used to analyze passenger station arrival behavior in Chapter 6 and passenger journey times in Chapters 8 and 9.

3.3.2 "Loadweigh" Train Payload Weighing Systems

The term "loadweigh" refers to electronic systems that estimate train payloads from measurements of air pressure in suspension systems (Interfleet Technology, 2004). All new Overground rolling stock are equipped with loadweigh systems for the explicit purpose of estimating passenger loads. Loadweigh data, although not yet tested on the Overground, are expected to allow estimation of actual loads on trains with reasonable accuracy, however saying nothing about boardings and alightings at each station. Experience with these systems at other Train Operating Companies is positive (Southern Railway LTD, 2009), but they have not tried to use these data in conjunction with other sources such as Oyster.

Loadweigh systems measure the weight of a train's contents, and thus report their measurements in units of mass (which of course is directly proportional to weight) (Interfleet Technology, 2004). As a result of variability in passenger weights and possible measurement error in the loadweigh system itself, there is expected to be some error associated with the estimates of passenger loads derived from loadweigh data. Chapter 4 will explore this issue further.

3.3.3 Station Gatelines

Entry and exit counts from station gatelines are typically used by the London Underground to scale up the results of manual origin-destination surveys (Maunder, 2003) or Oyster-based seed matrices (Wilson et al., 2008). In the Overground case, a large portion of traffic starts or ends at stations shared by multiple rail providers, so gateline data are ambiguous with respect to whether a given passenger used the Overground at all. Eleven stations (out of 56) are both fully gated and exclusive to the Overground. For the AM Peak period in the Spring of 2009, these eleven stations admitted an estimated 8,600 passengers out of the estimated 39,000 total Overground boardings (22%). Another 21 stations are gated but provide access to other London Underground or National Rail services.

3.3.4 Manual Passenger Counts

Under the current manual counting scheme, the boarding, alighting, and on-train load are sampled at most once for each scheduled service at each station. This presents the obvious statistical issue of assuming there is no day-to-day variation in demand which, lacking any other data, is remedied only at unreasonable cost. Nevertheless, these manual counts should be of use in testing and validating OD estimation methodologies, especially in the absence of complete loadweigh data.

Overground management expects to continue manual counts on non loadweigh-enabled portions of the network until new rolling stock is delivered. However, the concession contract allows LOROL to substitute loadweigh estimates for manual counts as loadweigh becomes available. This, combined with recent cost reduction initiatives at TfL, severely limits the ability of TfL to sponsor additional manual counts in the future (Smales, 2010).

3.3.5 Network Representations

TfL maintains (at least) two detailed representations of London's public transportation network. Corporate and strategic planning groups developed and use the RailPlan model, which represents all bus and rail modes in London, for long-term investment planning. RailPlan is implemented inside the proprietary EMME/2 transportation planning software package, and is regularly updated and modified by staff across TfL for various planning tasks.

The Strategy and Service Development group at London Underground has developed a rail-only model, focused on its own network but including competitive services, for use in its Rolling Origin and Destination Survey (RODS is in fact a combined OD estimation and network assignment model). It is implemented as a suite of custom in-house software tools and represents the transport network as a set of easy-to-share flat files. The RODS model and data are well explained in internal TfL documentation (Maunder, 2003).

The network representation from the RODS model will be used in this work because of its open and easy-to-share nature, and because it was designed and implemented with OD estimation in mind. It was not, at the outset, sufficiently detailed with respect to the Overground, but the additional detail was straightforward to add. Most of the necessary updates were to

- distinguish Overground services from other National Rail services at certain stations,
- reflect current Overground service patterns and frequencies,
- and explicitly represent entries, ticket halls, and exits at Overground stations lacking those features in the model.

3.3.6 Additional Data Sources

Train Control Systems – Modern train signaling, supervision, and control systems record extensive data indicating actual railway operations. Many systems, including the National Rail network used by the Overground, record the movements of every train on the network. The use of this type of data to support operating strategies and tactical planning has been researched extensively, for example by Rahbee (1999, 2006), Vescovacci (2001), Lee (2002), and Carrel (2009). This thesis does not deal with this type of data directly, but uses performance data and cites other analyses of Overground operations both derived from these type of data.

National Rail Performance Monitoring Systems The National Rail network has an elaborate performance monitoring and delay attribution framework in place. Its TRUST system uses data from signal and control systems to monitor and record all train movements (Office of Rail Regulation, 2010). These records are used to calculate the Public Performance Measure (PPM), a measure of train on-time performance at terminals (Office of Rail Regulation, 2008). A complex delay attribution methodology, requiring substantial manual inputs, is used to allocate delays to responsible parties (TOCs, Network Rail, etc) for the sake of performance monitoring and financial remuneration (Network Rail, 2009). Chapters 8 and 9 use some of these outputs to evaluate the proposed measures of service quality and the outcome of the tactical planning intervention on the North London Line.

The London Travel Demand Survey The London Travel Demand Survey (LTDS) is a London-wide household travel survey updated on a rolling basis over time. This survey, along with census data and other sources, underpins an area-wide transportation planning model (also called LTDS). One output of that model is an estimate of public transport travel in London on a zone-to-zone basis (zones in this case are traffic analysis zones, rather than fare zones).

Railplan Regional Public Transport Model The LTDS estimates of public transport travel become inputs into the RailPlan regional public transport assignment model.

Railplan uses a detailed representation of London's public transport network to estimate demand for existing services, and to model the effects of proposed changes to the network and of forecast economic growth (cf AECOM, 2006). It is used almost exclusively to model transport for the AM Peak period. Railplan does not use direct measurements of the sort discussed above to estimate current conditions. Instead, it is, from time to time, validated against other estimates of demand. As has been found (AECOM, 2006) and will be shown in Chapter 5, Railplan's estimates often diverge from other more believable estimates by 50% (or more) at fairly aggregate levels. Nevertheless, Railplan is currently the only source of origin-destination matrices used by the Overground for various analysis and tactical planning tasks.

Public Timetables TfL has access to public timetables for Overground services in plain-text formats. These timetables become an integral part of the work in Chapter 6, 8, and 9.

Chapter 4

Calibration of Loadweigh Systems

The term "loadweigh" refers to electronic systems that estimate train payloads from measurements of air pressure in suspension systems. Loadweigh systems measure and record the weight of a train's contents in units of mass. Passenger loads are inferred from these measurements of weight by means of an estimated or assumed average passenger weight (Interfleet Technology, 2004). As a result of variability in passenger weights and possible measurement error in the loadweigh system itself, there will be some error associated with the estimates of passenger loads inferred from loadweigh data.

This chapter develops a methodology by which to infer train loads in units of passengers from loadweigh measurements in units of weight. The basic idea of this methodology is to regress loadweigh data on corresponding manual passenger counts to estimate average passenger weight and vehicle tare (*i.e.* unladen) weight. These two parameters can then be applied to infer passenger load from new loadweigh data in the future. The methodology also provides an estimated bound on the magnitude of random error associated with these estimates.

Section 4.1 reviews the little academic and industry literature available on the topic. Section 4.2 discusses the various sources of error in loadweigh systems and develops a simple linear regression model for calibrating loadweigh systems. Section 4.3 presents an exploratory analysis of a sample of loadweigh data and manual counts from the London Overground. Section 4.4 applies the model to data from the London Overground and presents the results. Section 4.6 draws conclusions and offers some recommendations.

4.1 Literature Review and Industry Experience

Loadweigh systems were developed in the UK by British Rail Research and have since been commercialized by aftermarket vendors and by rail car manufacturers such as Bombardier (Interfleet Technology, 2004). While they have been used in the UK for over a decade (Smale, 2010), in some cases with documented positive results (*e.g.* Southern Railway LTD, 2009), very little analytical literature exists on the subject. Most of the accumulated knowledge and experience with these systems appears to be within industry parties, such as equipment suppliers, information service providers, and the railways themselves.

In a telephone interview, Smale (2010), director of UK railway information services provider Demon Information Systems, reports from internal company research that loadweigh data are "accurate to within ±20 passengers at a 95% confidence interval for a three-car train." He reports that this error is not related to the passenger load, and so is larger as a percentage of the load at smaller loads. He reports that average passenger weights of 75-85 kilograms have been estimated in previous calibration exercises, and recommends using 80kg for the London Overground. These exercises have used very careful and expensive manual counts taken by two surveyors on each rail *car*. Technical details as to how the calibration was accomplished from loadweigh data and passenger counts are not available. Finally, he reports that the loadweigh measurements recorded and provided by Bombardier, the manufacturer of the new Overground fleet, are in fact the average of a series of measurements taken at twenty cycles per second as trains travel between consecutive stations.

Bombardier Transportation (2004) conducted a systematic but small-scale study involving three different kinds of tests, each conducted on a single train trip with encouraging results. In the first test, they loaded a known amount of weight (17,970kg to be precise) onto a train which was then run along the length of the route. They observed some variation in the en route loadweigh measurements, likely as a result of the dynamics of the train's suspension, but all recorded measurements (*i.e.* the averages of the 20hz measurements) were within 5% of the (constant) true weight. In the second test, they ran a train in passenger service and "a number of observers" counted the number of occupants at each station. They multiplied these passenger counts, which were never more than 250, by an assumed average passenger weight of 80kg and compared the estimated weight to the measured weight. The estimated was at times larger than the measured weight and at times smaller, but they never diverged by more than 6.25%. In the final test, 34 passengers of known weight rode a test train along its run, circulating en route between cars on the train. The measured weight diverged from the known weight by more than 5% in only one instance (out of 12) when there was a temporary equipment failure.

This study concerned Bombardier class 375 rolling stock, which are very similar in design to the Overground's new class 378 fleet, and concluded that their loadweigh systems were fit for purpose. It acknowledged that assuming an average passenger weight of 80kg can introduce some error, especially for particular days, but that "fluctuations of this nature will even themselves out over a period of time and multiple journeys over the same route."

Researchers in Copenhagen, Denmark have analyzed data from built-in loadweigh systems and from expensive aftermarket infrared passenger-counting systems for their S-Train suburban rail network. Unfortunately, the S-Train operator has not yet allowed the authors to publish their work in detail, so personal communications, presentation slides (Nielsen et al., 2008*a*), and an abstract (Nielsen et al., 2008*b*) are the only available references for this work. The researchers compared measurements from both systems with corresponding highly accurate manual counts taken from recorded video footage (Nielsen, 2009*b*). They used regressions to make these comparisons (Nielsen, 2009*a*) but the precise form of the regressions has not been revealed. Their presentation slides indicate that the estimates of passenger loads derived from loadweigh data were unbiased and had a random error with a standard deviation of about 14 that did not vary meaningfully with the actual load. The infrared systems had a smaller standard deviation of 0.75 but for loads above 50 passengers had a negative bias of about 7%.

(Nielsen et al., 2008*b*) concluded that, for their particular client, loadweigh systems were preferable to infrared counters despite the much larger random error because loadweigh systems were unbiased and available on all trains at no extra cost. After analyzing the loadweigh systems, they went on to estimate OD matrices from an existing survey-based OD matrix and passenger loads inferred from loadweigh data (similar to what is proposed here). This is discussed further in Chapter 5. They have implemented a production system for the S-Train operator that automatically estimates passenger loads and OD matrices on a daily basis for use in management, planning, and even revenue allocation among competing services. To date, this system assumes a single average passenger weight across the entire S-Train network for all times of day and and days of the year. The researchers plan to further explore the variation in passenger weights across these dimensions.

4.2 Model Development

Any estimate of passenger loads from loadweigh measurements will at some level be based on assumptions or inferences about passenger weights. The model proposed here is consistent with the literature discussed in the previous section in that passenger weights are parameterized in terms of a single average value. This value can potentially change over time or across market segments, but any group of passengers is described only in terms of their average weight. This average weight, as well as other calibration parameters, will be estimated through pairwise comparison of loadweigh measurements with corresponding manual passenger counts. These estimates of passenger loads will thus be affected by the following four sources of error.

- *Random measurement error in the loadweigh system* – random error can occur in the loadweigh system through any number of known or unknown factors, including the dynamics of train motion and suspension systems.

- *Systematic measurement error in the loadweigh system* – a constant error associated with the loadweigh system, for example by a non-zero tare weight.

- *Variation in true average passenger weight* – the difference between the actual average weight of passengers and the assumed or estimated value.

- *Random measurement error in the manual passenger counts* – random error associated with the manual counts with which loadweigh data are compared.

For a single loadweigh measurement and corresponding manual count taken on a certain train at a certain time, let

W = the loadweigh measurement,

η = the random error associated with the loadweigh measurement system itself,

α = the systematic measurement error of the loadweigh system,

C = the true number of passengers on that train at that time,

ω = the random error associated with the manual count,

β = the true average weight of all passengers (on all trains),

ν = the random error associated with the actual average weight of passengers on the train in question as compared to β.

The relationship between loadweigh measurements, manual counts, and these various sources of error can be characterized by the equation

$$W - \alpha + \eta = \beta(C + \omega) + \nu. \tag{4-1}$$

Assume that η, ν, and ω are symmetrically and independently distributed each with mean zero. Let

$$\varepsilon = \beta\omega + \nu - \eta.$$

Then ε also has a mean of zero. Equation (4-1) can be rewritten as

$$W = \beta C + \alpha + \varepsilon. \tag{4-2}$$

This equation satisfies the form of a single-variable linear regression, and so its parameters (β, α, ε) can be estimated by the method of ordinary least squares (OLS) (Greene, 2007).

Let the OLS estimate for a parameter λ be denoted $\hat{\lambda}$. Substituting the estimated parameters into Equation (4-2) and solving for C yields

$$C = \frac{W}{\hat{\beta}} - \frac{\hat{\alpha}}{\hat{\beta}} - \frac{\hat{\varepsilon}}{\hat{\beta}}. \tag{4-3}$$

This equation is important for two reasons. Firstly, it makes apparent the interpretation of the two following terms. $\frac{\hat{\alpha}}{\hat{\beta}}$ is the estimate of the tare weight (*i.e.* the systematic error) of the loadweigh system in number of passengers. Likewise, $\frac{\hat{\varepsilon}}{\hat{\beta}}$ is the estimate of total random error for a particular observation in number of passengers.

Secondly, it can be used to estimate passenger loads from *new* loadweigh measurements (*i.e.* those lacking corresponding manual counts). Consistent with application of linear regression models, such an estimation assumes that ε for these measurements is equal to zero, which yields

$$C = \frac{W - \alpha}{\hat{\beta}}. \tag{4-4}$$

In other words, to estimate the number of passengers, subtract the tare from the loadweigh measurement and divide the result by the average weight per passenger.

The standard deviation of $\frac{\hat{\varepsilon}}{\hat{\beta}}$ is also an important quantity. It can be interpreted as an estimate of the random error in passenger loads inferred from loadweigh data. This random error consists of the random loadweigh measurement error and the variability in average passenger weight; factors which are beyond control for a given loadweigh system. In fact this estimate is an upper bound of the standard deviation of random error introduced by these factors. It is an upper bound because additional error is introduced in the calibration process by the manual counts (*i.e.* ε depends on ω). The bound is tight if and only if the manual counts with which it was calibrated were perfect (*i.e.* $\omega = 0$). This bound is important because it indicates the overall statistical reliability of passenger load estimates derived from loadweigh data. It will be referred to here as the "error bound" for loadweigh data, and will be signified by the variable δ.

Assume that the random error in passenger loads inferred from loadweigh data is normally distributed with mean zero and standard deviation equal to the error bound $\delta = \mathrm{sd}\left(\frac{\hat{\varepsilon}}{\hat{\beta}}\right)$. Then, by the properties of the normal distribution, passenger loads inferred from loadweigh data should be accurate to within at most $\pm 1.95\delta$ at the 95% confidence interval.

4.3 Exploratory Analysis

The London Overground provided a sample of 13,121 weekday loadweigh measurements from 13 different units (*i.e.* full 3-car trains) serving the North and West London Lines from 23 November through 6 December, 2009, inclusive. These data were extracted from on-board systems and processed into usable formats by Bombardier Transportation, the rolling stock provider. Each observation reports a number of elements, including train unit number, departing station, date and time of departure, and loadweigh measurement in kg. It should be noted that about half of the units on the North and West London Lines were from the new loadweigh-enabled fleet at the time of this data sample.

Corresponding manual counts were also provided for 1,253 of these observations over 80 different vehicle trips on 24 November and 1-2 December, 2009. This is the data set on which the model from the previous section will be estimated. 115 of these observations were taken on the West London Line, the balance on the North London Line. The manual counts were taken by a pair of observers at each platform, and are anecdotally reported by their provider to have an accuracy of ± 20 passengers at a 95% confidence interval. It should be noted that the provider of these counts explicitly recommends the use of a more intensive counting scheme, with two observers on *each car*, for the purposes of loadweigh calibration (Smale, 2010)

Figure 4-1 plots the distribution of these loadweigh measurements, which range from 0kg to 47,060kg (588.25 passengers at 80kg each). This plot illustrates that loadweigh data, at least from the Overground fleet, vary smoothly over a considerable range.

Figure 4-2 plots loadweigh measurements against time of day for a random 10% sample of the provided data. It illustrates that the highest loadweigh measurements indeed occur in the morning and evening peaks experienced throughout the TfL network.

Figure 4-3 plots loadweigh measurements against time of day for all observations on a single link in the Overground network. The selected link is from Canonbury to Highbury & Islington, generally held to be the peak load point of the North London Line during the AM Peak period. This plot shows directional effects, where loadweigh measurements are greater in the morning than in the evening, consistent with expectations.

It also illustrates some day-to-day variation, with variance in measurements taken at the same time of each day (*e.g.* just after 20:00). Being taken at the same time of day at the same location, these measurements are likely taken from the same scheduled service. As further discussed in Section 4.5 this indicates the presence of day-to-day variability in on-train loads that the Overground's current single-sample on-board counts do not capture.

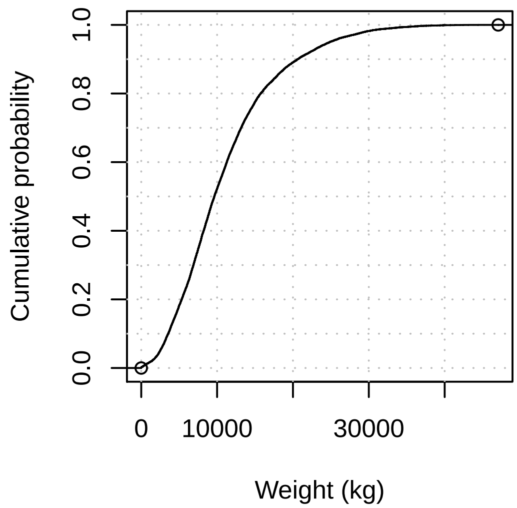


Figure 4-1: Cumulative distribution of London Overground loadweigh measurements

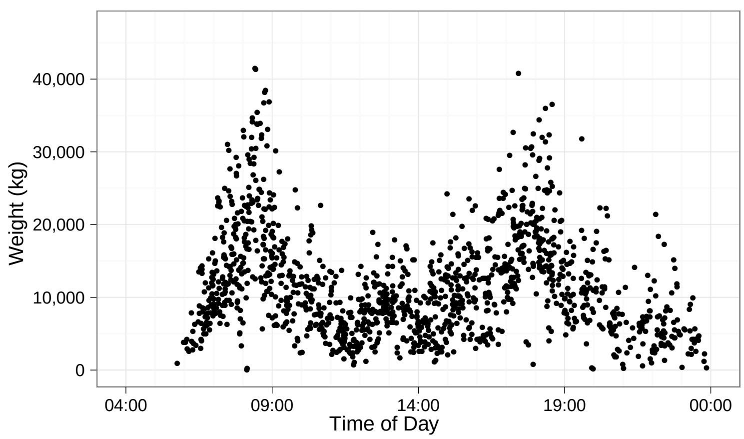


Figure 4-2: Loadweigh weight vs. time of day (random 10% sample)

Figure 4-4 plots loadweigh measurements against corresponding manual counts for all stations and for Stratford and Richmond, the two main North London Line terminals. It shows a clear but not perfect linear correlation between the two variables, with a correlation coefficient of 0.927 for the whole data set. This correlation is significantly higher for the observations from the terminals, with a coefficient of 0.992.

The higher correlation at terminals could be explained by more accurate passenger counts at those locations, which could result for the following reasons. Overground trains generally spend much longer at terminals waiting to depart than they do at stations en route, so observers have time to count accurately the number of boarding passengers. Moreover, at terminals there is little if any simultaneity in passenger boarding and alighting, which simplifies the passenger counting task. In the lexicon of the previous section, ω should be smaller for manual counts taken at terminals. The general shape of these plots suggest that a

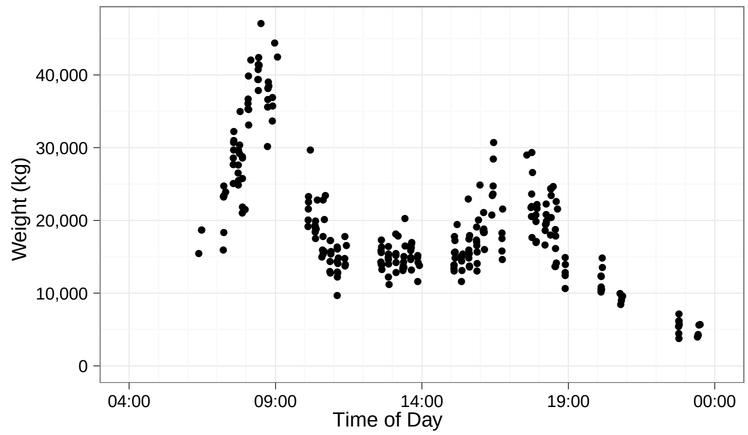
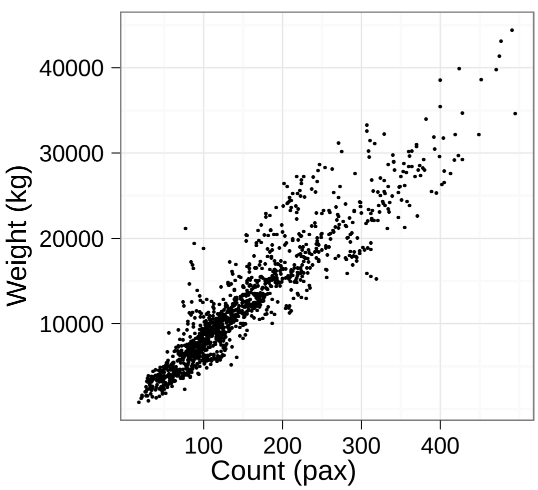
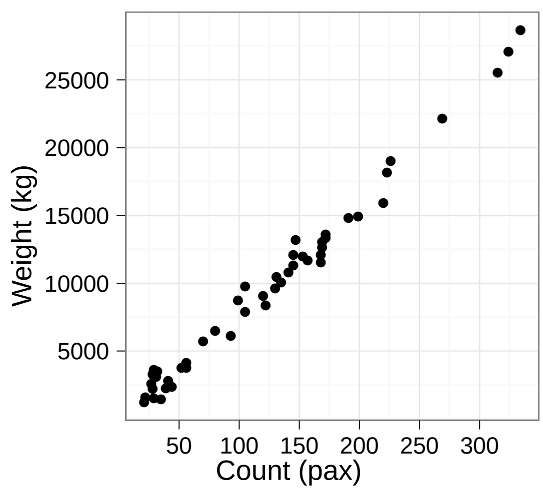


Figure 4-3: Loadweigh weight vs. time of day for peak load point of London Overground network (Canonbury to Highbury & Islington)

linear regression model is an appropriate modeling framework. They also suggest that such a model will have a better fit and produce more accurate parameter estimates for observations taken at terminals than for observations taken en route.



(a) All stations



(b) Terminals only

Figure 4-4: Loadweigh weight vs. manual count data

4.4 Calibration Results for the London Overground

The model of Equation (4-2) was regressed on the calibration data set in the following ways.

Pooled – the model was estimated on the entire calibration set of 1,253 paired loadweigh and manual count observations.

Unit Segmentation – the model was estimated on data for each of the 13 full-train units separately. This series of regressions explores the possibility of variability in the functioning of loadweigh systems across different trains. It can be considered an unrestricted version of the pooled model. The observations are not uniformly distributed across the different units. The number of observations per unit ranges from 15 to 267. The two summary statistics as well as the observation-weighted estimate of β were also estimated for the joint results of this entire series of regressions.

Terminals Only – the model was estimated on data from just the two main North London Line terminals. This reduced data set contains only 49 observations, 25 from Stratford and 24 from Richmond. This regression should indicate whether the manual counts from terminals are more accurate than manual counts taken en-route. The regression was also run on terminals-only data for the two units with at least 10 observations at terminals.

The regressions were estimated using the open source statistical programming language and environment R (The R Project, 2009). In addition to parameter estimates, the adjusted R^2 and the error bound, $\delta = \mathrm{sd}\left(\frac{\hat{\varepsilon}}{\hat{\beta}}\right)$, were calculated for each regression.

For all of these regressions, β is expected to be in the 75 - 85 kg range. There is no particular expectation on the sign or magnitude of α, nor is there an expectation as to how different the parameters from the unit segmented regressions should be from each other. Estimates of α that are small in magnitude or are not statistically significant will indicate small or non-existent systematic bias in the loadweigh measurement systems. Based on the literature, it is expected that the variance of the residuals of these regressions will be constant (*i.e.* homoscedastic). The statement of Smale (2010) that loadweigh data are "accurate to within ±20 passengers at a 95% confidence interval for a three-car train" would be supported by finding a δ of approximately 10.

Table 4-1 shows the results of these model estimations. All estimates of β, the average weight per passenger, are statistically significant at the 0.1% significance level. β for the pooled and terminals only model, 77.3 and 81.4, respectively, are both within the expected range but substantially different from each other. The estimates of β for the unit segmented regressions cover a wide range, from 62.4 to 85.0. In a standard F-test, the disaggregation of the pooled model into the joint unit segmented model is overall statistically significant at the 0.1% level.

The estimate of α, the tare weight, is statistically significant in some of the models estimated on all of the calibration data. The sign is positive in some cases, and negative in others. The magnitude ranges widely for the unit-segmented regressions, from 115kg (estimated 1.5 passengers) to 3,686kg (estimated 47.8 passengers).

The terminals-only regressions appear to provide substantially better results than those estimated on data from all stations. The estimates of β for the two units with at least 10 observations, 79.9 and 79.7, are close to the estimate over all terminal observations and are almost identical to each other. The estimate of α is, perhaps tellingly, small in magnitude (at most 6 passengers) and in all cases not statistically significant. In terms of the adjusted R^2 and the estimate of the error bound, δ, the terminals-only regression results are far superior: its R^2 is 0.98 compared with 0.86 for the pooled regression and 0.88 for the joint results of the unit segmented regression.

Perhaps most importantly, the terminals-only regression estimates a δ of 10.8, and 5.0 and 11.4 for the two segmented units, very much in line with industry expectations. The pooled and joint unit segmented regressions, both estimated using observations from all stations, estimate a δ of 35.3 and 31.9, respectively.

Model	Obs.	Trips	Avg Weight β	Tare Weight α	Tare in Pass. $\frac{\alpha}{\beta}$	Error Bound δ	$\bar{R}^2$
Pooled	1,253	80	77.3 ***	572 ***	7.4	35.3	0.86
Unit Segmentation							
378007	139	8	62.4 ***	2,070 ***	33.2	29.1	0.91
378008	15	3	66.1 ***	2,161 *	32.7	16.0	0.97
378010	237	12	85.0 ***	-835 *	-9.8	28.5	0.89
378011	223	16	78.4 ***	115	1.5	36.3	0.88
378015	191	12	82.7 ***	-905 **	-10.9	30.9	0.91
378016	56	3	77.1 ***	3,686 ***	47.8	25.3	0.81
378017	267	18	76.8 ***	1,442 ***	18.8	37.0	0.82
378018	125	8	80.4 ***	453	5.6	27.0	0.85
(joint)	1,235	80	78.2			31.9	0.88
Terminals Only	49	49	81.4 ***	-329	-4.0	10.9	0.98
Unit Segmentation (Terminals Only)							
378010	10	10	79.9 ***	-174	-2.2	5.0	1.00
378017	10	10	79.7 ***	475	6.0	11.4	0.99

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4-1: Loadweigh calibration results

Figure 4-5 plots the residuals, $\hat{\varepsilon}$, against the manually counted number of passengers for two of the estimated models. It shows the residuals from the regression on all data with the lowest δ, the joint unit segmented model, and on the pooled terminals-only data. For the regression on all data the variance of the residuals appears approximately constant at passenger counts above 100 or 150 passengers. The standard deviation (in kg) of the all residuals and for residuals only from observations at terminals are 2,502 and 951, respectively. In the terminals only regression the variance of the residuals appears constant throughout, and is clearly much less than the variance of the residuals from regressions estimated on all of the data. In this regression, the standard deviation of the residuals is 885.

These results indicate the desirable property of homoscedasticity in the error terms of the regression models. More importantly, they indicate that the fit between manual counts and loadweigh data is much tighter for observations at terminals that at other stations, even when the model is estimated with the entire data set. The implications of this difference is discussed further in the following section.

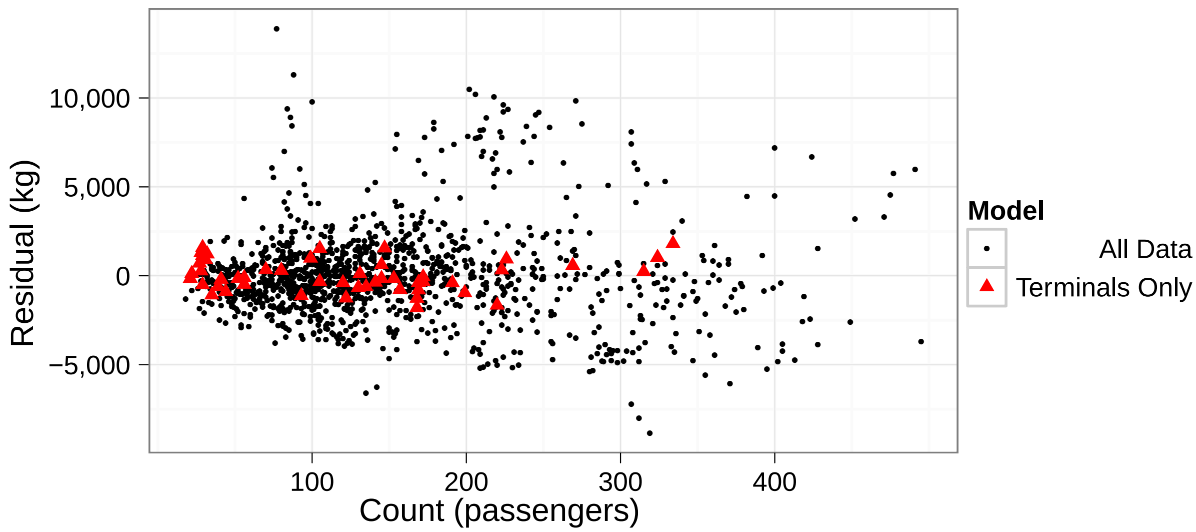


Figure 4-5: Residual vs. manual count for all data (unit segmented) and terminals-only (pooled)

4.5 Variability of On-board Loads

The key analytical quantity that the London Overground looks to loadweigh data to provide is on-board loads for each scheduled service between each pair of stations. Given the results of the previous section, it is possible to start to assess the statistical quality of the existing on-board counts, sampled precisely once per counting period, that loadweigh systems are intended to replace.

Figure 4-6 shows loadweigh measurements for trains traveling from Canonbury to Highbury & Islington between 08:00 and 09:00 – the peak load point of the North London Line in the peak hour. At the time these counts were taken, the 08:09 train from Stratford to Richmond was the only service in the timetable on this link between 08:20 and 08:30. It is thus reasonable to assume that those loadweigh observations on this link between those times were taken on that scheduled service. It is necessary to make this assumption because loadweigh data are not yet associated with specific scheduled services.

There are 7 observations (from 7 different days) in Figure 4-6 between 08:20 and 08:30, with measured loads from 37,870kg to 42,400kg. Using a β of 80kg, the load in number of passengers estimated from these measurements range from 473 to 530, with a mean of 506. The single manual calibration count associated with this link for this scheduled service was 475 passengers. For the trip on which this scheduled service was manually surveyed, this link was indeed the peak load point, as is generally assumed for westbound travel on the North London Line in the AM Peak period. If the mean of these estimated passenger loads (*i.e.* 506) is taken to be the true average, the single-sample manual estimate of the average peak load on this service (*i.e.* 475) was short by 31 passengers, or 6.1%.

The standard error of the 7 load estimates is 18.5 passengers. This can be interpreted, somewhat speculatively, as an estimate of the standard deviation of passenger load estimates

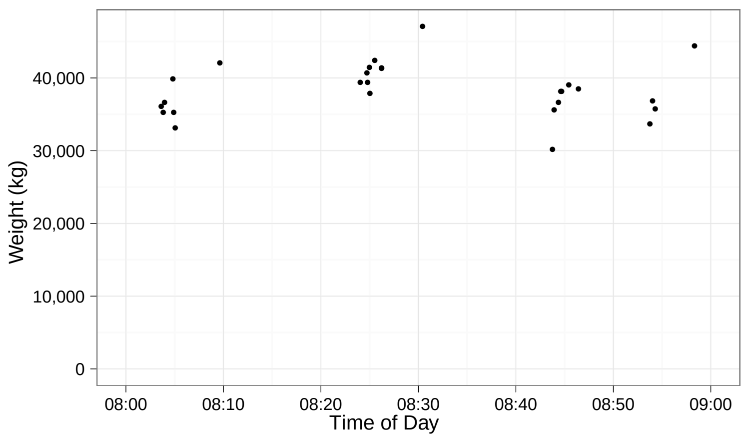


Figure 4-6: Loadweigh weight vs. time of day for peak load point data

on this link on this service *including* the random error introduced by the loadweigh system and calibration process. Consequently, an estimate of the "true" average load from a *single* loadweigh sample would be accurate to within ±37 passengers at the 95% confidence level. However, with 14 loadweigh samples the accuracy could be improved to within ±10 passengers at a 95% confidence level (*i.e.* a standard deviation of 5).

In the previous section, the best estimate of the error bound associated with passenger loads inferred from loadweigh data was 10.8. Under the assumption of independent and normally distributed errors this implies that the standard deviation of "true" passenger load with respect to the "true" average load is 7.7 passengers (*i.e.* $18.5 - 10.8$). The error bound of 10.8 was estimated using data only from terminals. The lowest estimate of the error bound using data from all stations was 31.9. This can be interpreted, again somewhat speculatively, to imply a standard deviation of just over 21.1 passengers for the error associated with manual counts at *non-terminal* locations. The total standard deviation of one manual count, with respect to the "true" average on this link for this scheduled service, can then be estimated at 28.8 (*i.e.* $21.1 + 7.7$), implying an accuracy of ±57.6 at the 95% confidence level. For reasons of cost, the Overground is unable to take more than a single manual count of each scheduled service per counting period.

The analysis in this section indicates that a single estimate of passenger loads from loadweigh data may, at non-terminal locations, be more accurate than a single manual passenger count. Moreover, the automatic and easily obtainable nature of loadweigh data may be exploited to obtain substantially more accurate results at effectively zero marginal cost.

The precise numerical results of this section are somewhat speculative. They should not be assumed to be generally representative of the variability of passenger loads or the error in passenger counts on the Overground network as a whole. A much larger sample of loadweigh data (*i.e.* for many more links and scheduled services) should be used to understand the statistical nature of these factors.

4.6 Conclusions and Recommendations

This chapter analyzed loadweigh data from the London Overground, and used corresponding manual passenger counts to calibrate a model for estimating passenger counts based on loadweigh data. This section presents first some conclusions drawn from the analysis in this chapter, and next some recommendations based on those conclusions.

Conclusions

The foremost conclusion drawn from these results is that there is no basis for rejecting the practices for using loadweigh data recommended by Smale (2010). Rather, analysis of a limited data set (from terminals only) suggests that the recommended assumptions of an average passenger weight of 80kg and a 95% confidence interval of ±20 passengers, with no tare weight, (Smale, 2010) is in fact reasonable.

A corollary to this conclusion is that the manual counts used here are generally of insufficient quality for calibrating loadweigh systems. This is acknowledged by the provider of the counts. The primary explanation is that the counts used here were taken by a single pair of observers on each *station platform*, whereas the preferred counting strategy for loadweigh calibration is to employ a pair of observers on each *car* (Smale, 2010).

That said, the regression results support the theory that counts taken at departing terminals are more accurate than counts at stations served en route. The likely explanations for this are that the surveyors have much more time to count passengers and passengers typically alight and board more sequentially at terminals than at stations en route. The number of paired observations of loadweigh measurements and manual counts taken at terminals is limited to only 49 out of the whole set of 1,253. The regression on this limited set produces results very close to prior findings, specifically:

- It estimates the average weight of passengers at 81.4kg, with high statistical significance. This is only 1.8% different from the recommended value of 80kg.

- It estimates an error bound of 10.8kg. This is the upper bound of the standard deviation of random error in passenger loads inferred from loadweigh data. Assuming normally distributed random error, this amounts to a 95% confidence interval of ±21.2 passengers, as compared to the recommended ±20 passengers.

- It estimates a relatively small tare weight of 328.6kg (4.0 passengers), but this estimate is not statistically significant. For individual units, the estimate of tare weight is of the same order of magnitude and also not statistically significant. That is, there is no evidence to suggest that the tare weight is other than zero.

In terms of methodology, the linear regression model used in this chapter appears suitable for comparing loadweigh data with manual passenger counts. Because the residuals in these regressions are of approximately constant variance, the method of ordinary least squares is adequate to estimate this model.

A preliminary analysis of the temporal variability of passenger loads on scheduled services found that fewer than 20 loadweigh samples, available at effectively no marginal cost, would be sufficient to estimate the "true" average passenger load to within ±10 passengers at the 95% confidence level.

Recommendations

Given the results and conclusions of this chapter, it is recommended that loadweigh data be pursued as a low-cost and (sufficiently) reliable source of data on passenger loads on trains. This applies to the railway industry in general and specifically to the London Overground. Naturally, multiple loadweigh samples should be used to estimate the average load on any particular scheduled service for any particular link in the network. The number of samples required depends on the desired accuracy.

Additional research should be conducted into the nature and magnitude of the various sources of error associated with passenger loads inferred from loadweigh data. The weakness of the analysis in this chapter stems primarily from the low quality of the manual counts against which loadweigh data were compared. To remedy this and other issues, the following are recommended.

- More and better calibration data should be gathered and paired with corresponding loadweigh data. The regressions from this chapter should be used to re-estimate the calibration parameters and to estimate the error bound.

- If it is believed that there is no bias in the calibration parameters at terminals, these data could be gathered at terminals under the same counting procedures used to gather the data analyzed here. If data is required at stations other than terminals, "calibration-quality" counts should be taken, with procedures to ensure greater accuracy than was found in the overall data used here.

- In addition, the calibration-quality counts or additional terminal counts should be used to explore variability in the parameter estimates and error bound across different rolling stock units. If such analysis indicates significant differences between individual rolling stock units, individual calibration of all future rolling stock units as they are delivered may be recommended. Additionally, it should be investigated whether the actual loadweigh equipment on each unit has calibration parameters that can be adjusted.

- As identified by Nielsen et al. (2008*a*), additional analysis should be conducted at different times of year to assess the systematic variation in average passenger weights correlated with seasons and weather. It is possible that such variation could be ignored, but this question should be explored.

- Controlled experiments, such as the first and third tests described by Bombardier Transportation (2004), should be conducted on Overground rolling stock. This would entail placing a known amount of weight (be it in sandbags or human participants) onto loadweigh-enabled trains, running them along the length of the line, and comparing the loadweigh measurements to the known value. The primary purpose of such experiments is to understand the pure measurement error associated with loadweigh systems.

Applied to the Overground, these recommendations should of course be tempered by cost considerations. For example, the Overground may not be able to afford multiple series of calibration-quality passenger counts in the future. In this case, they may assume that the parameter and error bound estimates derived from a first series of counts, or even the estimates from this chapter, apply to all rolling stock at all times.

Chapter 5

Origin-Destination Matrix Estimation

Passenger demand for a public transport network can be expressed as a matrix of passenger origin-destination (OD) flows (an *OD matrix*). OD matrices are one of the most important inputs into many public transport planning and management applications (Meyer and Miller, 2001). This Chapter is primarily concerned with the estimation of OD matrices for the London Overground network. It develops, applies, and validates a methodology to estimate OD matrices from multiple data sources, including (i) journey transactions from automatic fare collection (AFC) systems (such as Oyster), (ii) manually counted and/or automatically measured on-board passenger flows in the network, and (iii) station entry and exit counts from automatic station gatelines. The balance of this section introduces some of the basic concepts necessary to discuss this topic, and then describes the organization of this chapter.

In transport modeling, *assignment* refers to the process of estimating how passengers use a given network to travel between their respective origins and destinations. One outcome of the assignment process applied to an OD matrix is an estimate of the total utilization of each link in the transport network. Such estimation depends crucially on models of passenger behavior (*e.g.* path choice) and of the relationship between link flow and link performance (*i.e.* congestion effects) (Meyer and Miller, 2001). These models, along with the choices of how to represent the transportation network both spatially and temporally, will be collectively referred to here as "assignment models."

Estimation of an OD matrix *given* links flows ("OD estimation"), which is the primary concern of this chapter, can be understood as the inverse of the assignment problem (Bierlaire, 2002). As such, it too depends on the assignment model. This should be intuitively clear in that, to estimate an OD matrix from given link flows, it is necessary to know the links used by passengers traveling between each origin-destination pair.

In addition, OD estimation often takes as additional input some prior estimate of the OD matrix (also referred to as a "seed" matrix), which is important because there are often infinitely many ways to construct an OD matrix to match the given link flows (Bierlaire, 2002). Historically, these prior OD estimates were expected either to be OD matrices directly estimated from expensive manual surveys and requiring updating to account for changing travel patterns, or to be OD matrices that came from regional transportation models that were not calibrated on measured link flows (Cascetta, 1984).

To estimate time period level OD matrices for the Overground, a credible assignment model must be adapted or developed. That assignment model will be used as part of an

OD estimation process which takes as input manually and automatically measured link flows and an Oyster-based seed matrix. As discussed in Chapter 3, Oyster represents only a lower bound on the true OD matrix (thus referred to here as a "fractional seed matrix"). In that sense, the problem faced by the Overground is somewhat different from that addressed by most of the established methodologies.

Sections 5.1 and 5.2 review some well-known assignment models and OD estimation methods, respectively, with an eye towards practical application to the Overground. Section 5.3 draws on that review to propose a detailed strategy for estimating OD matrices from AFC-based seed matrices and a flexible set of measured link flows. This strategy includes an assignment model that is tailored to the context of the Overground. Section 5.4 presents an exploratory analysis of data relevant to the proposed strategy. Section 5.5 presents and validates the results produced by the proposed method, and compares the validation of the estimated OD matrix to a validation of the existing Overground OD matrix produced by the RailPlan regional model. Section 5.6 draws some conclusions and makes recommendations for implementation.

5.1 Public Transport Network Assignment Literature Review

Broadly speaking, assignment models are those models which are used to estimate the utilization of different parts of a transport network given an estimate of demand in the form of an OD matrix. Passengers can be assigned to certain paths through the network, specific services in the network, or individual links on the network. Nuzzolo (2003) provides a good review of public transport assignment models. What are referred to in this chapter as "assignment models" actually consist of a hierarchical family of models, only one of which is, in the literature, referred to by this name. At the lowest level of the hierarchy are *supply models* – those that represent the public transport network itself. The next level is *path choice models* – behavioral models that describe how passengers identify, evaluate, and choose among different paths through the network. The highest level is what are narrowly referred to in the literature as "assignment models" – those which build on the other types of models to estimate and report the utilization of the network in spatial, and possibly temporal, dimensions. Certain kinds of assignment models are associated with certain supply and path choice models, so this distinction is somewhat arbitrary. This section reviews selected literature on these members of the broad family of assignment models, focusing on those aspects relevant to the work in the balance of this chapter.

5.1.1 Supply Models

Supply models represent the public transport network itself, including the various services on the network and the physical infrastructure for pedestrian access and egress to and from these services. Transport networks are typically represented as *graphs* – data structures composed of *nodes* and *edges*, or *links*, that have an extremely wide set of applications in many fields beyond transport analysis (*e.g.* Ahuja et al., 1993). The two most common representations of public transport supply are *line*-based and *run*-based, where the latter is more granular in the time dimension (Nuzzolo, 2003).

Line-Based Supply Models

Nguyen and Pallottino (1988) and Spiess and Florian (1989) describe and use the line-based model of public transport supply in their seminal works on public transport assignment. They represent the network with two distinct subgraphs – one representing the various transport services and the other representing the pedestrian infrastructure. Figure 5-1 illustrates this supply model with an example which includes some aspects of the station environment, such as the entrance and the exit, with walking links to and from the platform. In this example, Line 1 serves Station A, B, and C, whereas Line 2 bypasses Station B on the way from A to C.

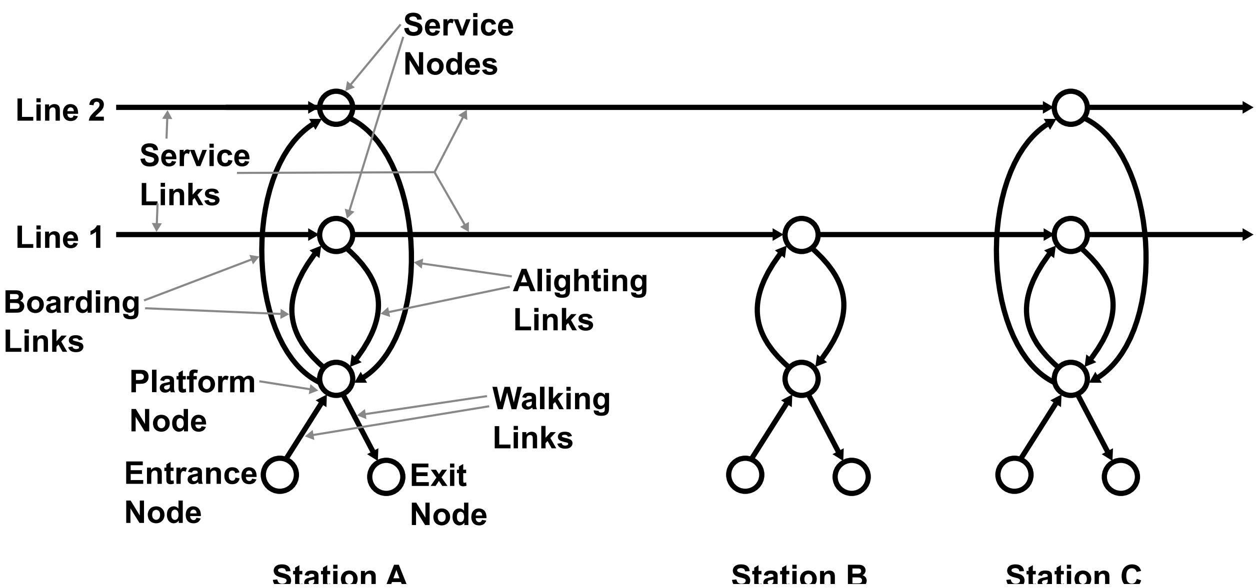


Figure 5-1: Example of line-based representation of public transport network

The *service subgraph* is the union of a set of graphs representing each of the network's lines. Each line, or service pattern, is represented by a series of nodes and links describing the stopping pattern of that service. The *service nodes* for a given line indicate which stops that line serves. The time or cost associated with a *service link* connecting a pair of service nodes indicates the time or cost of traveling on that line between the stops connected by that link. Each line also has certain attributes which can include the name of the line, frequency of service, the capacity of the vehicles on that line, the operating company, etc.

The *pedestrian subgraph* represents the physical aspects of the network that are used to access and exit the transport services. The nodes in this graph represent locations, including station entrances, exits, ticket halls, and platforms. The links in this graph represent pedestrian facilities (*e.g.* corridors, escalators, and stairs) between these locations.

The two subgraphs are connected by *boarding links* and *alighting links* at each stop. The time or cost of a boarding link includes the assumed or average waiting time for passengers boarding that service. In some cases it is assumed that passengers arrive at the service randomly and that the service runs at perfectly even intervals, implying that the average passenger waiting time is half the headway (the headway is the reciprocal of the frequency) (Spiess and Florian, 1989).

It should be noted that in this type of model, the attributes of each part of the network are assumed constant over a certain interval of time. It is possible for attributes to vary over time, but such variation is represented with a sequence of fixed time intervals with constant attributes. This is explained further later in this section.

Run-Based Supply Models

Nuzzolo and Crisalli (2004) present a good review of run-based supply models and their applications. The primary difference between these models and the line-based models is that run-based models represent points in time explicitly. The graphs for each line (service pattern) are decomposed into graphs for each run (scheduled service) on that line. The nodes in these run subgraphs indicate the time at which the run is scheduled to be at the given location. This representation demands that the nodes in the pedestrian subgraph have a temporal dimension as well. The OD estimation of this chapter does not require such a highly granular temporal representation, so this discussion is postponed until Chapter 6 which does have such requirements.

5.1.2 Paths and Path Choice Models

Path (or route) choice models describe how passengers identify, evaluate, and choose paths through the network from their origins to their destinations. The representation of paths in the network depends on the network itself. Line-based supply models facilitate *frequency-based* path representations while run-based supply models facilitate *schedule-based* path representations. While any path through a transport network has temporal components (*e.g.* the duration of a trip), the primary difference between these two types of paths is the degree to which *specific points in time*, rather than time intervals, are represented. This section describes frequency-based paths and path choice models. Schedule-based paths are discussed further in Chapter 6, along with run-based supply models.

Frequency-Based Paths

Nuzzolo (2003) reviews some of the literature on frequency-based path choice models. In this type of model, a path on the public transport network is evaluated with respect to the frequencies of the services used on that path, as well as other attributes. The primary relevance of frequency is that it is inversely proportional to waiting time (though the precise relationship between frequency and waiting time depends on certain aspects of passenger behavior). This framework accommodates walking links by representing them as services with infinite frequency (and thus zero waiting time).

Given a line-based representation of the network, frequency-based paths are typically identified via graph-based algorithms for finding the shortest path (or paths) according to some criterion (Prashker and Bekhor, 2004). Common algorithms for finding shortest paths are Bellman-Ford or Dijkstra (Bertsimas and Tsitsiklis, 1997). The total cost of a path is typically modeled as the weighted sum of the time (or cost) of its constituent links, and may be expressed in units of monetary cost or actual or weighted time. The cost of each link is determined by the attributes of that link (*e.g.* travel time, waiting time, and walking time) and a set of weights for those attributes that express user preferences. These weights can be considered relative monetary values of time (VOT) (cf Small and Verhoef, 2007, Chapter 2) for different components of a public transport trip. Wardman (2001) reviews a range of findings on travel time valuation.

Ben-Akiva et al. (1984) proposed the *path labeling* approach by which multiple different sets of weights are used to identify multiple different shortest paths. An example of such a labeled set of weights would be one that finds the path with the fewest interchanges.

Frequency-Based Hyperpaths

The type of model described above has (at least) one serious drawback. It assumes that passengers identify paths in which they select only a *single* line to board from a given stop or platform, whereas in reality they may wait at a boarding location and make decisions about how to travel depending on which line a vehicle comes first. This is referred to as the *common lines* problem. Spiess and Florian (1989) treat this problem by describing a more sophisticated strategy on the part of the passenger. They propose that passengers use *optimal strategies* to minimize the total expected travel cost. With such a strategy, passengers select, at each boarding location, an *attractive set* of lines which they are willing to board. Passengers then board, in a stochastic process, whichever of those lines departs first. Their probabilistic path through the network is thus a function of the optimal strategy.

Nguyen and Pallottino (1988) develop the same idea in a graph-theoretic context. They describe *hyperpaths*, the probabilistic superpositions of multiple paths through the network, which result from the application of the strategies described by Spiess and Florian (1989). The attributes of a hyperpath (*e.g.* travel time, wait time, total cost) are the respective expectations of the attributes of the constituent paths, weighted by the paths' relative frequencies. Both sets of authors describe an efficient algorithm with which to find the shortest (*i.e.* least expected cost) hyperpath from an origin to a destination.

This formulation and solution algorithm are extremely elegant, if not completely realistic, and the reader is encouraged to consult the original references for a more detailed presentations. Hyperpaths have found wide application in the field of public transport modeling. For example, they are implemented by two popular transport modeling software packages, EMME/2 and TransCAD.

Random Utility and Discrete Path Choice

Path Choice models describe how passengers choose between a set of identified paths (or hyperpaths). Prashker and Bekhor (2004) provide a detailed review, including simple simulations, of a wide range of path choice models. The simplest model for path choices is a deterministic one – that passengers choose only the single path with the least expected cost from their origin to their destination. Models that allow for probabilistic choice between *multiple* paths derive primarily from *random utility theory*, for example as described by Ben-Akiva and Lerman (1985).

According to random utility theory (in the context of path choices), the utility of each path for each passenger is assumed to be a deterministic quantity, and passengers are assumed to choose the single path with the highest utility (or least disutility, as the case may be). However, the utility of each path to each user cannot be measured directly, nor can the model account either for all attributes of each path or the preferences of each user. The utility of the k_{th} path to a given user is considered to be a random variable U_k, with a deterministic component V_k and a random component ε_k, related by the equation

$$U_k = V_k + \varepsilon_k. \tag{5-1}$$

For the case of paths through a public transport network, V_k is estimated as the weighted path time or cost as discussed above. Because each passenger chooses the path with the most utility (as he or she perceives it), the fraction of passengers that choose each alternative depends on the distribution of ε_k. The reader is referred to Prashker and Bekhor (2004) for details on the different path choice models (and extensions thereof) that result from different assumptions on ε_k. One of the most common assumptions, that ε_k are identically and independently distributed (iid) Gumbel variables, results in the *Multinomial Logit* (MNL) model. In the MNL, the probability of a passenger choosing path k (out of K alternatives) is given by the equation

$$\Pr(path_k) = \frac{e^{U_k}}{\sum_{i \in K} e^{U_i}}. \tag{5-2}$$

The MNL is easy to use, but suffers from at least one significant limitation for path choice modeling. The independence of irrelevant alternatives (IIA) property, which can be interpreted as a failure to account for similarities between alternatives, produces unrealistic path choice probabilities under some circumstances. This is a particular problem in the presence of multiple paths with overlapping segments. Prashker and Bekhor (2004) explore and discuss this issue in detail. Nevertheless, this model was used by Guo (2008) to model path choices of passengers in the London Underground, a network with multiple possible (and overlapping) paths for many of its OD flows.

5.1.3 Assignment Models

Assignment models, narrowly defined, depend on supply and path choice models to estimate the utilization of each portion of the network, given demand. Line-based supply models and frequency-based path choice models result in *frequency-based assignment*, as discussed in this section. Run-based supply models and schedule-based path choice models result in *schedule-based assignment*, as discussed in Chapter 6. In the former, the results of the assignment are estimated at the level of a certain piece of the network over a certain interval of time. In the latter, results are reported for individual trips or vehicles at different points in time.

Aside from this distinction, Nuzzolo (2003) categorizes assignment models along the following dimensions which are often used to characterize assignment models.

Deterministic vs. Stochastic Path Choice Models – as discussed above, whether passengers are assumed to choose paths deterministically, according to observed path attributes, or probabilistically depending on observed and unobserved path attributes and personal preferences.

Static vs. Dynamic – the granularity with which time is represented. Static models assume all attributes of the network and levels of demand and utilization are constant over the period of interest. In dynamic models, the period of interest is broken up into smaller intervals, between which the attributes of the network and levels of demand and utilization can vary.

Congested vs. Uncongested Network – whether levels of demand and utilization affect the performance of the supply. In a model which assumes no congestion effects, the link costs in the network (*e.g.* walking times, waiting times, travel times) are independent of the number of passengers assigned to each link. In a model which allows for congestion effects, the link costs are a function of the assigned volume. This can result in internally inconsistent results where, for a given link, the number of assigned passengers does not correspond to the link cost used in the assignment. As a result, it is typically necessary to take an iterative approach to finding a point of *equilibrium* where link costs, path choices, and assigned volumes are internally consistent.

Congestion, Capacity, and Equilibrium

Clearly, many different types of assignment models are possible. Even for a selected point in the dimensions described above – for example a static frequency-based model with deterministic path choices on a congested network – there are many possible ways to model the causes and effects of congestion. One of the causes of congestion in public transport networks as perceived by passengers is vehicle capacity. When vehicles are full, passengers waiting to board will be delayed.

Spiess and Florian (1989) and Nguyen and Pallottino (1988) describe early approaches to finding equilibrium in a frequency-based assignment model with deterministically selected hyperpaths when link costs are monotonic functions of assigned volume. The literature in this area since their work is extensive; a few examples are presented here. Cominetti and Correa (2001) use a simplified bulk queueing model to model congestion effects in a frequency-based assignment model, and find the resulting equilibrium. Kurauchi et al. (2003), Bell and Schmocker (2004), and Cepeda et al. (2006) present different ways to model the capacity constraints of public transport vehicles, for example using absorbing Markov chains, and the resulting congestion effects. Schmocker et al. (2008) extend some of this work into a quasi-dynamic context that is more granular in the time dimension. Schmocker and Bell (2009) use these models to analyze congestion on the London Underground.

5.2 OD Matrix Estimation Literature Review

The literature on OD estimation methods is quite rich. Cascetta and Nguyen (1988) provided an early synthesis and survey of the field and Abrahamsson (1998) provides a somewhat more recent and thorough review. This section is not intended to be a complete review of the literature, but rather to discuss the various approaches considered and/or tested for this particular application to the Overground.

Many methods for OD estimation are expressed in terms of finding the "best" OD matrix that corresponds, via an assignment model, to the given link flows. Most approaches reviewed here have been formulated as mathematical optimization problems. The difference between the various formulations rests in how "best" is defined (be it in isolation or with respect to the prior estimate or seed matrix) and whether the given link flows are viewed as deterministic constraints or as approximate targets to be matched as closely as possible.

In this discussion, the term 'link' has very general meaning. Specifically, it refers to any element of a transport network than can be modeled as an edge in a graph-based model. This includes boarding and alighting of public transport services as well as walking or riding between points in the network, as in the assignment models described above. It also includes entry into or exit from the system at certain points, as a station gateline can be modeled as a link between points outside and inside the paid area of the station.

It should be noted that the differences between various OD estimation methods are largely orthogonal to the differences between the types of assignment models described in the previous section. For example, the same OD estimation method can be used to estimate static or dynamic OD matrices, depending on whether the associated assignment model is itself static or dynamic. The OD estimation methods depend on the outputs of *some* assignment model but are not specific with respect to *how* exactly that assignment model works. The caveat to this is that when the assignment model includes congestion and thus requires equilibrium, the same is required of the OD estimation method, as discussed later in this section.

Section 5.2.1 presents a trivial OD estimation example to develop intuition around the general problem of OD estimation. Section 5.2.2 reviews many of the methods that have been proposed to solve the problem. Section 5.2.3 discusses the results of some simple simulations that were conducted to test the properties of some of these methods.

5.2.1 OD Estimation Example

Figure 5-2 presents a trivial example of the OD estimation problem concerning a railway line with four stations, A through D, with service only in the A $\rightarrow$ D direction. The loads on the three links of this line, A $\rightarrow$ B, B $\rightarrow$ C, and C $\rightarrow$ D, are known to be 5, 10, and 5 passengers, respectively. In this network, the assignment model is trivial. This example does *not* include a seed matrix.

The figure illustrates two OD matrices that are feasible solutions to the OD estimation problem. The first solution is the trivial solution where the flow on each link corresponds to an OD flow of the same number of passengers traveling only on that link. This is the solution with the maximum number of passengers, 20, and the shortest trip lengths. The second solution minimizes the number of passengers, 10, by maximizing the trip lengths.

In this example, the first solution has twice as many passengers as the second. Any convex combination of these two solutions is in fact a solution to the stated problem and will have a total number of passengers between 10 and 20. Naturally, in any solution, the total number of boardings or alightings is equal to the total number of passengers, and the respective locations of some of the boardings and alightings will depend on the solution OD

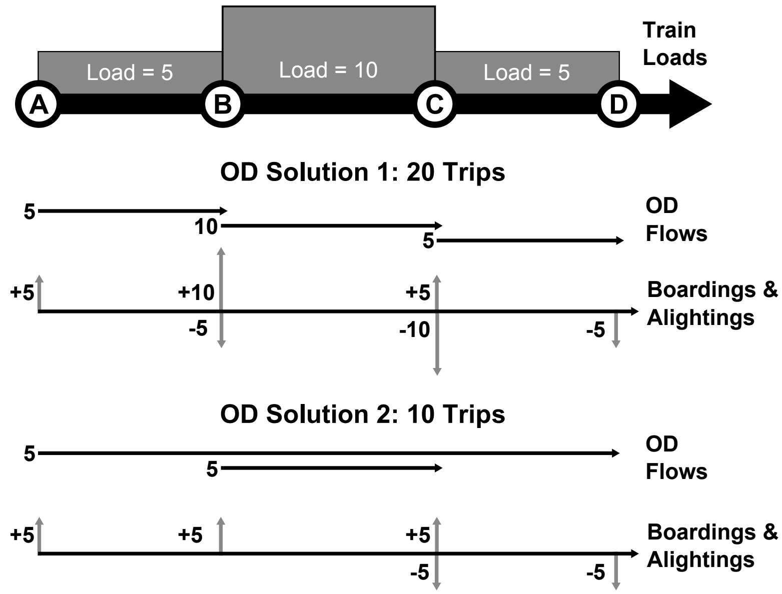


Figure 5-2: Example OD estimation problem

matrix. Any feasible solution will have 5 boardings at station A and 5 alightings at station D.

This example illustrates the following for the general OD estimation problem in which flows are given for an arbitrary set of links in the network.

- The OD matrix estimation process determines the estimate of the total number of passengers.
- The OD matrix estimate determines the estimate of boardings and alightings at each station.
- Generally, the given link flows do *not* uniquely determine the boardings and alightings. Terminals present a special case, where the link flow departing a terminal determines the number of boardings, and the link flow arriving a terminal determines the number of alightings.
- A seed matrix that offers some information about the true OD matrix is important, because there are infinitely many OD estimates that can match the given link flows.

These general attributes are important for understanding much of the balance of this chapter.

5.2.2 OD Estimation Methods

Iterative Proportional Fitting

Iterative Proportional Fitting (IPF, also Furness or Fratar or Bregman's balancing method) is a simple and widely used iterative procedure for expanding or adjusting a seed matrix to exactly fit one or more sets of marginal totals, *e.g.* total entries and exits at each station (Lamond and Stuart, 1981; Ben-Akiva et al., 1985; Ben-Akiva, 1987). These marginal totals are "complete" in that every OD flow contributes its full volume to exactly one element of each set of (entry and exit) marginal totals. Thus, this method takes the *total* volume of demand (*i.e.* sum of the OD matrix) as known, equal to the sum of one of the marginals (*e.g.* sum of station entries).

The IPF method has been used for the last decade or so by the London Underground (Maunder, 2003) to estimate dynamic OD matrices at the 15-minute level by scaling mail-back passenger surveys to meet marginal totals of entries and exits (estimated from gateline and manual counts) and age, trip purpose, and ticket type (from a separate in-person survey). Recent applications have used IPF to estimate time period level OD matrices from AFC-based seed matrices and entry/exit totals for urban rail networks. Specifically, Wilson et al. (2008) and Chan (2007) for the London Underground and Zhao et al. (2007) and Zhao (2004) for the Chicago Transit Authority rail network.

These applications of IPF can be seen as special cases of the general OD estimation problem, in which the set of links on which flows are given is *all* entry and/or exits to the network. In more general contexts, the measured link flows do not necessarily constitute complete marginal totals because a given OD flow can be assigned to multiple link flows, and so IPF is not applicable. Note that unlike the cited applications, the total volume of demand in the general case (*i.e.* given flows on an arbitrary set of links) is considered unknown and is *determined* by the OD estimation process.

Entropy Maximization and Generalized IPF

Van-Zuylen and Willumsen (1980) proposed to estimate an OD matrix by maximizing the "entropy" function (from information theory) of the prior (*i.e.* seed) and estimated OD matrices subject to the constraint that the estimated OD matrix, assigned to the network, reproduces the given link flows exactly. In this formulation, a given OD flow can be assigned to the network such that it contributes probabilistically to multiple link flows. Despite the fact that the Entropy Maximization (EM) method is formulated as a complex non-linear optimization problem, it can be easily solved by a simple iterative technique (Lam and Lo, 1991). On the other hand, it is possible for inconsistencies in the link flows to render this problem infeasible. This method was used by Wong and Tong (1998) to estimate time-dependent OD matrices (with a schedule-based assignment model) for the Hong Kong MTR metro system with high accuracy given a high-quality prior OD matrix.

The simple IPF method discussed previously is known to minimize a certain error function of the prior and estimated OD matrices subject to the constraints of the known marginal totals (Ben-Akiva et al., 1985). The IPF error function is simply the negative of the entropy function maximized by Van-Zuylen and Willumsen (1980). In addition, a generalized version of IPF (GIPF, also Generalized Iterative Scaling), which also allows for probabilistic assignment, has been studied by Darroch and Ratcliff (1972) and shown to maximize the same entropy function. In that sense, IPF is a special case of EM.

Paramahamsan (1999) and Maher (1987) have studied the EM method and found that, if the network structure and selection of link flows are such that the *total* volume of demand is known and constant (*e.g.* the case for which the IPF method is applicable), then uniform scaling of the seed matrix *will not* affect the estimates of the final OD matrix. In the general case, when total demand is not known, they find that uniform scaling of the seed matrix *will* affect the final OD estimates. These are very important properties to consider, in fact they should be cause for concern, when using a fractional seed matrix estimated from a AFC system used by only a sample of passengers (representative as that sample may be).

Information Minimization

Van-Zuylen and Willumsen (1980) also proposed to estimate an OD matrix by minimizing an "information" function (also from information theory) of the prior and estimated OD matrices subject to the flow constraints. The only difference between the Information Minimization (IM) and EM methods is the objective function in the resulting optimization problem. They also show that IM can be derived by applying EM to the link flow volumes themselves rather than to the OD flows directly. The IM formulation can be solved via the same algorithm developed for EM with one slight modification (Lam and Lo, 1991). It is susceptible to the same link flow inconsistency issues as EM.

Paramahamsan (1999) and Maher (1987) also studied IM and showed that, unlike EM, the final OD matrix estimate is *not* affected by uniform scaling of the seed matrix. This is a desirable property when using a fractional AFC-based seed matrix and a set of link flows such that total demand is not held constant.

Lam and Lo (1991) compared the performance of IM and EM using "complete" information on origin, destination, and path choice from over 13,000 roadside surveys of drivers in the AM and PM peak periods on a road network with 23 zones and 328 links in Shenzen, China. They used both methods to estimate the AM peak OD matrix in a range of scenarios defined by (i) varying the number of links with given flows (from one to forty), (ii) assigning based on all-or-nothing or observed link choice proportions, and (iii) using the PM peak matrix as the prior OD estimate, using the transpose of the PM peak matrix as the prior, or using no prior. Their primary conclusions were as follows.

- The better the prior OD estimate and the more link flows given, the better the results under IM and EM.
- IM outperformed EM with a good prior OD estimate, but lacking a prior estimate EM outperformed IM.
- The simplistic all-or-nothing assignment model did not introduce substantial additional error into the results, especially in the presence of a good prior OD estimate.
- Both methods underestimated total demand but by less than 5% and 7% in all IM and EM scenarios, respectively. With a good prior OD estimate and the observed path choice proportions this underestimate was less than 2% for both methods.

(Constrained) Generalized Least Squares

Cascetta (1984) proposed to estimate OD matrices using the familiar framework of Least Squares (LS) regression (Greene, 2007) by taking advantage of the fact that, for a given assignment, the link flows are a linear combination of the estimated OD flows. The use of Generalized Least Squares (GLS) is also proposed, to account for variance and covariance of the model's error terms. The primary benefits of this approach are that (i) it allows the estimated OD matrix to match the given link flows only approximately, (ii) it has known statistical properties, and (iii) it can account for varying statistical accuracy of the *inputs*. Bell (1991) extends this approach to include (necessary) non-negativity constraints on the OD estimates.

Unlike traditional linear regressions, this approach does not allow an intercept (constant) term in the model specification. Because linear regressions require error terms to be of zero mean, the lack of an intercept implies that the seed matrix is an unbiased estimate of the true matrix. This is *not* appropriate in the context of an AFC-based fractional seed matrix that is known to be a lower bound on the actual OD matrix. It should be possible to extend the GLS approach to deal with this bias, for example by adding constant terms to the model, but this confuses the interpretation of the model results and is beyond the scope of this work.

Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a very general and powerful type of statistical estimation with many desirable statistical properties (Greene, 2007). It is used, for example, to estimate discrete mode-choice models from household transportation surveys (Ben-Akiva and Lerman, 1985). In the context of OD estimation, the benefits of MLE are that it allows for flexibility in the types and accuracy of inputs, explicit assumptions about the degree and type of randomness in any of the data, and full flexibility of functional form.

For example, Hsu (1985), Ben-Akiva et al. (1985), and Ben-Akiva and Morikawa (1989) propose and test, on real-world data, a number of MLE formulations assuming Poisson distributed OD flows and different models of bias in the prior OD estimate. Their problems are similar to those solved by IPF in that they have a seed matrix and complete marginal counts at the entries and/or exits of the system. However, Ben-Akiva and Morikawa (1989) are able to explicitly estimate the bias in the prior OD estimate as a function of trip length as well as seasonal and trip-purpose characteristics, something impossible to do with IPF. On the other hand, some MLE formulations are equivalent to other methods listed here. Hsu (1985) proves that the Poisson MLE model with multiplicative bias parameters on all origins and destinations yields equivalent results to IPF. The GLS methods described above are known to be equivalent to MLE when it is assumed that all OD flows come from a multivariate normal distribution (Cascetta, 1984).

The advantages of MLE seem to be clearest when there is a strongly-held belief about the nature of the randomness in the problem or the functional form of the model (*e.g.* type of bias in prior OD estimate). However, the more complex the MLE model, the harder it can be to solve in real-world applications. For example, the complex MLE model proposed by Lo et al. (1996) can have a non-convex objective function, requiring the development of a custom algorithm that is still not guaranteed to find a global optimum solution (Lo et al., 1999). It is worth noting here that some of these same authors choose instead to use the much simpler Entropy Maximization method described above in subsequent work on complex real-world problems such as the estimation of combined OD matrices for Hong Kong's complex multi-modal public transport network (Wong et al., 2005).

Gradient Descent

Spiess (1990) takes an approach that is somewhat different from the others discussed here, in that it seeks only to minimize the difference between the given link flows and the link flows implied by the estimated OD matrix. It depends on a particular choice of optimization *algorithm* – in this case the *gradient method* (also known as *steepest descent*) – to change the prior OD matrix as little as possible. This method was designed with scalability in mind, as other methods apparently could not handle large real-world networks at the time. This has become substantially less of an issue with the expansion of computing power in the decades since.

The primary reason this method is discussed here is that it was implemented as part of the commercial transportation modeling software package EMME/2, updated versions of which are used by Transport for London for strategic modeling of London's public transportation network (Warner, 2010). Current versions of EMME/2 may have maintained the implementation of this OD estimation method, in which case it may be possible for Transport for London to estimate OD matrices without the purchase or development of any specialized software.

Multiple Path Matrix Estimation

Nielsen (1998) proposed a heuristic approach to estimating OD matrices, called Multiple Path Matrix Estimation (MPME), and used it for OD estimation of large urban traffic networks. It is similar to some of the other methods described here, but heuristic in the sense that it does not optimize any explicit objective function *per se*, and does not have well-defined statistical properties. More recently, Nielsen et al. (2008*a*) used MPME to estimate dynamic OD matrices of a large urban railway network in Copenhagen, Denmark, using complete train loadweigh data and existing survey-based OD matrices. This method is available as part of the commercial transport modeling software package TransCAD (Caliper Corporation, 2007).

The Total Demand Scale

The Total Demand Scale (TDS) is not an OD estimation method *per se*, but rather a measure proposed by Bierlaire (2002) to characterize the ambiguity inherent in a specific OD estimation problem instance. In short, it is proposed to find two OD estimates with the minimum and maximum *total* demand, respectively, that still satisfy the flow constraints (and of course have only non-negative values). The range between the minimum and the maximum provides an indication of how much the total estimated demand can vary, and thus can provide insight into the overall effect of the prior OD estimate. This measure is easily estimated using well-known linear programming (*i.e.* linear optimization) algorithms.

OD Estimation Under Congestion and Equilibrium Models

The formulations for the OD estimation methodologies reviewed here are generally expressed in terms of fixed path or link probabilities for each OD flow. When congestion and capacity constraints are considered using equilibrium assignment models, this approach can yield internally inconsistent results.

The widely accepted approach to OD estimation under equilibrium assignment, for example as proposed by Nielsen (1998) and Cascetta and Postorino (2001), is to use an iterative bi-level estimation process. In such a process, each iteration consists of two steps. First, the OD estimate is held fixed and the equilibrium path or link probabilities are estimated using the assignment model. Second, the path or link probabilities are held fixed and the OD matrix is estimated using one of the above methods. Such iterations are repeated until the OD estimate and path or link probabilities converge.

5.2.3 Simulation of OD Estimation Methodologies

A number of these OD estimation methods, including Entropy Maximization, Information Minimization, Least Squares, and Gradient Descent, were tested using simulations of a very simple network. The narrow goal of these simulations was to test the sensitivity of each method to different Oyster penetration rates.

In each simulation, a 'true' OD matrix is randomly generated and 'true' link flows are derived from the true OD matrix and the network structure. Next, an Oyster penetration rate is randomly generated for each OD flow. The 'Oyster' seed matrix is generated by multiplying each OD flow in the true OD matrix by the respective simulated Oyster penetration rate. Each OD estimation method is then used to estimate an OD matrix from the true link flows and the sampled seed matrix, and the results are compared with the true OD matrix.

In addition, each OD estimation method was tested on a seed matrix uniformly scaled in two different ways. First, the seed matrix was uniformly scaled so that the total number of trips was equal to the total number of trips in the true OD matrix. This is intended to simulate a situation in which, somehow, the total volume of trips is known *a priori* and this knowledge is used to adjust the seed matrix *before* the OD estimation process. Second, the seed matrix was uniformly scaled so that the sum of the seed matrix is one (*i.e.* it is reduced to a pure multinomial probability distribution over the OD flows). In this context, the seed matrix contains information *only* about the distribution of trips in the network, and not the volume of trips.

The finding from these simulations is that, as expected, the Information Minimization method is indeed unaffected by uniform scaling in the seed matrix. That is, it produced the same result regardless of whether the seed matrix was scaled up, down, or not at all. All other methods tested were sensitive to such scalings. Their performance generally improved when the seed matrix was scaled up to match the total volume in the true OD matrix, and degraded when the seed matrix was scaled down to a probability distribution.

5.3 OD Estimation Strategy for the London Overground

The many available assignment models and OD estimation methods cover a wide spectrum of complexity and sophistication. The strategy developed for the purposes of this application is quite simple, primarily as a result of the following circumstances which are described in further detail throughout the balance of this section.

- The required time scale for OD estimation is the time period level. For example, the three hour AM Peak from 07:00 to 10:00.

- High confidence is expected to be placed in the measured link flows, be they from loadweigh data or manual counts.

- The shape of the Overground network is such that for any trip *assumed* to have used the Overground, there is never more than a single reasonable path through this network.

- It is assumed that for most passengers, the choice of path (with respect to using the Overground or not) is largely insensitive to different (but reasonable) assumptions about how congestion and passenger preferences affect the choice among path alternatives. This assumption likely has some truth to it, but is adopted primarily as an engineering simplification.

These simplifications and other more specific derived assumptions allow the development of the assignment model and OD estimation methodology described in the following sections. A subsequent sensitivity analysis tests how sensitive the assignment model is to violation of most of these assumptions.

5.3.1 Assignment Model for the London Overground

This section describes in detail all aspects of the proposed assignment model. In the language of assignment models developed above, the proposed assignment model is a static frequency-based model that does not account for congestion and addresses the common lines in a limited way without the traditional hyperpath-based approach. The purpose of the model is to relate OD flows to link flows to support the OD estimation process. This model is for the most part derived from specific assumptions about passenger behavior, but is also a product of certain engineering simplifications.

Congestion and Capacity

It is assumed that for most passengers, the choice of path (with respect to using the London Overground or not) is insensitive to the effects of congestion, including vehicle capacity constraints. From the perspective of Overground managers, this is a conservative assumption because the likely effect of congestion would be to divert passengers from the highly-congested London Underground to the less congested Overground.

Static and Line/Frequency Based

Because OD matrices are desired at the time period level, this work takes a static approach to line- and frequency-based assignment as reviewed by Nuzzolo (2003). That is, the network is described in terms of infrastructure (*e.g.* stations, platforms, etc) that is connected by public transport lines running in fixed service patterns at specific frequencies. Nguyen and Pallottino (1988) and Nuzzolo (2003) provide extensive detail on this type of representation, including the specific graph structures to use.

The model is static in the sense that service and demand levels are assumed constant over the entire period of interest. This is reasonable in terms of service levels because the time periods over which OD matrices will be estimated are the same time periods used for scheduling, and reasonable in terms of demand because of the assumption regarding congestion and capacity.

Passenger Arrival Behavior and Waiting Time

It is assumed that the times of passenger arrival at stations are independent of any published timetable at headways (intervals between services) of ten minutes or less, but are influenced by the timetable at longer headways. This assumption is operationalized by the expression for average waiting time W as a function of headway H and a threshold T shown in Figure 5-3 and Equation (5-3). The "random incidence threshold" T is the headway below which passengers arrive randomly and thus should wait on average half the headway, assuming even headways[1] (Osuna and Newell, 1972). Beyond this threshold, behavior changes but no assumption is made regarding the change in behavior other than to say that the average waiting time will be as shown in the second case of Equation (5-3). This function was used by Casello and Hellinga (2008), with $T = 10$.

$$W(H) = \begin{cases} \frac{H}{2} & , H \leq T \\ T - \frac{T}{2}e^{\left(1-\frac{H}{T}\right)} & , H > T \end{cases} \tag{5-3}$$

It is a property of this function, illustrated in Figure 5-3, that the upper bound on waiting time is the value of the random incidence threshold, *i.e.* $W \in (0, T), \forall\, H > 0$. This function is adopted as an engineering simplification; more complex models exist and could be used. This simplification is thought to be reasonable because, as discussed in Chapter 6, the average waiting time (with respect to the timetable) on the London Overground was found to be approximately 10 minutes or less over the entire network at all times of day.

[1] See Chapter 6 for a detailed discussion of the relationship between headways, passenger behavior, and waiting time.

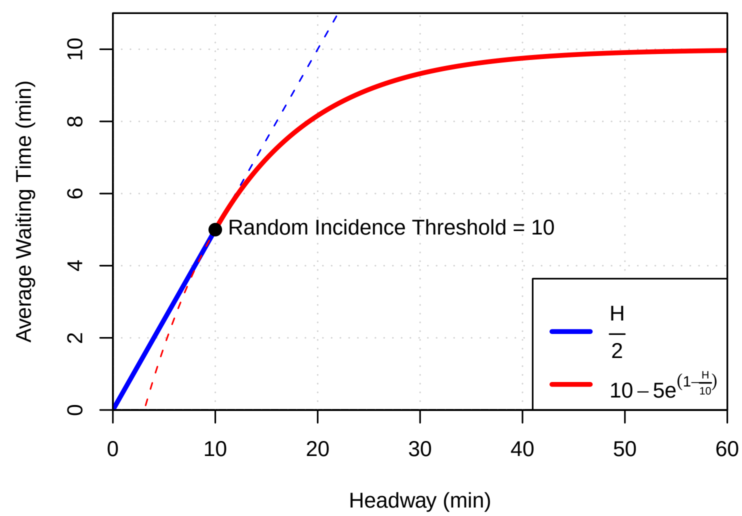


Figure 5-3: Waiting time as a function of headway

Indifference between Interavailable Services

It is assumed that passengers are indifferent between distinct but interavailable services (*i.e.* common lines) with identical stopping patterns on a given corridor. For example, on a line with a trunk segment and two branches, passengers traveling on the trunk are indifferent between services bound for either branch. Also, on a corridor of consecutive stations all served by multiple providers, passengers are indifferent between the different services.

This is a simplified version of the *optimal strategies* assumption made by Spiess and Florian (1989). It allows the use of a straightforward augmentation of the network structure to faithfully represent this indifference without depending on formal hyperpaths and related algorithms. The crux of this augmentation is the following. Between each pair of consecutive stations, a new "composite service link" is added for each possible combination of services that connect those two stations. Each link represents the superposition of a *possible* set of services that a passenger *could* decide to use to travel between the pair of stations. These links are analogous to the "attractive set" in the model of Spiess and Florian (1989).

The service frequencies, running times, and fractional assignment of passengers for these new composite service links are determined according to the model of Spiess and Florian (1989). The service frequency for each new link is simply the sum of the frequency of the combined services. It is assumed that the headway of the combined services is even and equal to the inverse of the combined frequency.[2] Since it is assumed that passengers are selecting between these services randomly, it is also assumed that passengers arrive randomly and so always experience a waiting time of half the combined headway. The running time for each new link is the frequency-weighted average running time of the services making up that link. When a number of passengers is assigned to one of these composite service links, the fraction of passengers assigned to the individual service links is determined by the respective

[2] This assumption, that the headways of the combined services is even, is coarse but, nevertheless, a common feature of frequency-based models.

frequency shares.

This augmentation is illustrated in Figure 5-4. In this example, the London Underground (LU) and London Overground (LO) both serve the link between North Wembley and Willesden Junction with a single service pattern each, with equal running times on this link. The Underground service is 6 tph (a 10 minute headway), so according to the above waiting time model passengers experience an average waiting time of 5 minutes. The LO service is 3 tph (a 20 minute headway), with an average waiting time of 8.1 minutes (as per the model described above). This results in a single composite service link with a combined frequency of 9 tph, a constant headway of 6.66 minutes, and an average waiting time of 3.33 minutes.[3] 33.3% of passengers assigned to this link will be assigned to the Overground service, with the balance going to the Underground service.

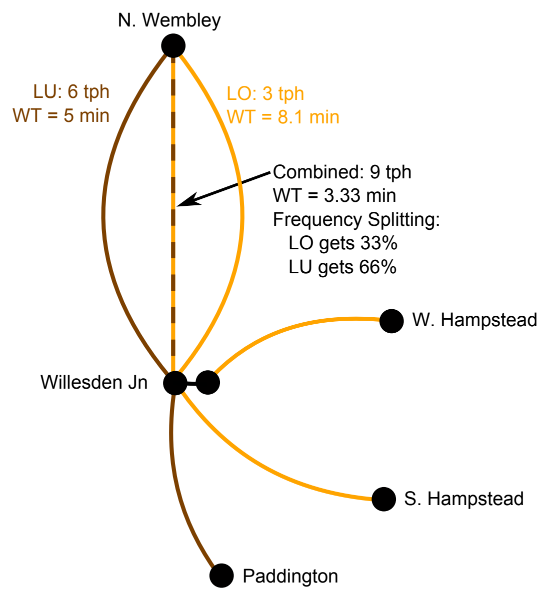


Figure 5-4: Illustration of network augmentation

This example also illustrates how this model reasonably captures overall passenger behavior, depending on passengers' eventual destinations. For passengers traveling from North Wembley to Paddington, the shortest path will not use the new composite service link because to do so would necessitate an additional waiting and boarding at Willesden Junction to continue on to Paddington. These passengers are better off on a path containing only the Underground-only link. Likewise for passengers traveling to South Hampstead, for whom the shortest path would include the Overground-only link.

Passenger Preferences and Path Choice

It is assumed that for most passengers, the choice of path (with respect to using the London Overground or not) is largely insensitive to different (but reasonable) assumptions about

[3] It is in fact impossible to combine two even headway services of 6tph and 3tph, respectively, into a single 9tph even headway service. This is an example of the types of approximations that occur in frequency-based models.

how passengers choose among alternative paths. Nevertheless, for each OD flow, different models for passenger preferences are used to identify up to three possible lowest-cost paths and then apply a discrete choice model to predict the fraction of passengers choosing each path. These methods are reviewed extensively by Prashker and Bekhor (2004) and Guo (2008).

In this model, the cost of a path is equated with the generalized journey time of that path. The *generalized journey time* (GJT) of a path is a weighted sum of the time of its constituent links. The GJT weights on each type of link (*e.g.* walking, boarding, riding a service) express passenger preferences. A range of different weighting schemes and values are used within TfL for different purposes (Transport for London, 2009*a*; Guo, 2008; London Transport, 1999). The following simple set of GJT weights has been selected, as prescribed for analysis of the London Underground by the TfL's *Business Case Development Manual* (BCDM) (Transport for London, 2009*a*). These weights reflect the information available in the RODS network representation and retain sufficient detail for the purposes described here.

$\beta_{InVehicle}$ = GJT weight for time spent riding (or dwelling) inside a transit vehicle.

β_{Wait} = GJT weight for time spend waiting outside a transit vehicle.

$\beta_{Interchange}$ = GJT weight (in minutes) for number of interchanges.

β_{Walk} = GJT weight for time spent walking.

Given a set of weights, the single lowest-GJT path is found using the Hao-Kocur shortest path algorithm (Hao and Kocur, 1992), but other algorithms such as Bellman-Ford or Dijkstra's (cf Bertsimas and Tsitsiklis, 1997) would work just as well. To identify multiple alternative paths, the *path labeling* approach is adopted, in which the GJT weights are modified and the shortest respective paths are found (Ben-Akiva et al., 1984). Three labels and respective sets of weights are used – one according to the BCDM, one to minimize travel time, and one to minimize interchanges – as shown in Table 5-1.

Label	$\beta_{InVehicle}$	β_{Wait}	$\beta_{Interchange}$	β_{Walk}
BCDM	1.0	2.5	3.5	2.0
MinTime	1.0	1.0	0.0	1.0
MinInterchange	1.0	1.0	1,000	1.0

Table 5-1: Labels and weights for identification of alternative paths

For each OD flow, the shortest path is found under the application of each label's weights. Duplicates are then removed, where two paths are duplicates if they visit the same stations in the same order. For each remaining path, the GJT is calculated using the weights of the BCDM label. The choice probabilities for each path are estimated using the simple Multinomial Logit (MNL) model, which was used by Guo (2008) to analyze path choices on the London Underground. According to the MNL model, for a given OD flow,

$$\Pr(path_k) = \frac{e^{GJT_k}}{\sum_{i \in K} e^{GJT_i}} \tag{5-4}$$

where K is the set of available paths and GJT_k is the GJT of the k^{th} path.

Finally, any path with probability less than a certain "path probability threshold," 10%, is dropped and all path probabilities are re-normalized to sum to 100%. This is done to prevent OD flows with only a very small likelihood of using the Overground from factoring into the subsequent OD estimation process, since that process will be based on link flows measured only on the Overground. This is effectively a convenient heuristic that attempts to eliminate spurious artifacts of the above modeling assumptions.

Operator Aggregation

One feature of the RODS network representation is that each service line is assigned a specific *operator code*. One operator code is assigned for the entire London Overground network. These operator codes are used in the following two types of aggregation in the assignment model. Both of these aggregations happen *after* the assignment process.

Operator links are the aggregation of all services by a given operator between a given pair of adjacent stations. This operation is designed so that outputs of the assignment model correspond to the given link flows, which will be used at the same level of aggregation in the static time period level OD estimation context. For example, the London Overground and the London Underground both provide service connecting the adjacent Richmond and Gunnersbury stations. Under the described aggregation, the assignment model produces results for two different operator links between these two stations. Only one of these links, that of the Overground operator, will have corresponding link flows in the OD estimation process (given the data available).

Operator clamps are a somewhat more subtle concept. The Oyster system allows passengers to interchange between services (of the same operator or different operators) without conducting additional validation. OD flows in this work are defined by pairs of Oyster-enabled stations, most of which contain at least one station outside the Overground network. Since the goal is to estimate an OD matrix for a single operator, the assignment model must provide a means by which to map end-to-end OD flows (*i.e.* first and last points of Oyster transactions) to the OD flow on a given operator's network (*i.e.* first and last stations at which that operator's services were used).

This process, referred to as "clamping" of OD flows, provides a "clamp" which describes the "clamped" or "inner" OD flow entirely on the given operator's network that would be used by passengers traveling on the "unclamped" or "outer" end-to-end OD flow. For example, the Overground clamp for the outer OD flow of Leyton (on the Underground Central Line) to Camden Road (on the Overground North London Line) is Stratford (where the two lines meet) to Camden Road, with a share of 100%. This means that all passengers traveling from Leyton to Camden Road are predicted to interchange at Stratford to the North London Line, and will be counted as part of the Stratford to Camden Road Overground-only OD flow.

The clamped OD flow can be smaller than the unclamped OD flow if that clamp is along a path predicted to be used by only a portion of passengers. For example, all passengers from Wembley Central (on both the London Underground's Bakerloo Line and Overground's Watford DC Line) to Oxford Circus (on the Bakerloo Line) are predicted to travel via Queen's Park (also on both lines). However, because the Bakerloo Line frequency is substantially higher between Queen's Park and Oxford Circus than between Wembley Central and Oxford Circus, all passengers are predicted to be indifferent between the Bakerloo Line and the Watford DC Line at Wembley Central. Those who, by chance, take the Watford DC Line will transfer at Queen's Park to the Bakerloo Line to continue their journeys. Since the frequency at Wembley Central of the Bakerloo Line is twice the frequency of the Watford DC Line, the Overground clamp for this OD flow is Wembley Central to Queen's Park with a 33% share. An algorithm for accomplishing this sort of computation is described in detail in Appendix D.

Out-of-Station Interchanges

The RODS network representation includes walking links between stations that are near each other but not directly connected. At the same time, the Oyster ticketing system allows selective free "out-of-station interchanges" (OSI's) wherein passengers validate when exiting one station in the OSI pair and then validate again when entering the other. For a number of reasons, these interchanges are not handled perfectly in the assignment model.

Unfortunately, the set of OSI's in the RODS network representation is not perfectly congruent with the set of OSI's in the Oyster system. Furthermore, the assignment model only records explicit entrances and exits, for a given OD flow, at the first station of entry and last station of exit. This will be a source of error in the assignment model, and should receive attention in the future.

Implementation and Outputs

The assignment model described here is implemented as a custom software package of nearly 3,200 lines of code developed in the widely-used free programming language Java (Sun Microsystems, 2009). This program, called *ODNet*, proceeds more or less as follows.

1. Read updated RODS network representation, including infrastructure and service patterns and frequencies.

2. Select service patterns and frequencies for a specific time period of interest (*e.g.* weekday AM Peak).

3. Modify network model to reflect separate physical and service networks, and add boarding and alighting links to connect the two.

4. Determine the presence of interavailability and augment the network with additional services and boarding and alighting links to reflect passenger indifference.

5. For each pair of Oyster-enabled stations, assign a single passenger on that OD flow to the network and output link and clamp flows for the London Overground operator.

It is worth discussing here the effects on the network size and algorithmic performance of the augmentation in step 3. The RODS network representation has 7,213 service links for the AM Peak period. The augmentation adds 22,453 common service links, effectively quadrupling the size of the graph representing the TfL rail network. This change has no perceptible effect on the performance of the assignment algorithms used in step 4.

The last step of this process is appropriate only because congestion effects and capacity constraints are not considered. Because each OD flow is treated separately and only a single passenger is assigned, the resulting link flows can be interpreted as link probabilities for a given passenger traveling on that OD flow. These link probabilities, for a given OD flow, depend on

- the paths through the network identified for that OD flow,
- the probabilities of each path,
- the London Overground links along each path,
- the relative combined frequency share of Overground services on each link.

Clamp flows are estimated as part of the same process, and likewise interpreted as clamp probabilities. These link and clamp probabilities, discussed in further detail in following sections, are the crucial input from the assignment model into the OD estimation process.

Sensitivity Analysis

A number of scenarios were used to test the sensitivity of the assignment model to the various assumptions described above. In each scenario, some aspect of the assignment model was modified and the link and clamp probabilities were re-calculated and saved. For the sake of this sensitivity analysis, two quantities are defined. The *competitive market* is the set of all OD flows for which the assignment model predicts a positive probability of using the London Overground. The *captured market* is the set of all OD flows of the competitive market after being clamped to the Overground network. The size of each of these markets is simply the sum of the respective market's OD flows.

Figure 5-5 shows the results of the sensitivity analysis. The competitive and clamped market sizes are based on a raw Oyster OD matrix estimated as the daily number of Oyster journeys on each OD flow departing during the AM Peak averaged over all weekdays of March, 2009. The *Baseline* scenario is as described in the preceding sections, using Weekday AM Peak service patterns and frequencies. The scenarios tested are as follows, where each is described with respect to the Baseline scenario.

- *LU Congested (10%)*: To simulate the increase in perceived travel time caused by congestion on the London Underground, running times on all London Underground services are increased by 10%. This is a poor approximation of congestion, since it assumes congestion does not vary spatially or temporally.
- *LU Congested (25%)*: Likewise, increased by 25%.
- *High Wait Cost*: $\beta_{Wait} = 5.0$.
- *RI Threshold = 5*: The random incidence threshold in the waiting time function is set to 5 minutes.

- *RI Threshold = 30*: The random incidence threshold in the waiting time function is set to 30 minutes.

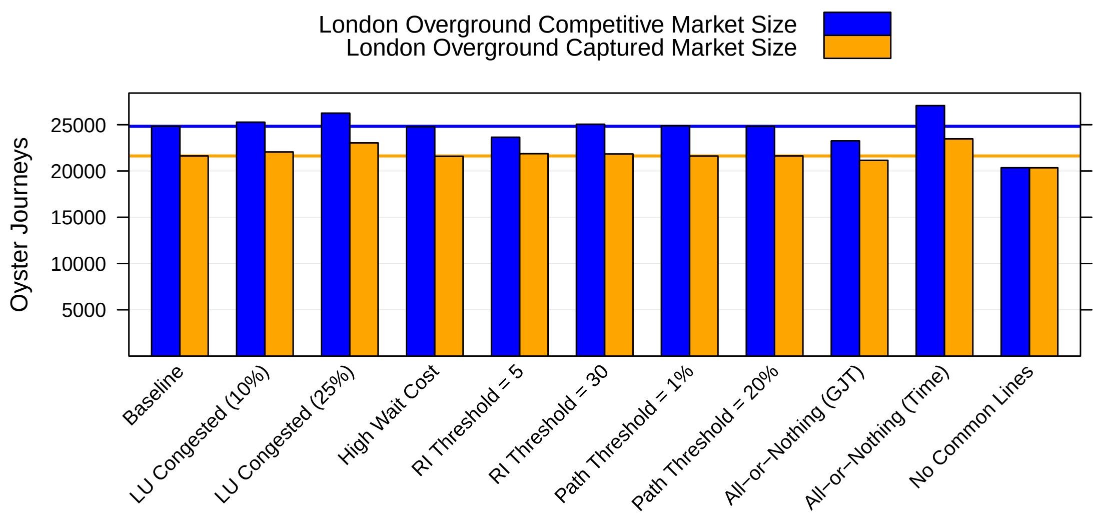
- *Path Threshold = 1%*: The path probability threshold is set to 1%.

- *Path Threshold = 20%*: The path probability threshold is set to 20%.

- *All-or-Nothing (GJT)*: A deterministic all-or-nothing (rather than stochastic MNL) assignment is used, taking the single lowest-GJT path for each OD flow.

- *All-or-Nothing (Time)*: A deterministic all-or-nothing (rather than stochastic MNL) assignment is used, taking the single lowest-travel time path for each OD flow.

- *No Common Lines*: The assumption of indifference to interavailable services is disregarded, and no common line links are generated.



Horizontal blue and orange lines show the size of the London Overground's competitive and captured markets, respectively, in the baseline scenario

Figure 5-5: Sensitivity of London Overground Oyster-only competitive market size and assigned trips

The assignment model does appear to be somewhat sensitive to the crudely simulated congestion on the London Underground, moreso at higher levels of simulated congestion. Even higher levels of simulated congestion would likely result in more drastic changes in the market sizes. That said, the simulation of congestion here is rather crude in that it assumes the congestion is spread uniformly across the entire London Underground network. In reality, the Underground is most congested in Central London, where it does not generally compete with the Overground.

The model appears to be generally insensitive to changes in GJT weights, the random incidence threshold, and the path probability threshold. This adds to confidence in the model because the variables tested in those scenarios are, in a certain sense, the most numerically arbitrary. The assignment model is somewhat sensitive to the two All-or-Nothing scenarios, but when passengers are assumed to take only the path that minimizes GJT, the captured market size does not change appreciably. This also adds to confidence in the model because it corresponds to modeling assumptions across TfL which hold that some weighted measure of GJT is a more realistic determinant of path decisions than is total travel time. Note that in the All-or-Nothing scenarios, the competitive market size and captured market size are still different because of the randomness associated with interavailability.

Probably most importantly, the drastic change in the No Common Lines scenario, in particular that the two markets are almost the exact same size, indicates the importance of the assumption of indifference to interavailable services. This adds to confidence in the model because it is believed by Overground management that this assumption is accurate for certain Overground corridors.

The overall results of this sensitivity analysis are generally coincident with the assumption that passengers using the Overground do not for the most part have reasonable alternatives, and gives confidence that the assignment model itself is reasonable. With this confidence, it becomes possible to use the formulation described in the next section to estimate OD matrices for the Overground.

5.3.2 Information Minimization Matrix Estimation

Of the many available approaches to the OD estimation problem, the Information Minimization (IM) formulation proposed by Van-Zuylen and Willumsen (1980) appears most appropriate for the circumstances faced by the London Overground. It is straightforward and simple to implement (given the outputs of the assignment model), and is a methodological cousin to the Entropy Maximization (EM) and IPF methodologies long used by the London Underground. As in Section 5.2, the term 'link' is used abstractly; it can refer to boarding, alighting, riding, walking in a public transport network as well as entry into or exit from the network.

In the Underground case, link flows are given at all entrances and exits to the network, and so total demand levels are fixed. In the Overground case, link flows are given primarily inside the network, and so total demand is a function of the OD estimation process. It has been found that, when total demand is not fixed, EM is sensitive to uniform scaling of the seed matrix (as will be the case when it is Oyster-based), but IM is not (Paramahamsan, 1999). Thus, IM is appropriate under the circumstances while being the OD estimation method most similar to those already used for similar problems within TfL.

The rest of this section shows the explicit mathematical IM formulation for OD matrix estimation proposed by Van-Zuylen and Willumsen (1980), and discusses several issues.

Formulation

Let

K = The set of links for which flows are available.

V_k = The observed flow on link k.

p_{ij}^k = The proportion of trips from i to j that use link k. This assumes the existence of a satisfactory assignment model which is used to generate p_{ij}^k.

t_{ij} = The measured flow from origin i to destination j in the seed OD matrix.

T_{ij} = The flow from origin i to destination j in the estimated OD matrix.

Using this notation, Van-Zuylen and Willumsen (1980) formulate the OD estimation problem as follows:

$$\min_{T_{ij}} \sum_{k \in K} \sum_{ij} T_{ij} p_{ij}^k \log \left(\frac{T_{ij} S^k}{V_k t_{ij}} \right) \tag{5-5}$$

where

$$S^k = \sum_{ij} p_{ij}^k t_{ij} \, , \forall k \in 1..K \tag{5-6}$$

subject to the flow constraints

$$V_k = \sum_{ij} p_{ij}^k T_{ij} \, , \forall k \in 1..K. \tag{5-7}$$

Through a Lagrangean analysis, they find that the OD flow estimates

$$T_{ij} = t_{ij} \prod_{k \in K} X_k^{\left(\frac{p_{ij}^k}{g_{ij}} \right)} \tag{5-8}$$

where

$$g_{ij} = \sum_{k \in K} p_{ij}^k \tag{5-9}$$

and X_k, a function of the known inputs and the Lagrangean multipliers, can be found through an efficient and simple iterative solution algorithm which is described in detail by Lam and Lo (1991). This algorithm, as well as the integration of the various data sources used in this chapter, was implemented in *R*, the free and open source statistical and graphical programming language and environment (The R Project, 2009). The convergence criteria for this algorithm were that Equation (5-7) be satisfied for each link to within ± 1% (*i.e.* approximately, rather than exactly).

Considerations

For this approach, the seed matrix should be an average of several days or weeks of Oyster data. When using the automatic loadweigh data, the link traffic counts should be the average of the loadweigh data over the same period of time as the Oyster data. When using the manual load counts, there is, unfortunately, only one measurement on each link, so no averaging is possible. By far the most complex aspect of this approach is the estimation of the p_{ij}^k from the assignment process.

The primary weaknesses of this approach, compared with the other approaches reviewed, are as follows.

- It treats the given link flows as deterministic constraints, which may not be ideal for counts coming from sources with small sample sizes (*i.e.* manual counts) and/or potential for significant measurement error.
- It does not explicitly account for the varying statistical quality among the inputs.
- It does not provide measures of statistical quality of the OD estimates.

Another limitation of this approach, common to all of the approaches reviewed, is that it does not ensure that the estimated OD flows are not smaller than the corresponding OD flows in the seed matrix. This is appropriate for the generic problem for which most of these approaches were formulated – *adjusting* prior OD matrices based on measured link flows. This application is somewhat different in that the Oyster seed matrix is considered a lower bound on the actual OD matrix, so it is possible that this bound will be violated by the IM estimation process. While it would be trivial to add this constraint to the *formulation*, such a constraint would likely invalidate the simple and efficient solution *algorithm* for which this method was selected. The addition of such a constraint should be the subject of further research.

5.4 Exploratory Analysis of OD Estimation Inputs

This section presents some brief exploratory analyses derived from the assignment model and the various available data sources.

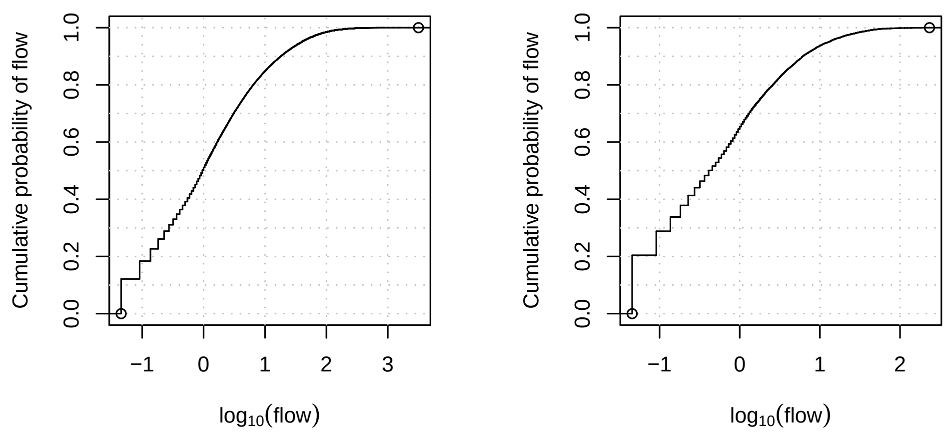
5.4.1 Oyster Seed Matrix

The seed OD matrix was estimated as the daily number of Oyster journeys on each OD flow departing during the AM Peak averaged over all 22 weekdays of March, 2009. This Oyster seed matrix has 733,087 total passengers over 85,444 non-zero OD flows, the distribution of which is shown in Figure 5.6(a). The smallest OD flows are 0.045 passengers (*i.e.* a single observed journey over the entire 22 days) and the largest is 3,160 passengers from Waterloo to Canary Wharf.

Of all the non-zero flows in this seed matrix, 8,742 of them have a positive probability of using the London Overground, according to the outputs of the assignment model. This matrix has 24,814 total passengers over these OD flows, the largest of which is 505 passengers from Stratford to Highbury & Islington. When clamped to the Overground network, the matrix reduces to 21,620 passengers over 1,763 flows between pairs of Overground stations. The distribution of unclamped flows in this matrix is shown in Figure 5.6(b). According to the assignment model, 85% of these flows and 78% of the passengers on these flows are, respectively, guaranteed to use Overground services at some point in the journey.

5.4.2 Link Flows

The manual counts conducted over March, 2009, give a single point estimate of "service link flows" – the number of passengers on board each individual scheduled London Overground



(a) Whole seed matrix

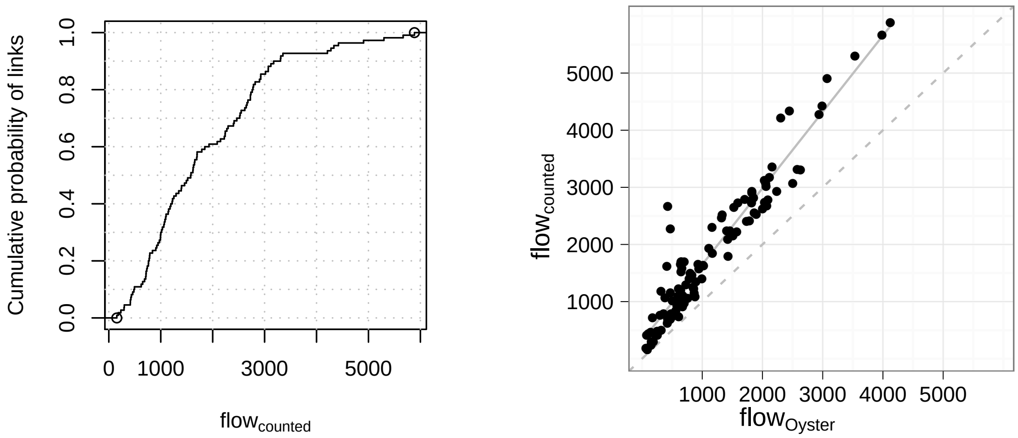
(b) Flows with positive probability of using the London Overground

Figure 5-6: Distribution of non-zero OD flows in Oyster AM Peak seed OD matrix

service on each link between consecutive stations (*e.g.* the 08:06 train from Stratford to Richmond, between Stratford and Hackney Wick). Consistent with general TfL practice, each scheduled service is classified into a specific time period (*e.g.* AM Peak) depending on the departure time of that service from its originating terminal. Aggregation of service link flows to the AM Peak level yields the link flows with which the OD matrix will be estimated. The distribution of aggregate counted AM Peak link flows, shown in Figure 5.7(a), ranges from 154 passengers (Gospel Oak to Upper Holloway) to 5,886 passengers (Canonbury to Highbury & Islington).

The assignment model, applied to the Oyster seed matrix, yields the "Oyster link flow" of each link as per Equation (5-6). Consistent with expectations, no Oyster link flow for the AM Peak is greater than the corresponding manual count. Figure 5.7(b) plots the counted link flows against the Oyster link flows. The two most apparent outlier link flows in this figure, with counted flows above 2,000 but Oyster flows below 1,000, are Clapham Junction to West Brompton ($\frac{flow_{Oyster}}{flow_{counted}} = 0.16$) and West Brompton to Kensington Olympia (0.21), the first two northbound links on the West London Line.

Table 5-2 aggregates the ratio of assigned Oyster flow to counted flow by line. Consistent with the two outliers mentioned, the assigned Oyster flow is substantially lower on the West London line than on other parts of the network. This is consistent with the expectation of a low rate of fully-validated Oyster journeys on the West London Line because of large numbers of non-Oyster interchange passengers using the West London Line at Clapham Junction (a major National Rail interchange).



(a) Counted link flows

(b) Counted link flow vs assigned Oyster link flow

Figure 5-7: Counted and assigned Oyster link flows for AM Peak, March 2009

Line	$\frac{\sum flow_{Oyster}}{\sum flow_{counted}}$
NLL	0.67
GOB	0.65
WAT	0.56
WLL	0.29

Table 5-2: Aggregate ratio of assigned Oyster flow to counted flow, by line

5.4.3 Gateline Data

Unfortunately, the gateline data it not so consistent. As of March 2009, 32 stations with London Overground services were gated, 11 of which are exclusive to the Overground. It is hoped that for these 11 stations, gatelines could provide automatic estimates of total passengers (*i.e.* Oyster and non-Oyster) entering and exiting the Overground network. This data is validated primarily by comparing the gateline entries and exits with the daily totals from the Oyster database, as shown in Figure 5-8.

At all these 11 stations, gateline entries are consistently greater than Oyster entries (with a few exceptions), which is as expected. The exceptions are 17 March at Homerton, 13 March - 25 March at Dalston Kingsland, and 31 March at West Hampstead, where gateline entrances drop off precipitously with no correlated change in Oyster data. On the other hand, gateline exits are consistently fewer than Oyster exits at Watford High Street and at Camden Road, where the gateline exit volumes are quite erratic. These discrepancies between Oyster and gateline data could be the result of a range of factors, including faults in the gatelines or station staff allowing passengers to validate their Oyster cards and then exit without using the gates.

As discussed above, the assignment model does not explicitly record, for a given OD flow, station entrances and exits during out-of-station interchanges (OSI's). Similarly, the Oyster

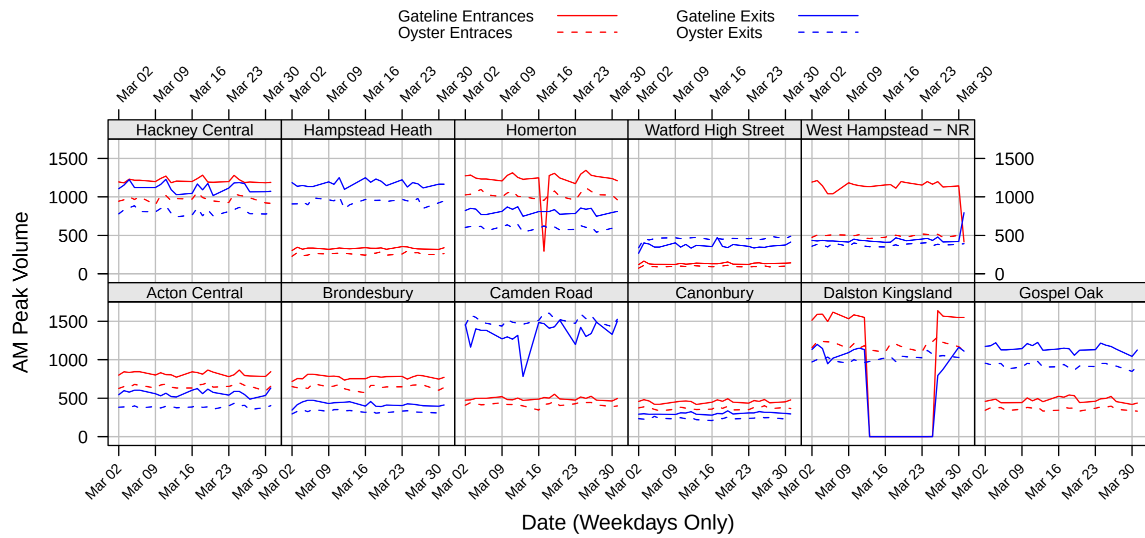


Figure 5-8: Oyster and gateline entry and exit counts at stations exclusive to the London Overground with recorded gateline data, March 2009 weekday AM Peak periods

data as provided indicates only the first station of entrance and last station of exit and not intermediate validations. This may or may not be the cause of the wide gap between gateline and Oyster entries at West Hampstead, where passengers can transfer from National Rail and from the Underground's Jubilee Line to the Overground's North London Line. These potential sources of error are set aside for now, but may be worth attention in the future, particularly if Oyster data is reported with intermediate validations.

Despite its obvious flaws, the following selective use of these gateline data should improve the OD estimation results. In this work, gateline data was averaged at the station level (for the AM Peak time period) to be used in the OD estimation process, with the following exceptions.

- Because there are two stations where all exit counts are apparently flawed, all exit counts were disregarded.

- The exceptions mentioned above, where gateline entrances are obviously but temporarily flawed, were manually removed from the data set.

If, in the future, gateline data is judged to be of higher quality, it can perhaps be used without such intervention.

5.5 OD Estimation Results

An AM Peak OD matrix for the London Overground was estimated from an Oyster seed matrix, manually-counted on-board link flows, and automatic gateline entry flows, using the Information Minimization formulation. The resultant OD matrix reproduces all given link flows (on-board and gateline) to within $\pm$ 1%. To show the effects of using selected gateline data, the OD matrix was also estimated without gateline entries.

Table 5-3 shows summary statistics for the two resulting OD estimates. The total number of Overground AM Peak passengers (*i.e.* the sum of the clamped OD matrix) is slightly smaller when estimated with gateline entry flows, but in both cases is between 37,000 and 38,000 spread over 1,763 different origin-destination combinations. This represents an overall expansion of between 71% and 75% over the Oyster-only seed matrix of 21,620 Overground passengers. The Total Demand Scale, the minimum and maximum *possible* sums of the clamped OD matrix such that the given link flows are reproduced, is 25,202 to 150,559. This indicates that the Oyster-based seed matrix plays a significant role in determining the total estimated number of Overground passengers.

Source of Link Flows	Sum of Clamped Estimated Matrix	Expansion Over Seed Matrix	Number of Flows > 0	Number of Flows $\geq$ 1.0
Onboard Link Counts	37,731	74.5%	1,763	1,245
+ Gateline Entries	37,124	71.7%	1,763	1,238

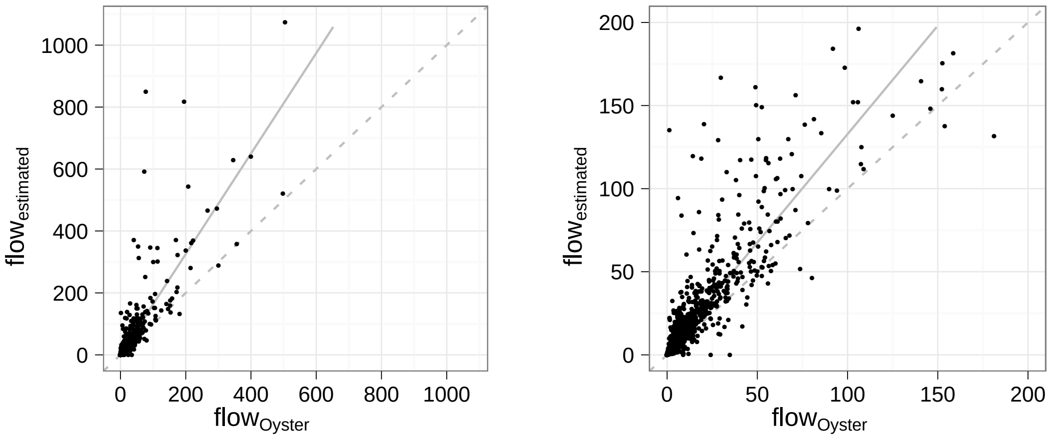
Table 5-3: Summary statistics for London Overground estimated AM Peak OD matrix

Figure 5.9(a) plots OD flows estimated with gateline data, clamped to the Overground network, against the respective flow in the Oyster-based seed matrix. Points above and to the left of the dashed line of unit slope are flows that were expanded under the estimation process. Table 5-4 identifies some of the largest estimated flows that expanded the most in absolute and/or relative terms. Consistent with expectations, these are flows from Overground terminals, primarily Stratford and Clapham Junction, that are large shared facilities with interchanges to other rail services. It is worth noting that because the origin stations of these OD flows are terminals, the estimated number of boardings at these stations will, by the nature of the IM estimation process, match the given values almost exactly (*i.e.* to within the convergence criteria).

		Clamped OD Flow			Alightings at Destination		
Origin	Destination	Oyster	Estimated	Expansion	Counted	Estimated	Error
Strat.	High. & Isl.	505	1,073	568 (112%)	3,555	3,184	-371 (-10%)
Strat.	Camd. Rd.	346	629	283 (82%)	1,778	1,799	-21 (-1%)
Clap. Jn.	W. Brom.	78	849	771 (988%)	1,115	1,560	445 (40%)
Clap. Jn.	Shep. Bush	195	818	623 (319%)	1,193	1,144	49 (4%)
Clap. Jn.	Kens. Oly.	73	592	519 (711%)	1,053	1,014	-39 (-4%)
Rich.	Gunn.	54	350	296 (548%)	1,211	1,290	79 (7%)

Table 5-4: Selected OD flows with large estimated values and large relative and/or absolute expansions

The expansion of the flows from Clapham Junction to points on the West London Line are especially large in relative terms – 988% to West Brompton, 711% to Kensington Olympia, and 319% to Shepherd's Bush. These are the first, second, and third stops (out of four) on the West London Line from Clapham Junction to Willesden Junction. This is a potential



(a) All OD flows (b) Smaller OD flows

Figure 5-9: Estimated OD flow vs OD flow from the Oyster seed matrix, clamped to London Overground network

cause for concern, because large relative expansions of short OD flows (*i.e.* those covering only a few links) can be a reflection of inconsistencies between the assignment model, the seed matrix, and the given link flows.

That said, these large expansions from Clapham Junction are generally consistent with anecdotal observation on the part of the author and the expectations of Overground management (Bratton, 2010). As was shown in Table 5-2, the West London Line has a particularly low Oyster penetration rate. Figure 5-10 plots the estimated flows against Oyster flows by Overground line. It is clear from this plot that, as a group, flows on the West London Line are expanded differently from flows on the other lines. Moreover, the estimated total number of alightings at the destination stations of these flows is within 10% of the surveyed value, with the exception of West Brompton. Comprehensive validation of this sort is discussed further in Section 5.5.1.

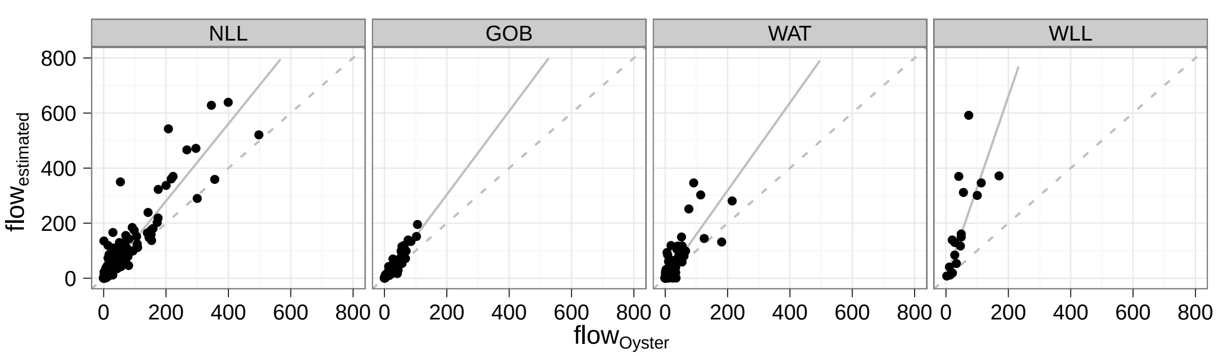


Figure 5-10: Estimated OD flow vs OD flow from the Oyster seed matrix, by London Overground Line

Even though the seed matrix should represent a lower bound, as discussed in Section 5.3, it is possible for flows to shrink in this OD estimation method. Figure 5.9(b) zooms in closer to the origin of Figure 5.9(a) for inspection of smaller OD flows. While 91% of clamped flows expand under the estimation process, the other 9% shrink, a few quite drastically in relative terms. The two most severe examples of this are Hatch End and Carpenders Park, respectively, to Harrow & Wealdstone. These flows had 24 and 35 passengers, respectively, in the clamped seed matrix but near zero passengers in the estimated matrix. The total estimated boardings at the origin stations of these flows match the respective surveyed values to within fewer than 10 passengers, but the total estimated alightings at Harrow & Wealdstone is estimated at only 41% of the actual value. These OD flows, like some of the OD flows that suffered very large relative expansions, are assigned to only two or three links.

Further investigation of the flows that were expanded or reduced beyond reasonable expectations is merited. It is possible that these unexpected results reflect inaccuracies in the otherwise-reasonable results of the assignment model. It is also possible that they reflect gaps (*i.e.* bias) in the seed matrix, likely the result of low Oyster penetration in certain segments of the traveling population. While it is certain that the methodology developed here is imperfect, it may represent an improvement on the current practice at the Overground, as explored in the following section.

5.5.1 Validation

This section presents a comparison of the new London Overground OD matrix estimate to the OD matrix estimated by the RailPlan regional model. The on-board link loads used here to estimate this new OD matrix came from a set of manual counts that also indicate the number of passengers boarding and alighting each service at each station. Since these boarding and alighting counts were not used in the OD estimation process, they can be used to validate the OD estimation results. This is consistent with the structure of the OD estimation problem as discussed in Section 5.2.1.

The assignment model translates the estimated OD matrix into a set of estimated boarding and alighting flows. These are aggregated to the station level, and compared against the manual counts according to different measures of performance, each with its own advantages and disadvantages. For two vectors (in this case sets of boarding counts) x_i, v_i, $i \in 1..N$, with x_i representing the estimated or experimental results and v_i representing the validation "ground truths," the following performance measures can be calculated.

- *Percent Error (%E)* – the percent error of the sum of the two sets of values, measuring the difference in total flow.

- *Percent Absolute Error (%AE)* – the sum of absolute errors over the sum of the validation values, shown in Equation (5-10). This is, in essence, the absolute percent error of each value, weighted by the respective validation values.

$$\%AE = \frac{\sum_{i=1}^{N} |x_i - v_i|}{\sum_{i=1}^{N} v_i} \tag{5-10}$$

The advantage of this measure is that it places more weight on larger flows. The disadvantage is that it masks errors in smaller flows.

- *Mean Absolute Percent Error (MA%E)* – the unweighted average of absolute percent errors, as shown in Equation (5-11).

$$MA\%E = \sum_{i=1}^{N} \frac{|x_i - v_i|}{v_i} \tag{5-11}$$

  The advantage of this measure is that it is sensitive to relative errors on smaller flows. The disadvantage is that for small flows, an error of very few passengers can equate to large percentage errors.

- *Root Mean Square Error (RMSE)* – the euclidean distance, in N dimensions, between the estimated and validation vectors, as shown in Equation (5-12).

$$RMSE = \sqrt{\sum_{i=1}^{N} (x_i - v_i)^2} \tag{5-12}$$

  The primary advantage of this measure is that it places weight on large errors, regardless of the size of the validation value. The disadvantage is that as a quantity it is hard to interpret in a physical sense, and thus is only useful as a relative measure.

Table 5-5 shows the comparison of the two different OD estimates with the manually-counted boardings and alightings across all Overground stations. The total estimated boardings or alightings is in all cases fewer than the total counted boardings or alightings, but by less than 5%. This discrepancy is made worse by the addition of the gateline entry data. This is contrary to expectations, as additional information should improve the accuracy of the estimate. However, all the station-level error measures do improve when gateline entry flows are included in the estimation.

Boardings	Total			By Station		
Source of Link Flows	Estimated	Counted	%E	%AE	MA%E	RMSE
Onboard Link Counts	38,496	38,800	-0.8%	13.4%	17.8%	158.7
+ Gateline Entries	37,941	38,800	-2.2%	10.6%	15.7%	136.9

Alightings	Total			By Station		
Source of Link Flows	Estimated	Counted	%E	%AE	MA%E	RMSE
Onboard Link Counts	38,496	39,748	-3.1%	15.0%	19.8%	167.3
+ Gateline Entries	37,941	39,748	-4.5%	13.0%	18.2%	153.5

Table 5-5: Comparison of estimated and counted boardings and alightings

The OD matrix estimated from on-board link flows and gateline entry flows is thus taken as the best Oyster-based OD estimate. The OD matrix currently used by Overground planners and managers was estimated using the RailPlan regional public transport model in 2008, before the opening of the new Shepherd's Bush station. Table 5-6 presents a comparative validation of these two matrices at the line segment,[4] line, and network levels. For the sake of this analysis, the two stations served by multiple Overground lines – Gospel Oak and Willesden Junction – are separated into the "interchange" (INT) line. The Oyster-based estimate is validated against 2009 boarding and alighting counts (as above) while the RailPlan estimate is validated against 2008 boarding and alighting counts.

		RailPlan, 2008						Oyster-based, 2009					
Line	Segment	Est.	Count	%E	%AE	MA%E	RMSE	Est.	Count	%E	%AE	MA%E	RMSE
NLL	NLLE	11,935	12,691	-6%	19%	32%	322	12,435	13,093	-5%	6%	8%	177
NLL	NLLC	3,198	4,297	-26%	54%	60%	503	4,022	3,694	9%	16%	20%	134
NLL	NLLW	2,382	2,696	-12%	30%	28%	207	3,024	2,830	7%	8%	10%	85
NLL	(Total)	17,515	19,684	-11%	28%	39%	363	19,481	19,617	-1%	8%	12%	148
WAT	WATN	1,787	1,809	-1%	44%	43%	171	1,699	1,734	-2%	4%	7%	17
WAT	WATC	3,518	4,457	-21%	34%	32%	249	3,862	4,428	-13%	13%	14%	111
WAT	WATS	962	876	10%	47%	54%	157	1,066	793	34%	34%	54%	112
WAT	(Total)	6,267	7,142	-12%	38%	39%	212	6,627	6,955	-5%	14%	18%	91
WLL	(Total)	2,784	2,876	-3%	7%	20%	93	4,079	3,580	14%	16%	43%	246
GOB	(Total)	2,977	4,035	-26%	31%	34%	135	3,664	3,985	-8%	9%	9%	44
INT	(Total)	3,007	3,201	-6%	13%	18%	236	4,090	4,663	-12%	12%	14%	298
(Total)	(Total)	32,550	36,938	-12%	27%	36%	267	37,941	38,800	-2%	11%	16%	137

Table 5-6: Line and line segment level validation on counted AM Peak boardings for RailPlan (2008) and Oyster-based OD estimates

At Transport for London, the RailPlan model is typically validated against total boardings (*i.e.* %E) at the line or line segment level (Warner, 2010). By this measure, the Oyster-based OD estimate is more accurate than RailPlan for all lines but one. For the North London Line (NLL), Watford DC Line, and Gospel Oak to Barking Line (GOB), it is more accurate by 10, 8, and 18 percentage points, respectively. For these lines, the new OD matrix underestimates total boardings by only 1%, 5%, and 8%, respectively, but RailPlan understimated total boardings by much more. The West London Line is the exception here, with boardings overestimated by 14%, as compared to RailPlan's 3% under-estimate. For the entire Overground network, the new OD matrix underestimates total boardings by 2% compared to 12% for RailPlan.

For the more disaggregate station-by-station performance measurements, the Oyster-based matrix is also substantially more precise in most cases. For example, consider the Eastern portion of the North London Line (NLLE), between Stratford and Kentish Town West, inclusive – the busiest segment of the Overground network. The %AE and MA%E for this segment are a third (or less) of the values for the RailPlan matrix. The same can be observed for most other segments and lines on the network, with the primary exceptions of the West London Line and the Southern portion of the Watford DC Line. It should be noted that if West Brompton is excluded from the West London Line, the %AE and MA%E improve to 3% and 10% respectively, much better than RailPlan's results. Over the whole Overground network, these two measures are 11% and 16% for the new OD estimate compared with 27% and 36% for RailPlan.

[4] Line segment definitions can be found in Appendix B.

It is clear that while the OD estimation methodology developed here is not perfect, it is substantially better in most cases for estimating *current* demand levels and travel patterns than that currently used by the Overground. This should not be surprising, since RailPlan is estimated based only on regional household travel surveys and does not incorporate any direct measurements of travel patterns or volumes on the network.

5.5.2 Loadweigh Sensitivity Analysis

This section uses simulation to explore the sensitivity of the OD estimation process to random error in the measured link flows. The intent is to assess the robustness of the OD estimation process to the measurement error associated with loadweigh data described in Chapter 4. This data source is the only one addressed because the other inputs – the Oyster system and automatic gatelines – are considered to have little if any random measurement error. For the sake of this analysis, it is assumed that the link flows on which the OD matrix was estimated are the true average link flows. That they were actually derived from manual counts rather than loadweigh data is immaterial here.

The results of Chapter 4 are interpreted to mean that each measurement of the number of passengers on a given train is subject to a random normally distributed additive error term with mean zero and standard deviation of 10 passengers. In other words, the loadweigh estimate of number of passengers $\hat{L}$ is a random variable that is the sum of the true number of passengers, L, (a non-random quantity) and an error term, ϵ, according to the following equations.

$$\hat{L} = L + \epsilon \tag{5-13}$$

$$\epsilon \sim N(\mu = 0, \sigma = 10.0) \tag{5-14}$$

In a certain sense, the strongest part of this assumption is that the measurement error is *not* correlated with the number of passengers, which is consistent with the findings in Chapter 4 and in Nielsen et al. (2008*a*).

Simulations were used to test (i) the effect of this type of measurement error on the outcome of the OD estimation process and (ii) the degree to which the availability of larger volumes of loadweigh data can be exploited to minimize these effects. The goal of these simulations is to test *only* the effects of the loadweigh measurement error and *not* the effects of other sources of stochasticity such as day-to-day variation in demand. Consequently, it is assumed here that the manual on-board passenger counts represent the true average number of passengers on each link on each scheduled service and that this number does not change over time. Exploiting this assumption and the properties of the normal distribution, the simulated average of loadweigh measurements taken over d days, $\hat{L}_d$, can be written as

$$\hat{L}_d = L + \epsilon_d, \tag{5-15}$$

$$\epsilon_d \sim N\left(\mu_d = 0, \sigma_d = \frac{10}{\sqrt{d}}\right). \tag{5-16}$$

Two sets of thirty simulations each were run based on these assumptions. The first set simulates the use of five days (*i.e.* one week) of loadweigh data to estimate on-board link flows. The second set simulates the use of the average of forty days (*i.e.* eight weeks) of loadweigh data to estimate on-board link flows. Each simulation run consists of the following steps.

1. Each individual manual on-board passenger count is perturbed randomly according to Equations (5-15) and (5-16).
2. Any counts that are randomly perturbed to be less than zero are set to zero.
3. Individual counts are aggregated, as in the previous section, to the link and time period (*i.e.* AM Peak) level.
4. The OD matrix is re-estimated, using the same Oyster seed matrix and gateline entry counts as previously.
5. The resultant OD matrix is clamped to the London Overground network and the validation measures from Section 5.5.1 are calculated, at the level of OD flows, with respect to the clamped OD matrix estimated on unperturbed data in the previous section.

The error introduced into the OD estimate by the simulated loadweigh measurement error is summarized in Table 5-7. It appears that the simulated random loadweigh error does effect the estimated OD matrix, but that averaging over forty days of data reduces the effect significantly. For example, the average %AE and MA%E under five days of data are 20.1% and 39.1%, respectively[5]. These are unacceptable deviations, but are reduced to 1.4% and 2.2%, respectively under forty days of data.

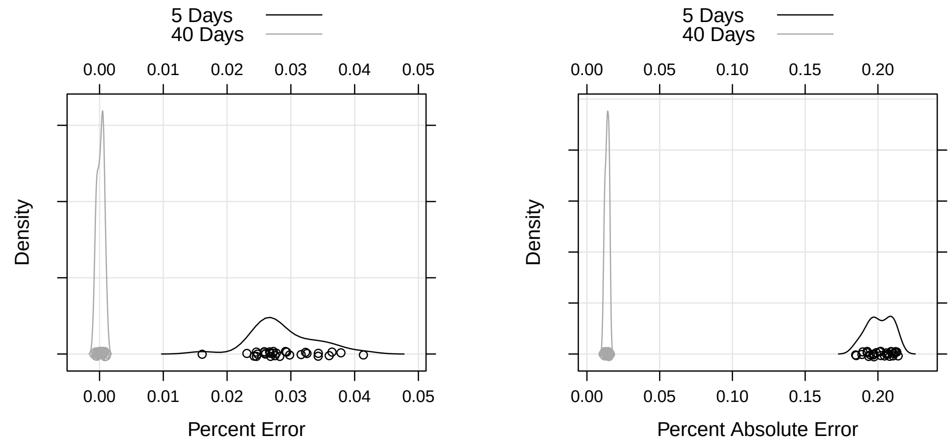
days	% Error	%AE	MA%E	RMSE
5 days	2.88%	20.1%	39.1%	15.0
40 days	0.02%	1.4%	2.2%	0.9

Table 5-7: Average values of validation measures for OD estimation under simulated loadweigh error

Figure 5-11 plots the distribution of two of the validation measures – the total percent error (*i.e.* the difference in the sum of the matrices) and the percent absolute error (%AE). On average, the simulations do not change the total number of passengers substantially, but the distribution of this change is much tighter for the simulations of forty days of loadweigh data. Likewise, the %AE is both lower on average and more tightly distributed for the forty-day simulation.

The simulations presented in this section show that the OD estimation process is in fact sensitive to errors in the measurement of on-board link flows, but that the errors can be reduced to acceptable tolerances by averaging over larger amounts of data. This is clearly the strength of loadweigh data – that it is continuously available in large quantities at low

[5] The calculation of MA%E excludes certain edge cases which caused huge absolute percentage errors. These were OD flows that had very small values (*i.e.* $< .01$) in the non-perturbed estimate so any increase (*e.g.* to 0.05) in the perturbed was a huge change in percentage terms.



(a) Percent Error (sum of matrices) (b) Percent Absolute Error (flow-by-flow)

Figure 5-11: Smoothed densities of OD validation measures under simulated loadweigh error

cost – as compared to manual counts which are taken only once per counting period at high cost.

This section did not consider day-to-day variability in loads on trains, which must exist in practice. Chapter 4 analyzed the load on a single scheduled service on a single link, estimating a standard deviation of the true load of 7.7 passengers. Considering this an independent error term would approximately double the standard deviation of error in loadweigh estimates, requiring an approximate quadrupling of the sample size to achieve the level of accuracy found here for the simulated 40 day sample.

This should be considered a worst case scenario. The analysis in Chapter 4, for a single peak hour service at its peak load point, was fairly speculative. Moreover, the day-to-day variability in loads on individual services are likely *not* independent of each other. If one train has more than its average number of passengers because it is running late (and thus has a long leading headway), the train following it is likely to have fewer than average. In other words, the total number of passengers on a given link in the whole AM Peak period (on which the OD matrix is estimated) can be stable even when those passengers change their distribution between individual trains because of variability in operating conditions.

5.6 Conclusions and Recommendations

This chapter developed a methodology to estimate OD matrices for railway networks from multiple automatic data sources, including aggregate journey data from AFC systems, automatic entry/exit counts from station gatelines, and on-board passenger loads estimated from loadweigh data, and applied this method to the London Overground network. This section presents first some conclusions drawn from the analysis in this chapter, and next some recommendations based on those conclusions.

5.6.1 Conclusions

The following conclusions are drawn in this chapter.

- Link flows from loadweigh measurements and/or manual on-board link counts can be combined, through a mathematical estimation process, with aggregate transactional data from the Oyster smartcard ticketing system to estimate time period level OD matrices for the London Overground network.

- The overall accuracy of the OD estimate is improved by the addition of automatic entry and/or exit totals from gatelines at stations exclusive to the Overground. However, these gateline totals must be compared with corresponding Oyster totals to identify and remove faulty data.

- Of the wide range of available mathematical models and methods, a relatively simple approach is sufficient to use these data sources to improve the accuracy and timeliness of OD matrices for the Overground network compared with the existing OD matrix from the RailPlan regional model.

- The key outputs of the network assignment model developed here, which does not account for congestion or capacity constraints, are relatively insensitive to most embedded assumptions regarding passenger path choice. Specifically, the choice for most passengers of whether or not to use the Overground network does not change when most of the model's assumptions are violated. One assumption to which the model is sensitive is indifference between interavailable services. The assignment results do change significantly when this feature of the model is disregarded, so it is concluded that this is an important feature that should not be neglected.

- The Information Minimization method for OD estimation from link flows and a seed matrix is suitable to the problem faced by the Overground. It is simple to implement, is conceptually very similar to the matrix estimation method used by the London Underground, and has the very important feature that its results are not sensitive to overall scaling of the seed matrix when total number of passengers (*i.e.* the sum of the OD matrix) is not fixed. This final feature is key to the Overground application because the Oyster seed matrix is a lower bound on the true OD matrix and, unlike in the London Underground case, the available data are such that the total number of passengers is determined by the OD estimation process.

- The OD estimation process developed here is insensitive to measurement error in the loadweigh data under two conditions. First, that the measurement error is unbiased and uncorrelated with the actual number of passengers, which has been found to be the case in this and other work. Second, that there is a sufficient quantity of loadweigh data (*i.e.* at least eight weeks) over which to estimate average loads.

- There is evidence that this OD estimation process *may* expand some shorter OD flows beyond reasonable limits in order to match the given link flows. This is most apparent for certain OD flows on the West London Line, but it is impossible to conclude that this behavior is in fact erroneous in most or all cases. One reasonable explanation of this behavior is the low penetration rate of Oyster along this corridor, especially for National Rail passengers interchanging at Clapham Junction.

- The OD estimation method developed here does not treat the seed matrix as a lower bound of the estimated matrix, and some estimated OD flows are in fact lower than their respective values in the seed matrix. In this sense, the method lacks a constraint that is necessary to faithfully model the real-world phenomenon. The violation of this constraint is not considered a serious problem for this method. It affects only a very small portion of the OD flows, and those it does affect are for the most part relatively small flows to begin with.

Broadly speaking, it is concluded that the methodology developed here would represent an improvement with respect to the Overground's current practices, but that there is also potential for further improvement.

5.6.2 Recommendations

The methods developed in this chapter should be adopted by the London Overground, and should be considered by other railways with similar available data. The assignment model is custom tailored to the specific circumstances of this particular railway, and so may not be applicable to other situations. The OD estimation method, drawn directly from existing literature, is more likely of use in a broader range of contexts. It should be considered in other circumstances where AFC systems provide a high quality (if not perfectly representative) seed matrix and where link flows can be estimated with sufficient confidence that they can be considered deterministic constraints.

One particular aspect of the OD estimation methodology used here merits further research. The constraint of the seed matrix as a lower bound on the final OD estimate should be added to the Information Minimization formulation. It is trivial to add this constraint to the formulation, but it may make the model much more difficult if not impossible to solve efficiently. It is possible that a lagrangean analysis similar to that developed by Van-Zuylen and Willumsen (1980) would yield an efficient algorithm as it has for the existing formulation.

Implementation for the London Overground

In terms of application to the London Overground, the methods developed in this chapter should be applied to the network as it exists today and extended as the network expands. The forthcoming East London Line will be served entirely by loadweigh-enabled rolling stock, will be fully Oyster-enabled, and most of its stations will be gated. The RODS network representation will need to be expanded to include the new East London Line.

Unfortunately, the implementation of the proposed assignment model and estimation method is a complex undertaking. Fortunately, the capacity to execute such a project exists within TfL, most obviously within the Strategy and Service Development group of the London Underground. It is that group which maintains the RODS network representation, for the purpose of estimating OD matrices for the London Underground. Collaboration of

London Underground and Overground analysts on such a project would also be a first step towards the longer-term goal of estimating integrated OD matrices for the entire London railway network (including the Underground, Overground, DLR, and National Rail) from Oyster and other automatic data sources.

Alternatively, a third party could be contracted to operationalize the prototypes developed for this report. Ideally this would be a party experienced in transportation modeling as well as custom software development. It is possible that this third party would be a new TfL modeling software and data analysis group, or it could be an external contractor.

Regardless of who takes on the task of operationalizing the methods developed in this chapter, the working prototypes should be utilized as much as is beneficial. They are developed in the free and open source programming languages *Java* and *R*, and so are readily available for inspection, modification, or partial or complete re-use.

Chapter 6

Passenger Incidence Behavior

This chapter is concerned with certain behavioral elements of passenger arrival *to* public transport services. Because passengers can also arrive *at* certain locations *via* public transport services, a lexical convention is established to avoid ambiguity of exposition. *Passenger incidence* is defined here as the act or event of being incident *to* a public transport service with intent to use that service. Passenger *arrival* is defined here as the act or event of arriving *at* a certain location having used public transport services. When making an interchange, a passenger can arrive *at* the interchange location using one service and simultaneously be incident *to* the next service he or she intends to use.

This chapter is primarily concerned with the relationship between the times of passenger incidence and published timetables. It proposes a method to study this relationship by integrating disaggregate passenger journey data from automatic fare collection (AFC) systems with published timetables using schedule-based assignment. The purpose of this chapter is three-fold. Firstly, to develop a method that contributes to the study of passenger incidence behavior across a railway network with heterogeneous service patterns and frequencies using published timetables and AFC data. Secondly, to shed light on the incidence behavior of London Overground passengers. Lastly, to help set the stage for the following chapters which are concerned with measuring service quality with respect to passenger expectations, some of which are reflected in their incidence behavior.

Section 6.1 reviews literature relevant to the purposes of this chapter, include analysis of passenger incidence behavior and methods for schedule-based assignment. Section 6.2 defines certain analytical quantities through which incidence behavior can be studied. It proposes a method to derive those quantities from the integration of disaggregate passenger journey data and published timetables, and describes the means by which this method was implemented for the Overground. Section 6.3 presents results that describe the incidence behavior of Overground passengers. Section 6.4 offers some preliminary conclusions and recommendations.

6.1 Literature Review

This section reviews first some of the literature on passenger incidence behavior, next the literature on certain aspects of schedule-based assignment.

6.1.1 Passenger Incidence Behavior

Random Incidence

One characterization of passenger incidence behavior is that of *random incidence* (Larson and Odoni, 2007). The key assumption underlying the random incidence model is that the process of passenger arrivals to the public transport service is independent from the vehicle departure process of the service. This implies that passengers become incident to the service at a random time, and thus the instantaneous rate of passenger arrivals to the service is uniform over a given period of time. Let W and H be random variables representing passenger waiting times and service headways, respectively. A classic result of transportation science is that under the random incidence assumption[1]

$$\mathrm{E}[W] = \frac{\mathrm{E}[H^2]}{2\,\mathrm{E}[H]} = \frac{\mathrm{E}[H]}{2}\left(1 + \mathrm{cv}(H)^2\right) \tag{6-1}$$

where $\mathrm{E}[X]$ is the probabilistic expectation of some random variable X; $\mathrm{cv}(H)$ is the *coefficient of variation* of H, a unitless measure of the variability of H defined as

$$\mathrm{cv}(H) = \frac{\sigma_H}{\mathrm{E}[H]}; \tag{6-2}$$

and σ_H is the standard deviation (the square root of the variance) of H (Osuna and Newell, 1972). The second expression in Equation 6-1 is particularly useful because it expresses the mean passenger waiting time as the sum of two components: the waiting time due to the mean headway (*i.e.* the reciprocal of service frequency) and the waiting time due to the variability of the headways (which is one measure of service reliability). When the service is perfectly reliable with constant headways, the mean waiting time will be simply half the headway.

The following less well-known result, first derived in the literature by Friedman (1976) and explored further by this author and others (Frumin et al., 2010), describes the variance of passenger waiting times under the same random incidence assumptions:

$$\mathrm{Var}(W) = \frac{\mathrm{E}[H^3]}{3\,\mathrm{E}[H]} - \left(\frac{\mathrm{E}[H^2]}{2\,\mathrm{E}[H]}\right)^2. \tag{6-3}$$

These authors have noted that the variance of waiting time is thus a function of the symmetry of the headway distribution, which is reflected in the $\mathrm{E}[H^3]$ term.

Behavioral Incidence

It is often assumed that the random incidence assumption holds at "short" headways (Furth and Muller, 2006). The balance of this section reviews six studies of passenger incidence behavior which are motivated by understanding the relationships between service headway, service reliability, passenger incidence behavior, and passenger waiting time in a more nu-

[1] The given result also depends on the assumption that vehicle capacity is not a binding constraint – that all passengers are able to board the first desired departing vehicle.

anced fashion than is embedded in this assumption. Three of these studies depend on manually collected data, two use data from AFC systems, and one studies the issue purely theoretically. These studies reveal much about passenger incidence behavior, but all are found to be limited in their general applicability by the methods with which they collect information about passengers and the services those passengers intend to use.

Jolliffe and Hutchinson (1975) studied bus passenger incidence in South London suburbs. They observed ten bus stops each for one hour per day over eight days (for a total of 80 hours of observation), recording the times of passenger incidence and actual and scheduled bus departures. They characterized the reliability of the service by the standard deviation of the difference between observed and scheduled bus departure times. They limited their stop selection to those served by only a single bus route with a single service pattern so as to avoid ambiguity about which service a passenger was waiting for. The authors found that the actual average passenger waiting time was 30% less than predicted by the random incidence model. They also found that the empirical distributions of passenger incidence times (by time of day) had peaks just before the respective average bus departure times, and that on individual days there were spikes in the incidence rates coincident with actual bus departures. In other words, passengers adjust their incidence behavior, both planned and in real time, based on knowledge of the bus timetable and historical bus performance. They hypothesized the existence of three classes of passengers:

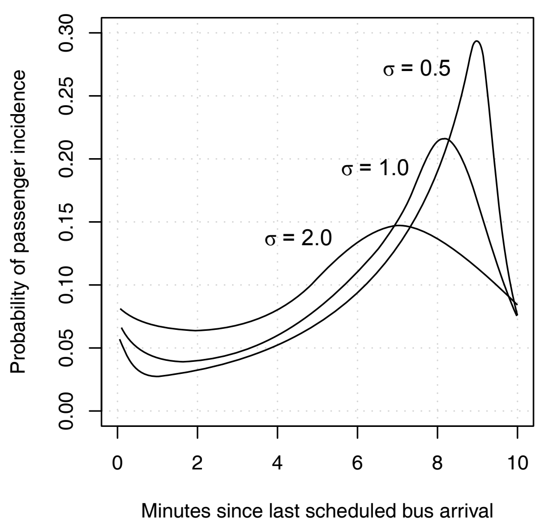
- With proportion q, passengers whose time of incidence is causally coincident with that of a bus' departure (*i.e.* because they saw the approaching bus from their home, a shop window, etc).
- With proportion $p(1-q)$, passengers who time their arrivals to minimize expected waiting time (*i.e.* based on some awareness of the timetable and reliability).
- With proportion $(1-p)(1-q)$, passengers who are randomly incident.

Estimating these proportions, Jolliffe and Hutchinson found that p was positively correlated with the potential reduction in waiting time (compared with arriving randomly) resulting from knowledge of the timetable and of service reliability. Namely, that p was correlated with the headway and with the reliability of departure times. They also found p to be higher in the peak commuting periods rather than in the off-peak, indicating more awareness of the timetable and/or historical reliability by commuters. This study did not propose any model by which to estimate distributions of passenger incidence times (for example under different headways or levels of reliability).

Bowman and Turnquist (1981) build on the concept of "aware" and "unaware" passengers (of proportions p and $(1-p)$, respectively) described by Jolliffe and Hutchinson (1975). They propose a utility-based model to estimate p as well as the distribution of incidence times (and thus the mean mean waiting time) of "aware" passengers over a given headway as a function of the headway and reliability of bus departure times. They observed seven different bus stops in Chicago, each served by a single (different) bus route, between 6:00 and 8:00am for 5-10 days each (approximately 105 hours total observation). The bus routes had headways of 5-20 minutes and a range of reliabilities. As in the previous study, the authors found that actual average waiting time was substantially less than predicted by the random incidence model.

Estimating their model, Bowman and Turnquist found that p was not statistically significantly different from 1.0, which they explain by the fact that all observations were taken during peak commuting times. The estimated model predicts that the longer the headway and the more reliable the departures, the more peaked the distribution of incidence times will be and the closer that peak will be to the next scheduled departure time (*i.e.* the end of the headway). This demonstrates what they refer to as a "safety margin" that passengers add to reduce the chance of missing their bus when the service is known to be somewhat unreliable. Such a safety margin can also result from unreliability in passengers' journeys *to* the public transport stop or station. Life is random – if a passenger is unsure of exactly how long it takes to walk to the station, he or she may leave a few additional minutes to be sure to arrive at the station before the scheduled departure time.

The model of Bowman and Turnquist is illustrated in Figure 6-1 for a 10 minute headway and different levels of reliability of departure time. They conclude from their model that, in general, the random incidence model underestimates the waiting time benefits of improving reliability and overestimates the waiting time benefits of increasing service frequency (*i.e.* lowering the headway).[2] This is because, as reliability increases passengers can better predict departure times and so can time their incidence to decrease their waiting time. Moreover, as frequency increases incidence may become more random, thus lengthening waiting times.



σ indicates reliability of departure time. Adopted from Bowman and Turnquist (1981).

Figure 6-1: Distributions of passenger incidence for different levels of reliability of departure time

Furth and Muller (2006) study the issue in a theoretical context and generally agree with the above findings. They are primarily concerned with the use of data from automatic

[2] This analysis does not consider other benefits of increasing service frequency, for example decreasing the "schedule delay" experienced by passengers whose preferred departure or arrival times do not align perfectly with the timetable (*e.g.* Bates et al., 2001).

vehicle tracking systems to assess the impacts of reliability on passenger incidence behavior and waiting times. They propose that passengers will react to unreliability by departing earlier than they would with reliable services. Randomly incident "unaware" passengers will experience unreliability as a more dispersed distribution of headways and simply allocate additional time to their trip plan to improve the chance of arriving at their destination on time. "Aware" passengers, whose incidence is not entirely random, will react by timing their incidence somewhat earlier than the scheduled departure time to increase their chance of catching the desired service. The authors characterize these reactions as the costs of unreliability.

Luethi et al. (2007) continue with the analysis of manually-collected data on actual passenger behavior. They use the language of probability to describe the two classes of passengers. The first is "timetable-dependent" passengers whose incidence behavior is affected by awareness (possibly gained through their own experience with the service) of the timetable and/or service reliability (*i.e.* the "aware" passengers). The second is "timetable-independent" passengers whose incidence behavior is random and so does not reflect any such awareness (whether or not they have it). The language of timetable-dependency is adopted for the balance of this chapter to describe the randomness of passenger incidence behavior regardless of what exactly is driving the behavior on the part of the passengers. It is preferred because it expresses the observed probabilistic association between two variables (*i.e.* incidence times and scheduled departure times) rather than some unobserved passenger state of mind.

Luethi et al. observed passenger incidence at 28 bus, tram, and commuter rail stations in and around Zurich, Switzerland, with headways of 2.33-30 minutes, during morning and evening peak hours and midday off-peak hours (for a total of approximately 200 hours). To avoid ambiguity, they limited their station selection to non-terminal non-interchange stations served by a single route with constant headways over the period of observation. Exploratory analysis of the observed distributions of incidence times concurred with the finding of Bowman and Turnquist (1981) that longer headways result in more peaked distributions of passenger incidence over a given headway. The authors observed that a substantial share of passengers appear to be timetable-dependent for headways as low as five minutes. They also observed spikes in the distribution at the beginning of respective headways, which they attribute to the assumption "that some passengers know very well the reliability and average delay of the public transit service and therefore arrive regularly a short time after the scheduled departure time."

The primary goal of this study was to estimate distributions of passenger incidence times over a given headway which they propose to be the weighted superposition of two distributions. The incidence times of timetable-independent passengers, with weight $1 - p$, are distributed uniformly over the headway. Timetable-dependent passengers, with weight p, are distributed according to a Johnson S_B (JSB) distribution, which is similar to the Normal distribution but skewed to the right for certain values of its parameters. The authors modify the JSB distribution to admit an additional parameter indicating a rightward shift for those passengers who regularly arrive shortly after the scheduled departure time. The authors found the fit of their bespoke distribution to the observed data to be statistically significant. This distribution is parameterized in terms of the headway, but, unlike the model of Bowman and Turnquist (1981), it is not parameterized in terms of the reliability of the service; it is a descriptive rather than predictive model. Estimations of their model yield values of p for the different time periods of the day, which they find to be highest in the morning peak and lowest in the off-peak, supporting the conclusion that the incidence of commuters is more timetable-dependent. Their data collection also included brief interviews with passengers, which indicated that timetable-dependence is also correlated with experienced service reliability.

Csikos and Currie (2007, 2008) study this phenomenon, first cross-sectionally and then longitudinally, using data from the AFC system of the Melbourne, Australia, heavy rail network. In their first study they use four weeks worth of data from 07:30 to 15:00, but limit themselves to analyzing seven particular stations (out of 209), for a total of 38,000 observations over approximately 1,470 hours. The stations are, as in the other studies, selected to avoid ambiguity regarding which scheduled service each passenger intended to use. They also obtained high level data about the aggregate reliability (6-minute terminal on-time performance) of the train lines serving the selected stations. Their findings generally confirm those of the other studies – that passenger incidence is more timetable-dependent with a more peaked distribution during peak hours, in longer headways, and for more reliable services.

The authors found less evidence of the "safety-margin" phenomenon than in previous studies. The suppose that this is because the previous studies focused on bus (rather than rail) services which may be more prone to early departures. At the station on the least reliable rail line in their study, the authors found that the distribution of passenger incidence times peaked exactly at the time of scheduled departure and had a high fraction of passengers incident just after this time (*i.e.* at the beginning of the successive headway). They hypothesize the existence of passengers with "late running awareness" who time their incidence to account for regular late departure of trains. They point out that such a behavior pattern would result in overestimates of actual passenger waiting time when comparing passenger incidence times to the schedule *per se*.

In their second study, Csikos and Currie (2008) use the same four week data set as in their first, but this time track individual ticket holders over time to study the consistency of behavior. They focus on the 15,000 trips made between 06:00 and 10:00 by 1,043 individual passengers who, as morning commuters, are expected to exhibit the most consistent behavior patterns. They characterize the passengers by the times of incidence and the "offset" times until the next scheduled departure. They classify passengers into four distinct archetypes exhibiting various levels of consistency in these two variables, finding roughly equal numbers of passengers in each of the four categories. On one end the spectrum are "like clockwork" passengers who exhibit fairly consistent behavior that often minimizes their "offset" (*i.e.* waiting) time with respect to the schedule. On the other end, "largely random" passengers have very little consistency with respect to "offset" time, exhibiting largely timetable-independent behavior. All classes of passengers used (according to the timetable) numerous different scheduled services over the observation period. A small fraction of passengers (less than 10%) with median incidence times just after scheduled departure times exhibit serial behavior indicating "late running awareness." The authors' overall conclusion in this work is one of heterogeneity in passenger behavior, even under homogeneous conditions (*i.e.* at the same station at the same time of day served by the same line).

Discussion

Previous research has identified a rich set of passenger incidence behaviors, and related them to certain aspects of public transport services. It has done so using manually and automatically collected data sources, and has used automatic data to investigate the consistency of such behaviors longitudinally over time. It has been found that the randomness of passenger incidence behavior is highly dependent on the service headway and the reliability of the departure time of the service to which passengers are incident.

Passenger incidence behavior has been characterized primarily in terms of how random it appears to be with respect to the timetable and to actual vehicle departure times. The appearance of randomness (or lack thereof) has been used to indicate the degree to which passengers have and use knowledge of the published timetable and of actual departure times. At longer headways, passengers have more to gain by gaining and using knowledge of the timetable; their behavior tends to be less random, peaking somewhat before the scheduled departure time. Passengers also appear to gain and use knowledge of the actual, rather than scheduled, departure times. When departure times are reliable, even if they are reliably late (or early) by a particular amount, incidence behavior tends to be less random with more passengers being incident shortly before the reliable departure time. When departure times are inconsistent (*i.e.* unreliable), passengers have less to gain from choosing any particular time of incidence, so their behavior tends to be more random.

Passenger incidence behavior has been studied primarily for the sake of understanding how changes to a public transport service will affect passenger waiting times. It is also of interest because it affects the relationship between the departure times of public transport vehicles and the passenger loads on those vehicles. In the context of managing an urban railway, it is thus important to understand passenger incidence behaviors so that management interventions (including tactical planning) will be based on realistic assumptions going forward. As pointed out in one of the seminal investigations on the topic by Bowman and Turnquist (1981), the effects on passenger waiting time of one particular intervention (*i.e.* increasing frequency) could be overestimated compared with a different type of intervention (*i.e.* improving reliability), depending on what assumptions about incidence behavior are made.

None of the research reviewed here has studied the effect on incidence behaviors of delivering timetable or real-time information to passengers via the internet or mobile devices. This subject is becoming increasingly important as such technologies are rapidly adopted by public transport providers and passengers world wide.

6.1.2 Schedule-Based Assignment

The authors of all of the studies reviewed in the previous section chose their data samples such that the linking of passengers to scheduled or actual services was straightforward. In more complex environments – for example where passengers at a given station have a choice of services – a more sophisticated approach is needed to study passenger incidence behavior. One aspect of the approach that will be used in this research is that of *schedule-based assignment*, which was introduced in Section 5.2. This section completes the review of the relevant aspects of this methodology. Nuzzolo and Crisalli (2004) present a good review of schedule-based assignment and the various sub-models on which it depends.

Run-Based Supply Models

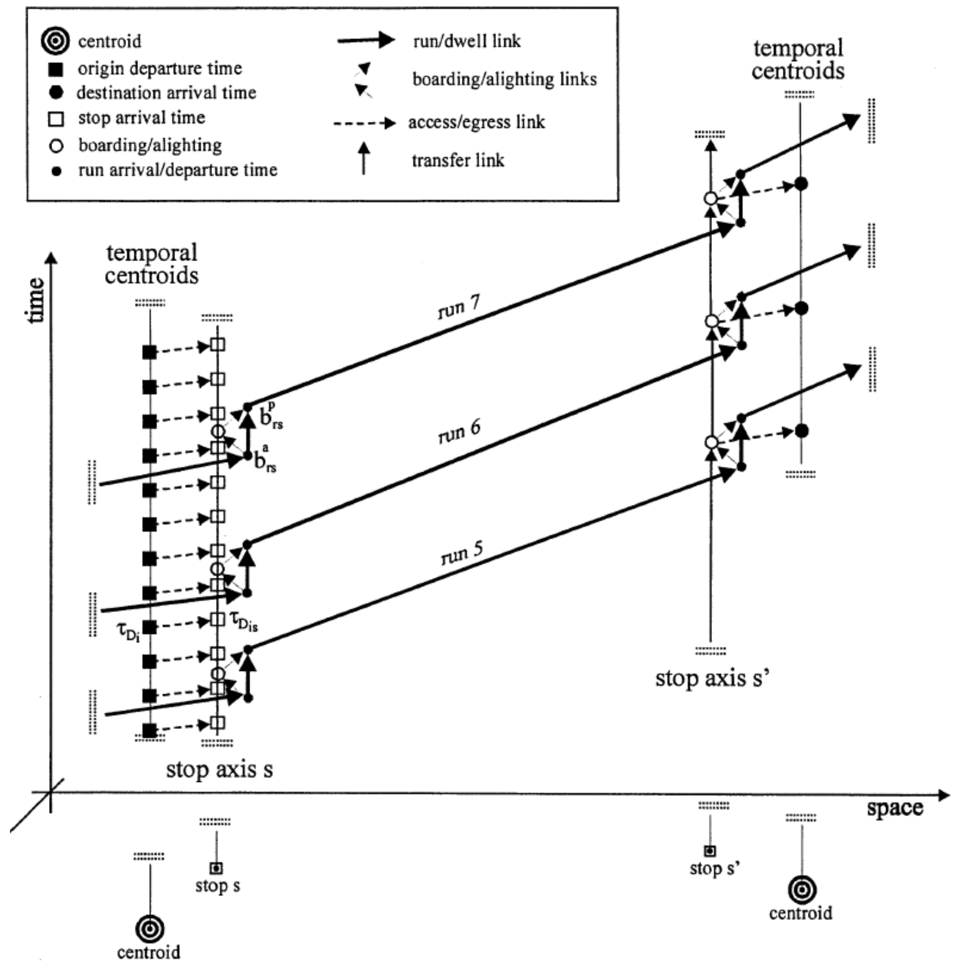
As mentioned in Section 5.2, schedule-based assignment depends on a run-based model of public transport supply. Figure 6-2 illustrates the run-based supply model described by Nuzzolo and Crisalli (2004). This model is very similar to the line-based model of supply, but unfolded in the temporal dimension. In such a model, each individual scheduled or actual run (or trip) of the public transport service is represented individually by its own subgraph. In the subgraph for a given run, the nodes represent the arrival, departure, or transit of that run at a specific location at a specific time. The links represent travel (or dwelling) on that run between specific points in time and space. The combination of the subgraphs of all runs is referred to as the *service subgraph*.

In this run-based model, demand is also modeled with temporal as well as spatial dimensions in the *demand subgraph*. Nodes in this subgraph represent centroids of demand in time, according to user departure and arrival times, and space, according to the physical network. The *access/egress subgraph* joins the service and demand subgraphs with boarding and alighting links. The union of these three subgraphs is referred to as the *diachronic graph* representation. One benefit of such a representation is that shortest travel time paths can be found via standard shortest-path network algorithms such as Bellman-Ford or Dijkstra's (cf Bertsimas and Tsitsiklis (1997)).

This highly granular representation can cause an explosion in the number of nodes and links in the graph representing the public transport network. This presents a problem for some large-scale applications in which requirements for high temporal granularity necessitate a run-based representation. Specialized algorithms and data structures have been developed to treat these problems in practice. Tong and Wong (1999) and Florian (2004) describe optimizations of the above graph representations and algorithms. The core idea of these optimizations is to represent the public transport network in a more concise line-based manner and to unfold lines into runs only as needed.

Schedule-Based Path Choice and Assignment Models

Use of a run-based supply model naturally results in paths which include the temporal dimension. However, the path choice and assignment models used for schedule-based assignment are quite similar to those described in Section 5.2 for frequency-based assignment (excepting the hyperpath model, which does not apply here). Some proposed variations account for hypothesized differences in passenger behavior on low and high frequency services, respectively. Others account for the stochasticity of the transport service itself. Nuzzolo and Crisalli (2004) review these models, with many references to additional literature on the subject. The vagaries of these models are beyond that is needed for the methodology and application described in this chapter, so the reader is referred to the literature for further detail.



From Nuzzolo et al. (2001)

Figure 6-2: Illustration of run-based supply model

6.2 Methodology

This section proposes a method for studying passenger incidence with respect to scheduled departure times (and implied headways). It focuses on scheduled times rather than actual times as a starting point of analysis because actual departure times are themselves often compared to their scheduled values. The proposed method depends on the following concepts, for a given passenger journey.

- *Attractive Departure*[3] – a departure scheduled from the passenger's station of incidence that the passenger is or would have been *willing* to board, however "willing" is defined. This concept makes explicit the possibility that some scheduled departures may not be viable alternatives for a given passenger as a function of that passenger's destination and of the subsequent itinerary of those departures. For example, on a line with a

[3] The use of the word "attractive" is in the tradition of Spiess and Florian (1989) and Nguyen and Pallottino (1988) in their work on hyperpaths and optimal strategies as discussed in Chapter 5. They defined the "attractive set" of lines as the set that a passenger is willing to board at a given location.

trunk and branches, passengers bound for one of the branches may experience longer headways than those traveling only on the trunk.

- *Scheduled Waiting Time* (SWT) – the time the passenger should have to wait according to the schedule, given his or her time of incidence and attractive departures. This is defined as the length of time between passenger incidence and the next attractive departure.

- *Incidence Headway* – the (scheduled) headway applicable to the passenger given his or her time of incidence and set of attractive departures. This is defined as the length of time between the last attractive departure prior to the time of incidence and the next such departure after the time of incidence.

The studies of passenger incidence behavior reviewed above all selected places and times of observation so as to avoid ambiguity with respect to each passenger's attractive departures. They trivialized the measurement of SWT and incidence headways by selecting stations served by only a single service pattern and, in some cases, with a constant headway. While this may be sufficient for academic studies and modeling of passenger behavior, it is clearly inadequate for understanding behavior across an entire network. In many real-world public transport networks, the largest numbers of passengers are incident at large stations or terminals that provide access to heterogeneous services.

In the case of the London Overground, this is most problematic on the North London Line (NLL). Consider, for example, passengers incident to the NLL at Stratford, one of the Overground's busiest stations. In 2008 peak period timetables, the NLL was running a mostly (but not perfectly) regular 15-minute (*i.e.* 4tph) service all day from Stratford to the end of the NLL at Richmond. This was augmented with occasional irregular services – a "shuttle" that ran only as far as Camden Road, and one "special" that ran on the NLL to Willesden Junction but then on the West London Line to Clapham Junction. It is not immediately obvious which of these services would be attractive to a given passenger at Stratford, and thus not clear what incidence headway each passenger experiences. To address this issue requires knowing that passenger's eventual destination, and possibly understanding that passenger's travel preferences. The previous literature avoided this issue by avoiding stations such as Stratford altogether.

The method proposed here is designed to support the study of passenger incidence behavior while accounting for the type of ambiguity described above. It does so by estimating SWT and incidence headway automatically from the integration of published timetables with disaggregate AFC passenger journey data via schedule-based assignment. These two quantities are necessary, but not always sufficient, to characterize passenger incidence behavior. They describe much about the relationship between passenger incidence and the published timetable (with implied headways), but clearly do not reflect any explicit information about service reliability. This method is developed as a tool to aid in study of passenger incidence behavior in general.

6.2.1 Behavioral Assumptions

Assume that, for a given origin, destination, and time of incidence, all passengers plan to use the single schedule-based path (*i.e.* set of scheduled services) through the network that minimizes total travel time. Additionally, assume that passengers plan itineraries to minimize the number of total boardings up (and only up) to the point where total travel time is not increased (*e.g.* in the trunk-and-branch example, branch-bound passengers won't board a train bound for the wrong branch just to get to the end of the trunk).

These assumptions are necessarily a simplification of the true behaviors and perceptions of passengers. The degree to which they hold is a function of the attributes of the particular network to which they are applied and of the behavioral preferences of the passengers in question. In any case, these assumptions are sufficient to determine, for each passenger journey, the attractive departures prior and subsequent to the time of incidence. SWT and incidence headway can be determined once the times of these two departures are known.

6.2.2 Algorithm

For a given passenger journey on a given public transport network, let

SWT = the scheduled waiting time for the given journey;

H_I = the incidence headway for the given journey;

I = time of passenger incidence for the given journey;

L_O = location of incidence of the journey in question (*i.e.* the origin);

L_D = destination of that journey;

D_{prior} = time of last attractive departure prior to I;

D_{next} = time of first attractive departure after I;

H_{max} = the maximum normal headway (*i.e.* time between any two successive departures in the same direction from the same location) on the network;

H_{min} = the minimum normal headway on the network;

Path($from, to, time$) = a function that finds the shortest weighted travel time path from location $from$ to location to with departure time strictly greater than $time$, with all travel time weights equal to 1 except for a transfer or boarding penalty that is positive but less than H_{min};

Departure($path$) = a function that returns the scheduled departure time of path $path$.

The Path() function encapsulates the complexity of conducting a schedule-based assignment for a single passenger trip. Embedded in this function is some efficient algorithm for finding shortest weighted travel time paths in a schedule-based network. The travel time weightings are such as to enforce the assumption that passengers minimize the number of boardings (without affecting total travel time).

Algorithm 6.1 can then be used to find H_I and SWT for the journey in question under the stated assumptions. This algorithm appears quite simple because most of the complexity in the problem is encapsulated by the Path() function. It will either find the prior attractive departure time D_{prior} or determine that there is no prior attractive departure in at most $\frac{H_{max}}{H_{min}}$ steps. It if does find D_{prior}, it uses that result to determine H_I.

Lines 1 through 3 accomplish the simple task of finding the next attractive departure and thus determining the scheduled waiting time (SWT). Lines 4 through 9 search backward in time in increments of H_{min} until either a new attractive departure time d is found or the time has been moved by more than H_{max}. H_{min} is the largest step possible such that the search will never skip over a possible attractive departure. In theory, the algorithm could use the timetable to determine the next departure time in this backward search process rather than blindly stepping in increments of H_{min}. However, the algorithm is unaware of particular departure times since the Path() function encapsulates all knowledge of the timetable itself. This particular algorithmic design is motivated primarily by implementation concerns, discussed in the following section.

Algorithm 6.1 Algorithm to find scheduled waiting time and incidence headway for a given passenger journey

1: $p := \text{Path}(L_O, L_D, I)$
2: $D_{next} := \text{Departure}(p)$
3: $SWT := D_{next} - I$
4: $i := I$
5: $d := D_{next}$
6: **while** $d = D_{next}$ **or** $D_{next} - i \leq H_{max}$ **do**
7:     $i := i - H_{min}$
8:     $d := \text{Departure}(\text{Path}(L_O, L_D, i))$
9: **end while**
10: **if** $d \neq D_{next}$ **then**
11:     $D_{prior} := d$
12:     $H_I = D_{next} - D_{prior}$
13: **else**
14:     $D_{prior} := \text{null}$
15:     $H_I := \text{null}$
16: **end if**

This algorithm improves on the previous approaches to finding SWT and H_I by considering the timetable in the context of each individual journey. The origin and destination of each journey determine which departures will be attractive. For example, if a passenger is traveling from one end of a public transport line to the other, no scheduled short-turn services[4] would be attractive to that passenger because they would increase the number of boardings without improving the total travel time.

[4] A *short-turn* is a service that, either in the timetable or as a result of real-time control actions, turns around before reaching the usual end of the line.

6.2.3 Implementation

Conceptually, it is simple to implement the above methodology using Oyster journey data and published timetables. The transaction time of the Oyster entry can be taken as the time of incidence, and the origin and destination of the journey can be used along with the timetable to estimate which services the journey was incident to. This section describes some of the finer points of such an implementation.

Data Considerations

The data available for analyzing passenger incidence behavior on the London Overground network have a number of limitations that must be considered:

- Overground timetables indicate departures in minutes but not seconds;
- the Oyster system truncates seconds from timestamps in Oyster journey data (*e.g.* all transactions between 08:00:00 and 08:00:59 are recorded at 08:00:00);
- for stations with entry/exit gatelines *and* platform validators, available Oyster data do not indicate where exactly in a station the validation took place;
- data on distance (or typical walking times) between station entry/exit gatelines and train platforms was not conveniently available.

As a result, the following assumptions are made in order to apply the proposed methodology.

- It is assumed that trains are scheduled to depart at the *beginning* of the minute indicated on the timetable (*e.g.* if the timetable shows a departure at 08:00, the train is expected to depart at 08:00:00).
- It is assumed that an Oyster journeys cannot be assigned to a departure scheduled for the same minute as their respective entry transactions (*e.g.* a passenger recorded to have entered a station at 08:00 cannot be assigned to a train that is scheduled to depart at 08:00).
- The walking time between station entry/exit gatelines and train platforms is negligible. In other words, incidence to a station constitutes incidence to the services at that station.

The first two assumptions are justified. Bratton (2009) indicated that the first assumption reflects the convention understood by British railways and their passengers about the meaning of times in published timetables. Since Oyster timestamp data are truncated, most transactions recorded on a given minute will actually have occurred over the course of (rather than at the beginning of) that minute – after the understood departure time of any departure at that minute in the timetable.

The final assumption is motivated primarily by the final two data limitations stated above. That said, it is largely justified in that (i) at most gated Overground stations, access times between gates and platforms are generally relatively short (*i.e.* less than one minute), and

(ii) at ungated stations, passengers use Oyster validators directly on platforms. That said, for gated stations with substantial access distances this assumption may introduce a bias into the estimations of incidence time. Specifically, passengers may appear to be incident to particular scheduled services earlier than they actually were (*i.e.* have an additional "safety margin" equal to the access time).

Open Standards and Open Source Software

The difficulty of implementing Algorithm 6.1 is in practice a function of the difficulty of implementing the Path() function. Fortunately, a robust implementation of such a function is available in the free/open source software library Graphserver (Graphserver, 2009). This package was developed primarily to support web-based public transport and multi-modal journey planner applications, which depend on exactly the same sort of Path() function as needed in this research. Graphserver has the added advantage that it can import multiple timetables for different days of the week or times of year and will choose between them appropriately depending on the date and time of passenger incidence. This is particularly useful if analyzing a sample of data that covers multiple timetables, for example when studying *changes* in passenger incidence on a given network over time.

Graphserver reads timetables in the widely used General Transit Feed Specification (GTFS) (Google, 2009). This specification was defined by Google to facilitate transfer of public transport schedules from operators to Google to power its own web-based journey planning software. It has become a de-facto standard for public distribution of public transport timetables. Unfortunately, Overground timetables do not (yet) come in GTFS format, so a simple (302-line) bespoke Perl (Perl.org, 2010) script was written to do the transformation.

Another simple (272-line) Perl script was written to implement Algorithm 6.1 for one or many individual Oyster passenger journey records (using Graphserver for the Path() function). This approach minimized the amount of complex network models and algorithms that needed to be implemented, instead taking advantage of an existing piece of free, fast, robust, and well-supported software.

6.3 Passenger Incidence Behavior on the London Overground Network

This section examines passenger incidence behavior on the London Overground network using a large sample of passenger journey data from the Oyster smartcard ticketing system. Section 6.3.1 describes the sample of data; Sections 6.3.2 and 6.3.3 present, validate, and analyze the results of applying the proposed methodology to the given data. Some of the results in this section are interpreted further in the following chapter.

6.3.1 Data

The data analyzed here is a 100% sample of all Oyster journeys between all pairs of London Overground stations for the 52 business days from 31 March, 2008 through 10 June, 2008, inclusive. This chapter is concerned primarily with the behavior of Overground passengers, and public timetables were obtained only for the Overground network. Consequently, the data set was filtered to include only those journeys for which it can be assumed with relative certainty that the passenger in question used only Overground services. This filtration was accomplished using the outputs of the assignment model of Chapter 5. The resulting data set contains nearly 1,670,000 journeys from 54 stations on 1,442 origin-destination pairs made by over 290,000 passengers. It constitutes approximately 53,000 station-hours of observation of passenger incidence to Overground services.

The methodology described in the previous section was applied to each Oyster journey in the data set. This processing took some number of hours for the entire data set, but was entirely automated. It results in a large set of observations for which the following are measured or estimated:

- the date and time of incidence,
- the location (*i.e.* station) of incidence (*i.e.* the journey's origin),
- the journey's destination,
- the scheduled waiting time (SWT),
- the incidence headway,
- the Overground line to which the passenger was incident.

For Overground-only journeys that require an interchange (of which there are relatively few), the above is measured or estimated for only the first incidence event. It should be noted that this data set does not include journeys that interchanged *to* the Overground but with initial Oyster validations elsewhere in the railway system (*e.g.* Underground passengers interchanging from the London Underground's Central or Jubilee lines to the North London Line at Stratford). Nor does it include journeys initially incident to the Overground but interchanging to other railway services before final validation. Journeys with interchanges to and/or from buses will be included here, since the Oyster system effectively separates the recording of bus and rail journeys.

6.3.2 Validation

As a point of validation, Figure 6-3 plots distributions of incidence headway for passengers on the Gospel Oak to Barking (GOB) and North London (NLL) lines. The findings are consistent with expectations. On the NLL, the mode of all the distributions is 15 minutes, reflecting the core service. The distribution is more concentrated during the Inter-Peak period, when there are no scheduled "shuttles" or "specials." The opposite is true on the GOB which runs a regular 20 minute service in the peak periods but transitions to and from a 30 minute service in the Inter-Peak period. The AM Peak distribution is somewhat more dispersed than that of the PM Peak because it includes a transition from 30 minute headways in the early morning.

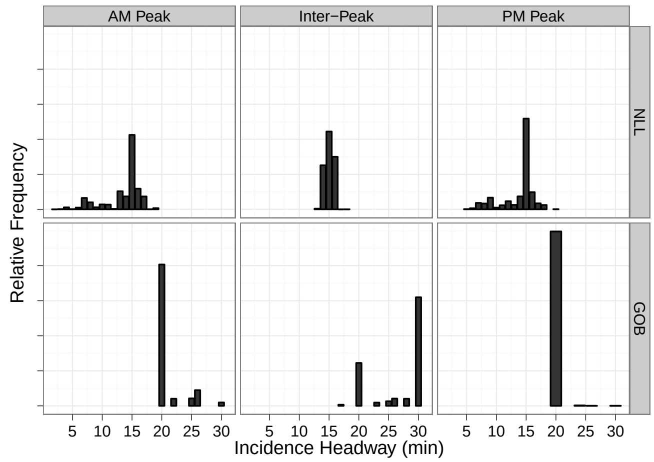


Figure 6-3: Distributions of passenger incidence headways, by line and time period

Some values in Figure 6-3, for example 3 minutes on the NLL in the AM Peak, are positive but observed very infrequently. This is because of slight variations in the timetable where a pair of services that are, for example, 4 minutes apart at most stations are only 3 minutes apart at one or two stations.

6.3.3 Results

Consistent with the reviewed studies of passenger incidence behavior, the first results of interest are distributions of passenger incidence time over a given headway. Figure 6-4 plots these distributions for the London Overground network by line and by time period. In this plot, incidence times are normalized by the incidence headway because different passengers experience different incidence headways. Prior to this normalization, incidence times (from which seconds were truncated by the Oyster system) are randomly perturbed by between 0 and 59 seconds so as to smooth the plots.

It is clear from Figure 6-4 that passenger incidence behavior, with respect to published timetables, varies spatially and temporally across the Overground network. For example, passenger incidence during AM Peak commuting hours is much more peaked (*i.e.* timetable-dependent) on the GOB and Watford DC (WAT) lines, each with 20 minute headways, than on the NLL, with 7-15 minute headways. Also, the NLL is acknowledged by Overground management to have the most serious reliability problems (Bratton, 2008). These variations are generally consistent with what has been found in the literature – that passenger incidence is more timetable-dependent with a more peaked distribution during in longer headways and for more reliable services.

Also consistent with the literature is that, for all lines, the distribution is more peaked in the AM Peak period than in the PM Peak or midday Inter-Peak periods. It appears that Overground commuters in the AM Peak period, likely with knowledge of the timetable and the service, exhibit less random incidence behavior than passengers in other time periods despite the shorter headways and less reliable service found in the AM peak.

The peaks of all of the distributions are somewhat before the very end of the headway,

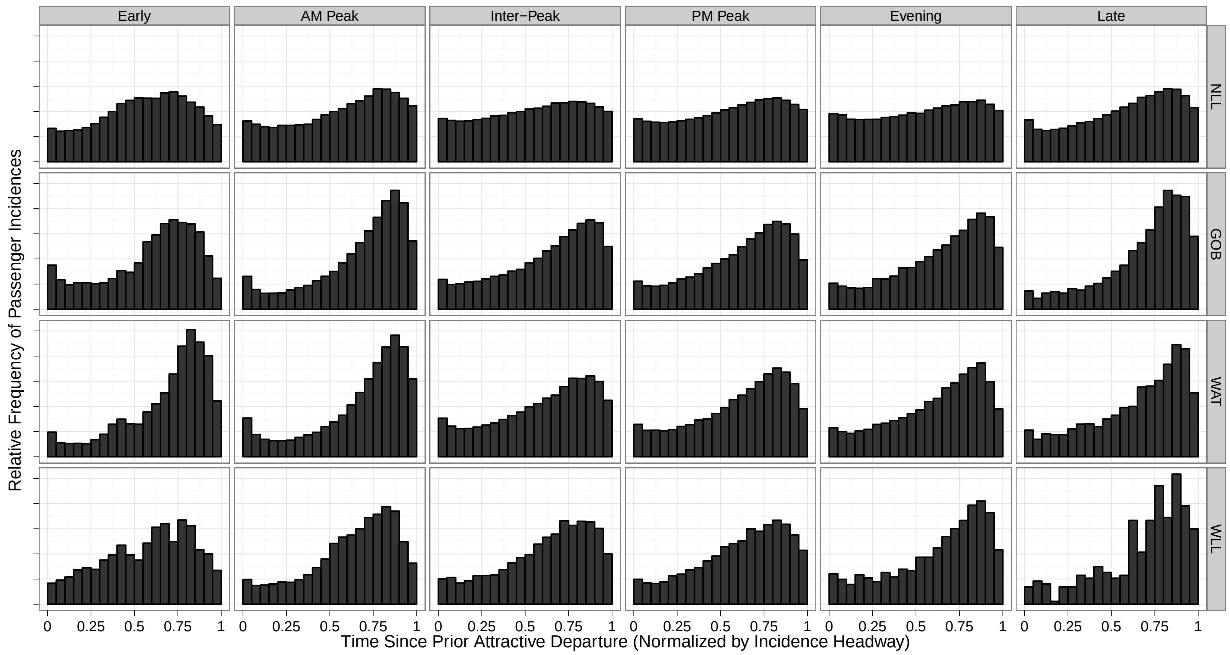


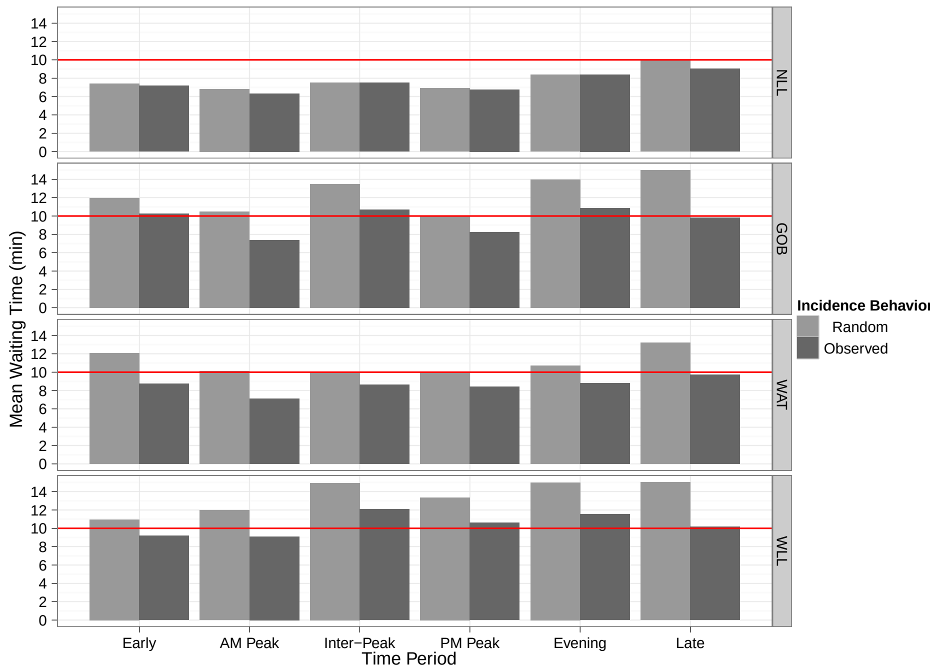
Figure 6-4: Distributions of passenger incidence, by London Overground line and time period

indicating some type of "safety margin" or waiting time-minimization behavior on the part of passengers. Many of the distributions have small spikes at the beginning of the headway, indicating possible "late running awareness" among some passengers. While such awareness may in fact be found on the Overground, it is also possible that it is the passengers themselves who are running late. They may be planning to take a train scheduled to depart at a certain time but, because of uncontrollable circumstances or just their own poor planning, arrive at the station shortly after that departure time.

Figure 6-5 shows the mean waiting time, for each line and time period, under two different models of passenger behavior and train operations. First, it is assumed that all passengers are randomly incident to constant headway services, so the mean waiting time is calculated as half the mean incidence headway. Second, no behavioral assumption is made but all trains are assumed to run as per the timetable, so the "Observed" mean waiting time is calculated as simply the mean SWT. Figure 6-5 thus indicates the effects of the observed incidence behaviors (compared with random incidence) on passenger waiting times.

On the NLL, the relatively slight skew in the incidence distributions translate into relatively small impacts on waiting time. In the AM Peak timetable dependence decreases waiting time by 7.2% from 6.82 minutes to 6.33 minutes (about 30 seconds). In the Inter-Peak and PM Peak periods, the reductions are only 0.2% and 2.2%, respectively. On the GOB, on the other hand, the implications of timetable-dependence are substantial. In the AM Peak, the waiting time decreases by 29% from from 10.5 minutes to 7.4 minutes (3.1 minutes). In the Inter-Peak and PM Peak periods, the reductions are still 20.4% and 17.5%, respectively.

In the model for average waiting time used in the assignment model of Section 5.3.1, no service would be assigned a mean passenger waiting time over 10 minutes. The results



Red horizontal line highlights a waiting time of 10 minutes

Figure 6-5: Mean scheduled passenger waiting time by London Overground line and time period

of this section lend support to that model. Regular weekday headways on the Overground network go as high as 30 minutes, but the observed mean scheduled waiting time is above 11 minutes in only two cases (on the WLL during the Inter-Peak and Evening time periods) and is never above 12.1 minutes.

6.4 Conclusions and Recommendations

This chapter developed a methodology to relate disaggregate AFC journey data to published timetables for the purpose of studying passenger incidence behavior, and applied this methodology to the London Overground. This section presents first some conclusions drawn from the analysis in this chapter, and next some recommendations based on those conclusions.

6.4.1 Conclusions

The following conclusions are drawn about the methodology developed in this chapter. Firstly, that it can be used to study passenger incidence behavior using large samples of disaggregate journey data from AFC systems such as the Oyster smartcard system. It is able to efficiently process thousands or millions of such data records. Secondly, that it can for each passenger journey estimate scheduled waiting time and incidence headway, two of the most important quantities for studying passenger incidence, even under quite heterogeneous conditions. This estimation is dependent on certain assumptions that are believed to be reasonable for the London Overground network, but may not be appropriate in all circumstances. Finally, that this methodology can be easily implemented using open standard timetable formats and free software tools.

With respect to the Overground, the following can be concluded from the results in this chapter. Broadly, that passenger incidence behavior is heterogeneous across the network and across times of day, and that the differences are broadly reflective of what has been found in the literature to date. Specifically, that incidence appears to be much less timetable-dependent on the North London Line (NLL) than on the other Overground lines. This chapter has not attempted to rigorously identify the causes of these differences. Hypotheses drawn from existing literature on the subject and knowledge of the Overground network include (i) shorter headways (*i.e.* higher frequencies) and (ii) less reliable service on this line as compared to others. On the lines with timetable-dependent incidence behavior (*i.e.* other than the NLL), passengers reduce their mean scheduled waiting time by over 3 minutes, or up to 30%, during daytime hours compared with random incidence behavior. On the NLL, such reductions are much smaller, in some cases nearly zero, in both relative and absolute terms.

6.4.2 Recommendations

The method developed in this chapter, which builds heavily on some of the basic concepts of schedule-based assignment, should be used to support further study of passenger incidence behavior. The work of Bowman and Turnquist (1981) has been influential in shaping the understanding of the relationships between headway, reliability, passenger behavior, and waiting time. Their work should be updated using the method of this chapter to easily analyze large samples of passenger data across heterogeneous networks. Their work also depends on measurements of service reliability, which should be gathered from automatic vehicle-tracking systems. The London Overground network represents an ideal opportunity to conduct such a study – its passengers can clearly be studied via Oyster data, and its trains are tracked by a computerized signaling system. Once the East London Line opens, the network will have headways ranging from 5 to 30 minutes during most hours of the day.

Nearly three decades have passed since the work of Bowman and Turnquist. In that time, many strides have been made towards informing passengers in real time about the status of public transport services. Such information is now often distributed via in-station signs and announcements as well as over the internet to passengers' computers and, more importantly, mobile devices. It is crucial to advance the understanding of passenger incidence to include the effects of real-time information. This requires careful thinking and research designs, but should be able to take advantage of the methodology developed here.

The methodology and results of this chapter, in particular the disaggregate application of schedule-based assignment and the degree of timetable-independence of incidence behavior on the North London Line, have certain service quality measurement and tactical planning implications for the Overground. These are explored in the following chapters.

Chapter 7

Service Quality Measurement from AFC Data

The Transit[1] Capacity and Quality of Service Manual defines *service quality*[2] as "the overall measured or perceived performance of [public transport] service from the passenger's point of view" (Kittelson & Associates, Inc et al., 2003*a*). A related guidebook (Kittelson & Associates, Inc et al., 2003*b*) points out that the public transport agency or operator[3] often perceives system performance from a different perspective, one more concerned with the quality of the operations than of the service as experienced by passengers. It defines *service delivery* in terms of how well "an agency deliver[s] the service it promises on a day-to-day basis." In this and the following chapters, *service quality* will refer to the passenger's perspective on system performance while *service delivery* will refer to the operator's perspective. As an example of the difference, consider a hypothetical "right-time" railway which, despite running every train exactly to the timetable, happens to have insufficient capacity to serve all of its demand at all times. The railway may consider its performance to be perfect, but passengers riding cheek-to-jowl or on occasion left standing on the platform would likely perceive the situation differently.

This discrepancy motivates the study of service quality itself, which is the subject of this and the following chapter. Some aspects of service quality have traditionally been *modeled* using operational data such as vehicle movement records (*e.g.* Wilson et al., 1992). With the introduction of automatic fare collection (AFC) systems and the data they produce about individual passenger journeys, it is now possible to *measure* certain aspects of service quality directly. Some AFC systems (*e.g.* Oyster) control entry to and exit from the public transport network. In this case, actual passenger journey time through the network can be estimated as the difference between the timestamps of the exit and entry transactions.

Direct and automatic observation of passenger journey times creates many opportunities for measuring service quality. One particular measure that is explored in this chapter is *excess journey time* (EJT). At the level of a single journey, EJT is the difference between actual journey time and some pre-defined journey time standard, however that standard is

[1] Public transport is also known, particularly in the United States, as "transit."

[2] They use "quality of service."

[3] *Operator* here refers to the operating company or organization, not, unless otherwise stated, to the individual vehicle driver.

defined. A positive value indicates that the journey took longer than the standard allows; a negative value indicates that it was shorter.

Section 7.1 reviews some of the literature on public transport service delivery and service quality measurement, including EJT. Section 7.2 discusses the implications of incidence behavior for establishing EJT standards, and proposes an analytical framework with which to analyze EJT under different incidence behaviors. Section 7.3 uses a rigorous probabilistic analysis to prove, under the proposed framework, that a single means for measuring *aggregate* EJT appropriately can accommodate a range of incidence behaviors. Section 7.4 discusses some important considerations for applying EJT in practice, including the interpretation of individual and aggregate negative EJT measurements. Section 7.5 draws conclusions, and Chapter 8 applies the method developed here to the London Overground network.

7.1 Service Delivery and Service Quality Measurement Literature Review

The literature on measurement of public transport performance, including service delivery and service quality, is rich. Interest in the subject has renewed since the introduction of systems for automatically monitoring various aspects of public transport operations and, most recently, public transport passenger journeys. This section has a dialectic quality, which should not come as a surprise since it is about two subjects which are formulated as perspectives of two very different entities. Section 7.1.1 and Section 7.1.2 discuss measures of service delivery and service quality, respectively, which are found to be somewhat at odds with each other. Section 7.1.3 discusses measures of *relative service quality*, which resolve some of the tensions between pure measures of service delivery and service quality. Excess journey time (EJT), the subject of the final sections of this chapter, is one such measure.

Before discussing the literature on specific measures of service delivery and service quality (*i.e.* from the operator's and passenger's perspectives) this section discusses first the notion of reliability, and next some of the goals behind conducting such measurement in the first place.

Reliability

Reliability is a much-discussed topic in academia as well as in real world management of public transport systems. Abkowitz et al. (1978) conducted a seminal study on reliability, which they define as "the invariability of service attributes which influence the decisions of travelers and transportation providers." The definition offered by Abkowitz et al. is useful in the context of this chapter for three reasons. First, it focuses the discussion on attributes of public transport service outcomes, rather than inputs (*e.g.* mechanical components, staff) which may have their own measures of reliability. Secondly, it defines reliability in terms of the higher-order moments (*i.e.* "variability") of these attributes of service outcomes. Lastly, it indicates that these attributes are experienced by passengers and/or operators.

Under this definition, the discussion of reliability is quite naturally subsumed by discussions of service delivery and quality *if and when* they consider higher-order moments of the attributes perceived by operators and by passengers, respectively. Consequently, much of what has been said about reliability applies to both service delivery and quality, and thus applies to the balance of this chapter and indeed this thesis.

Abkowitz et al. go on to investigate public transport service reliability from a number of different angles, including (i) how it is perceived by passengers and affects their attitudes and behavior, (ii) how it is perceived and acted upon by operators, (iii) reasons and ways to measure it, (iv) the factors that affect it, and (v) strategies to improve it. This study appears to have set the stage for much of the work on the subject in the intervening decades, some of which is described here. Uniman (2009) provides a detailed review of the work by these and other authors on the subject of reliability.

Goals and Applications of Service Delivery and Quality Measurement

A guidebook on public transport performance measurement prepared for the US Federal Transit Administration (Kittelson & Associates, Inc et al., 2003*b*) notes that public transport providers measure their performance, broadly defined, because (i) regulation requires it, (ii) it is useful for internal management purposes, and (iii) external stakeholders, including the riding public and funding bodies, depend on accurate information to support advocacy and decision-making processes. This chapter and thesis are primarily concerned with the second of these reasons – internal management, including tactical planning. The guidebook also describes many other stakeholders, motivations, goals, guidelines, and outcomes related to public transport performance measurement which are beyond the scope of this thesis.

Understanding "reliability" as a proxy for overall performance, including service delivery and service quality, Abkowitz et al. note that measuring performance from the operator's and passenger's perspectives should help public transport providers to:[4] "(i) identify and understand performance problems; (ii) identify and measure actual improvements in performance; (iii) relate such improvements to particular strategies; (iv) modify strategies, methods, designs to obtain greater performance improvements." In the context of this thesis, this description is useful in that it establishes the measurement of service delivery and service quality as elements of an *iterative* analytical management and planning process.

Cham (2006), also studying reliability, distinguishes between two primary means by which public transport providers should be able to improve performance through the use of automatic data sources. Firstly, through improved monitoring and direct management tasks, including evaluation of operational staff. Secondly, through improved tactical planning, which is the focus of this thesis.

7.1.1 Service Delivery Measurement and The Operator's Perspective

Kittelson & Associates, Inc et al. present a wide array of measures covering different aspects of public transport system performance. Many of these measures, particularly those that are not concerned directly with transport service outcomes, are beyond the scope of this thesis. Those related to transport service outcomes are typically relative in that they indicate the

[4] In this quotation, "performance" is substituted for "reliability," consistent with the understanding of reliability as one aspect of overall performance.

degree to which service delivery adhered to the service plan. Furth et al. (2006) describe two such classes of service delivery measures that can be developed from automatic data sources – those measuring adherence to timetables and those measuring adherence to headways.

Timetable-Based Measures

Timetable-based measures are often based on observations of *schedule deviation* – the difference, for a given service, between the scheduled and actual time of arriving, passing, or departing a given timepoint.[5] The most popular measure of timetable adherence is *on-time performance* (OTP), the fraction of services with schedule deviation within some threshold (Kittelson & Associates, Inc et al., 2003*b*). Under the name of Public Performance Measure (PPM), this is the current measure of performance on the London Overground and all other National Rail services in the UK, with a train considered "on time" if it is less than 5 minutes late at the destination terminal (Office of Rail Regulation, 2008). London Buses also measures OTP, with an on-time window of 2.5 minutes early to 5 minutes late, for "low frequency"[6] routes (Camilletti, 2003).

Henderson et al. (1990) and Henderson, Adkins and Kwong (1991) offer a number of criticisms of OTP, primarily for its lack of passenger orientation. Among these criticisms are (i) OTP measures performance at terminals, which for many networks are remote from the locations to which most passengers are bound, (ii) OTP typically counts as late services which have missed part of their trip or skipped stops, even if these adjustments don't affect many passengers, (iii) passenger waiting times are not accurately reflected, (iv) focusing on OTP can incentivize dispatch actions that favor schedule adherence over regular headways and can make passengers worse off, and (v) OTP, while offering a probabilistic measure, does not represent the odds of on-time arrival realistically.

A related measure is terminal-to-terminal running time. Statistics of the distribution of running time indicate the reliability of a service and are important inputs into the scheduling process. When running times are too short, some vehicles will not be in place to make subsequent trips; when they are too long, resources are not used efficiently and terminals may be congested (Furth et al., 2006; Furth and Muller, 2007) (see Rahbee, 1999, 2006, for studies of the Boston and Chicago metro systems, respectively).

Another common timetable-based service delivery measure is en-route schedule adherence (ESA), which can be defined as the fraction of services with schedule deviation within some threshold at a given set of timepoints. This is similar to OTP, but applied at multiple points along a line. The distribution of schedule deviation and segment (*i.e.* point-to-point) running times can also be studied (Furth et al., 2006; Hammerle et al., 2005) (see Cham, 2006, for a study of Boston bus services).

ESA (or lack thereof) is not necessarily a problem for passengers *per se*, since, as shown in Chapter 6, in some cases they can adjust their incidence behavior based on their knowledge of observed system performance. Passengers with non-random incidence behavior can adjust their behavior to account for predictably late (or early) services, while random incidence passengers may not notice any problem at all if headways remain even because *every* service

[5] A *timepoint* is simply a location at which service arrivals, passings, or departures are timed. Timepoints are used in scheduling as well as in performance measurement.

[6] London Buses' distinction between low and high frequency routes is discussed in Section 7.1.4.

is late (or early) by the same amount. It is when the degree of ESA is unpredictable (*i.e.* the service is unreliable) that passengers suffer the most negative effects. However, to truly understand those negative effects, the situation must be considered from the passenger's perspective.

These measures of schedule adherence treat each service in isolation, ignoring the attributes of service (waiting time) that depend on the headways between successive services at a given location (Reddy et al., 2009). In some circumstances the headways are more important than the specific arrival and/or departure times in the timetables, thus motivating the headway-based measures described in the following section.

Headway-Based Measures

Some public transport networks or lines publish service headways but not timetables. In some cases, particularly for higher frequency services, it is assumed that passengers do not use the timetable even if it is available (Kittelson & Associates, Inc et al., 2003*a*). As discussed in Chapter 6, mean headway and variability of headway both affect waiting times, particularly for randomly-incident passengers. For these reasons, headway is typically evaluated in terms of regularity, which can be defined in a number of ways (Furth et al., 2006).

Kittelson & Associates, Inc et al. recommend measuring the mean observed headway and the coefficient of variation with respect to the mean scheduled headway. Henderson, Kwong and Heba (1991) proposes two measures of headway regularity, one based on Gini's ratio and the other based on the coefficient of variation, that have the benefit of being normalized on a zero to one scale for comparison across services with different mean headways. These measures are all unitless, and thus hard to interpret in physical terms relevant to operators or passengers (Furth et al., 2006).

Reddy et al. (2009) and Hammerle et al. (2005) define headway regularity in terms of the fraction of observed headways that are within some absolute or relative deviation from the scheduled headway. These have the benefit of being easy to interpret by operators, but still fail to translate easily into passenger terms (Furth et al., 2006).

The adoption of headway-based measures is motivated by the effect of headways on passenger waiting times, and so is a real step towards representing the passenger's perspective. Nevertheless, they are still an indirect proxy for the passenger experience, since waiting times are related to but not equal to headways. Moreover, headway-based measures do not account for the entire duration of passenger journeys, which are important. Finally, as discussed in Chapter 6, headway at a given location depends on which services one is willing to board at that location (*e.g.* for trunk-and-branch services), which depends on where one is headed – headways cannot be accurately measured without considering the passenger's perspective.

7.1.2 Service Quality Measurement and The Passenger's Perspective

Service quality has many aspects, some easier to measure than others. Some aspects, for example those related to travel or waiting times, can be expressed in quantitative terms, while others, such as the customer experience dealing with staff, are more clearly understood qualitatively (Kittelson & Associates, Inc et al., 2003*a*). Qualitative aspects are typically measured through surveys (*e.g.* Lu et al., 2009). The quantitative aspects of service quality can be interpreted objectively or subjectively. Total journey time would be an objective quantity, whereas journey time weighted to reflect passenger preferences would be a subjective one.

Strictly speaking, service quality is absolute in nature, at least with respect to service delivery. For example, the service quality of a public transport network can be judged on its waiting and travel times. Even when every passenger experiences perfect service delivery – more frequent and faster service is always better. The work described in this section seeks to measure service quality in absolute terms.

For randomly incident passengers, Osuna and Newell (1972) describe how mean waiting times can be modeled given observations of actual headways. Friedman (1976) extends this result to model the variance of waiting times (see Section 6.1.1 for more information on both). Larson and Odoni (2007) describe how the complete distribution of waiting times of randomly incident passengers can be derived from headway observations.

Bates et al. (2001) provide an in-depth investigation of how passengers value reliability (expressed as the variability of total journey time) and how it may affect their behavior. Furth and Muller (2006) operationalize some of this analysis by proposing to measure the effect of reliability as additional waiting time costs perceived by passengers. Their analysis is based in part on the literature discussed in Chapter 6 which found that passengers adjust their incidence behavior based on knowledge of schedule and headway adherence and reliability. For short headway services (on which they assume all passengers are randomly incident), they propose to use headway observations to measure the "potential waiting time" as the difference between the "budgeted" 95^{th} percentile waiting time and the mean waiting time. This is intended to represent additional waiting time (with respect to the average case under observed circumstances) that passengers would have to budget to assure late arrival at their destination on at most 5% of trips (assuming en-route travel times are constant). It represents an additional penalty that passengers pay for the unreliability of the service headways, albeit a penalty paid in most cases by arriving early at their destination. A similar measure is developed with respect to mean and 95^{th} percentile schedule deviation for timetable-dependent passengers.

Chan (2007) and Wilson et al. (2008) extend the potential waiting time concept to the entire journey. They use data from the Oyster smartcard ticketing system to measure (rather than model) the journey times of London Underground passengers. They estimate the distribution of end-to-end journey times for each origin-destination (OD) station pair and find the "reliability buffer time" (RBT) [7] as the difference between the 95^{th} and 50^{th} (median) percentiles. This metric is aggregated from the OD pair to the line or network level by means of an OD flow-weighted average. It is interpreted similarly to the measure proposed by Furth and Muller (2006) but with some advantages. Chief among those advantages are that it (i) analyzes the entire journey, rather than just the waiting portion, and (ii) automatically and accurately accounts for the effects of congestion, including delays suffered by passengers who suffer additional waiting time because of trains that are too crowded to board. Indeed, these are advantages of any approach using direct measurements of observed journey time (OJT).

[7] These authors used the term "reliability factor;" Uniman (2009) later used "reliability buffer time," a more fitting term, which is used here.

Uniman (2009) notes that some "irreducible" amount of variability in passenger journey times is to be expected because of randomness in waiting times, variation in walking speeds, and normal but acceptable variability in service outcomes. Uniman proposes to divide observation periods into two classes of reliability levels – "recurrent" and "incident-related." Passengers experience normal levels of journey time variability in the former, and abnormal levels in the latter. Also studying the London Underground, Uniman makes this classification using a statistical technique that did not consider the perspectives of the managers of the system under study. Uniman then proposes as a measure of service quality "excess reliability buffer time" (ERBT) – the difference in RBT for journeys from all observation periods together and RBT for only those journeys in periods of recurrent reliability. In other words, a measure of how far the tail of the travel time distribution is extended as a result of abnormal operating conditions.

Uniman proposes ERBT as an attempt to create insight into the causes of unreliability by isolating the effects of incidents. However, it is not clear that the world can be so easily and neatly divided into recurrent or incident-related conditions. The author of this thesis spent several months working in the control center of the New York City Subway, and observed first hand that incidents on a continuum of severity occur constantly. It is often unclear whether or not some perturbation to the service is normal and what in fact constitutes an "incident." Moreover, this measure applies cleanly only to randomly incident passengers. Additional research is necessary to understand how to apply it to passengers whose incidence is dependent on the timetable and/or experience with observed departure times.

The measures discussed in this section are developed entirely with reference to actual operating conditions and passenger journeys, not with reference any service delivery commitments (*i.e.* the timetable, and headways and travel times implied therein). This author has found no evidence that any of these measures are regularly used in practice by public transport providers.

Such measures may not yet have been adopted because they do not provide information in terms that operators can easily relate to. That said, they still have a place in the tactical planning process, particularly when timetables are changed. In this case, measures of absolute service quality may be the only way to meaningfully capture the effects on passengers of a particular tactical planning intervention.

7.1.3 Relative Service Quality

A compromise between measures of service delivery and measures of service quality is described in this section, which measures what is referred to here as *relative service quality.* They measure service quality not in absolute terms, but rather with respect to certain standards derived from service delivery commitments (*e.g.* the timetable).

Wilson et al. (1992) estimate mean passenger waiting time from automatic headway measurements in the Boston subway system using the random incidence model described in Section 6.1.1. They estimate the mean waiting time according to the scheduled headways. The difference between these two estimates, called "excess waiting time" (EWT), indicates the waiting time experienced by passengers above and beyond what they would have waited had all headways been exactly as scheduled. London Buses uses this measure to monitor performance of high frequency routes (Camilletti, 1998). Historically, it used manual surveyors to measure headways for estimating EWT at selected points in the network. With the delivery of the new GPS-based "iBus" system, EWT will be calculated automatically (London Buses, 2009).

London Transport (1999) extended the EWT concept to the entirety of journeys on the London Underground, comparing mean actual and schedule values of each component of passenger journeys. Automatic data from train control systems is used to estimate EWT, as in Wilson et al. (1992), under random incidence model. The EWT estimate is augmented by models for estimating the fraction of passengers, based on static demand data, who are left behind by overcrowded trains. Automatic train movement data is also used to estimate excess on-train time, where the scheduled on-train time between any given pair of stations is as per the timetable. Manual sampling at 27 major stations is combined with pedestrian flow models to estimate access, egress, and interchange (*i.e.* walking) time as a function of pedestrian congestion and availability of escalators and elevators. The scheduled values for pedestrian movements are determined from manual samples under free-flow conditions.

The sum of these components is referred to as "excess journey time" (EJT). It is estimated in unweighted and weighted forms, where weights indicate relative passenger preferences as described in Chapter 5. Line or network level average EJT is weighted by static estimates of passenger flows. This approach to estimating EJT depends on a number of models to characterize the passenger experience from a range of direct automatic operational measurements and a small number of samples of pedestrian conditions. Despite the complexity of the system used to estimate it, EJT is perceived as easy to interpret and is used to this day as the primary means of performance measurement and management on the London Underground.

Chan (2007) and Wilson et al. (2008) use Oyster journey data to directly estimate (rather than model) unweighted EJT for individual journeys on the London Underground. They measure actual journey times directly from Oyster transactions, and derive scheduled journey time from the values in the Underground's existing EJT measurement system (continuing to use the random incidence model to derive mean scheduled waiting time). As discussed in the previous section, this approach automatically and accurately accounts for the effects of congestion on passenger journey times. For reasons that are unclear to this author, they use twice the scheduled waiting time as in the Underground's system (*i.e.* one full headway), but exclude all Oyster journeys with measured journey times less than the respective scheduled value.[8]

Chan and Wilson et al. estimate Oyster-based EJT for single-line journeys (to avoid ambiguity of journeys with multiple routing options through the Underground network). They estimate mean aggregate EJT at the line level, finding that these estimates do not correspond to the model-based estimates from the London Underground. While they do not reach any definitive conclusions as to why this would be, there is no reason to believe that Oyster-based estimates of actual journey time should be less accurate than those derived from a set of complex models and a variety of operational data sources and random samples. Consequently, there is reason to believe that, given common scheduled journey time values, unweighted Oyster-based EJT will be more accurate than model-based estimates. In any

[8] The scheduled time for the full journey is simply the sum of the scheduled values of the individual components (*i.e.* walking, waiting, on-train).

case, this work estimates service quality for individual passenger journeys relative to the timetable based on the assumption of random passenger incidence behavior.

Buneman (1984) uses schedule-based assignment (reviewed in Chapter 6) to estimate *passenger* on-time performance for the BART railway network in the San Francisco Bay Area. On that network, at that time, the AFC system recorded the origin and destination station of each passenger journey, but only the time of exit (not of entry). The network's train control systems also recorded all train movements. Buneman uses these data sources to conduct, for each passenger, a schedule-based assignment *in reverse* with respect to *actual train movements* (rather than the timetable). This results in estimates of the times that passenger departed their respective origin stations. Then, based on some hybrid assumptions about passenger incidence behavior (*i.e.* 25% timetable-dependent, the rest random) and known vehicle departures, he estimates probabilistic times of entry into the system. Given these times, a schedule-based assignment with respect to *the timetable* yields probabilistic scheduled arrival times for each passenger journey at its respective destination. The difference between actual and scheduled arrival time, in the parlance of this section, is an estimate of schedule-based EJT.[9] Buneman does not calculate aggregate EJT, but rather compares EJT to a 5-minute threshold window to estimate passenger OTP. It appears that this measure, perhaps in a modified form, is still used by BART over two decades later (BART, 2010).

All of the measures of relative service quality discussed in this section were developed with the intent of representing the passenger's perspective. However, they all make certain assumptions about passenger incidence behavior, from which they derive the standards against which measured or modeled service quality is compared. The next section assess the state of practice for service quality measurement, including these assumptions, and sets the stage for the method proposed in the following sections of this chapter.

7.1.4 Discussion

The primary client of public transport performance measurements are the managers and planners of the transport networks themselves. Ideally, they should be motivated to improve the service quality as experienced by their primary customers, the passengers. However, the levers over which they have the most tangible understanding and direct control are planning and operational service delivery. Consequently, it is proposed that measures of public transport performance should find a balance between the passenger's and operator's perspectives. They should strive for fidelity to the passenger experience, but not so far that they are not useful or interpretable by operators.

One benefit of the service delivery measures described in Section 7.1.1 is that they are easily interpreted, and hopefully acted upon, by transport operators who deal in schedules and headways on a daily basis. Operators can act on these measurements to improve system performance. However, these service delivery measures can be and have been criticized for a lack of passenger orientation; there is no guarantee that a certain improvement in operational service delivery will benefit passengers proportionally. It is with this recognition that the measures of service quality presented in Section 7.1.2 have been developed. These measures

[9] Had AFC transaction times at origin stations been available as they are in the Oyster system, Buneman would likely have opted for a single forward schedule-based assignment as described in Chapter 6.

maximize fidelity to the actual experiences of passengers, but to such a degree that they become difficult or impossible for operators to interpret or use.

The measures of relative service quality described in Section 7.1.3 represent a compromise between the pure operator and passenger perspectives. One of these measures, excess journey time (EJT), has found lasting application in large urban railways such as the London Underground and BART networks. It presents a compelling alternative to the train on-time performance (OTP) measure currently used by the London Overground. It measures the actual passenger experience in terms of end-to-end journey time, but reports it with respect to certain service quality standards. In the work reviewed here, those standards are derived from the timetable. There is a definite risk to this approach – in reviewing the work of Abkowitz et al. (1978), Uniman (2009) points out that measures using the timetable as a standard can be misleading because "a change in the [timetable] might artificially lead to measured improvements in performance, without any changes perceived by passengers." In other words, relative measures of service quality may be difficult to use for longitudinal analysis of timetable revisions.

On the other hand a standard based purely on the passenger experience could be overly generous to operators. As discussed by Bates et al. (2001) and in Chapter 6, passengers often adjust their behavior to account for their knowledge of service delivery. In the worst case, this could lead to a feedback cycle where poor service delivery results in a worsened passenger experience which leads to a looser standard, and so on – the so-called "crumbling edge of quality."

It is likely that, in practice, measures of relative and of absolute service quality are both useful in tactical planning, especially since one of the primary outcomes of tactical planning is changes to the timetable. Relative measures should provide cross sectional understanding relative to the existing timetable, and absolute measures should be useful for longitudinal analysis of changes to the timetable.

Given AFC data such as that produced by the Oyster smartcard system, the average observed journey time, a measure of absolute service quality, is straightfoward to estimate. Still remaining is how exactly to use the timetable to derive standards, against which to measure relative service quality, which reflect realistic passenger expectations to the maximum degree possible. This is, in part, a question of passenger incidence and of what can be assumed about passenger's expectations based on their incidence behavior.

Journey Time Standards and Passenger Incidence Behavior

The goal of measuring EJT is to most faithfully represent the passenger's perspective and experience while supporting the efforts of operators to improve service delivery and quality. To that end, measurement of EJT has two requirements:

1. to accurately estimate actual passenger journey times,
2. to set timetable-based journey time standards that match passengers' expectations as closely as possible.

With the advent of AFC ticketing systems, actual journey times can now be measured simply and directly as in Wilson et al. (2008) and Uniman (2009). One issue that remains unresolved, particularly as EJT is applied to networks with lower service frequencies, is how passenger behavior and expectations relate to the published or unpublished timetable, and thus how the timetable should be used in setting journey time standards.

Industry manuals (*e.g.* Kittelson & Associates, Inc et al., 2003*b*; Furth et al., 2006) typically recommend timetable-based measures for lower frequency services with a headway greater than 10 minutes, where passenger incidence is assumed to be timetable-dependent, and headway-based measures for higher frequency (*i.e.* shorter headway) services, where passenger incidence is assumed to be random. London Buses, for example, follows this pattern, classifying bus routes as "high frequency" at frequencies of 5 or more buses per hour (a 12 minute or lower headway), and "low frequency" otherwise (Camilletti, 1998, 2003).

Most of the relative service quality measures discussed here, including EJT on the London Underground, use the random incidence assumption to derive waiting time standards. The model of Buneman utilizes mixed assumptions about passenger incidence behavior to derive waiting time standards, but he acknowledges that they are arbitrary. These various approaches depend, explicitly or implicitly, on assumptions regarding how passengers' knowledge of the timetable affects their arrival behavior at rail stations and their expectations of waiting and travel time (and distributions thereof).

The stated intent of these recommendations and practices is to match journey time standard to the concerns, experiences, and expectations of passengers. The standards against which measured or modeled service quality is compared have been explicitly derived from these simplifying assumptions about passenger incidence behavior. However, as discussed in Chapter 6, passenger incidence behaviors, let alone passenger expectations, are in many cases not so clear cut. As has been shown, it is possible to have a mix of timetable-dependent and timetable-independent passengers using the same service at the same time. In cases when behavior is homogeneous across some segments of passengers (*e.g.* those traveling between a given pair of stations), it still possible to have varying conditions across the network or even at a given station. Trunk-and-branch services, which provide different service frequencies to different passengers, are a prime example. Moreover, incidence behaviors are likely to change over time as a function of changes in relevant attributes of the service (*e.g.* headway and reliability). Even where the random incidence assumption has historically been justified by a lack of posted timetables (*e.g.* the London Underground), the reality may be changing as a result of internet and mobile delivery of timetable information.

This heterogeneity of incidence behavior is an additional reason, not often mentioned, that existing measures of service delivery and (absolute or relative) service quality often fail to appropriately account for the passenger's experience. It presents a particular problem in measuring EJT, where different assumptions about incidence behavior could lead to very different journey time standards. The balance of this chapter proposes and explores a methodology for estimating aggregate EJT that, it turns out, applies equally well under a range of assumptions regarding passenger incidence and implied journey time standards.

7.2 Analytical Framework and Assumptions

For clarity of exposition, the following lexical convention is adopted. The *expectation* of a given quantity refers to the expected value of that quantity in the probabilistic sense. The *standard* for a given quantity refers to some individual's supposition of what that quantity should be. Standards can be random or deterministic.

This section establishes an analytical framework for analyzing EJT. First it defines a set of mathematical quantities for representing different temporal quantities related to a given passenger journey, including time of incidence, total journey time, and the journey time standard. Next it defines values for these quantities under different passenger incidence behaviors.

In this discussion, random variables will be shown as capitals, X, known quantities as lowercase, x, and standards as capitals with tildes, $\widetilde{X}$. The following analysis considers only trips along a single line without interchanges. Judgment with respect to how much this limits the general applicability of the results should be deferred until intuition is developed through the analysis.

7.2.1 Journey Time Components and Standards

For a given passenger, let

I = the time that passenger is incident at his or her boarding station,

$\widetilde{W}$ = the standard for waiting time, also referred to as the *scheduled waiting time*,

$\widetilde{V}$ = the standard for in-vehicle travel time, also referred to as the *scheduled travel time*,

$\widetilde{A}$ = the standard arrival time at the alighting station, also referred to as the *scheduled arrival time*,

$\widetilde{J}$ = the standard for end-to-end journey time from incidence at the boarding station to arrival time at the alighting station, also referred to as the *scheduled journey time*,

J = the observed or actual journey time,

X = the Excess Journey Time (EJT).

With these definitions, the following equations establish the intuitive analytical framework:

$$\widetilde{A} = I + \widetilde{W} + \widetilde{V} \tag{7-1}$$

$$\widetilde{J} = \widetilde{A} - I \tag{7-2}$$

$$X = J - \widetilde{J}. \tag{7-3}$$

Equation (7-1) says that the arrival time standard is the incidence time plus some standard for waiting time plus some standard for in-vehicle time. Equation (7-2) says that the journey time standard is the arrival time standard less the incidence time. Equation (7-3) simply formalizes the definition of EJT. Naturally, the first two equations imply that the journey time standard is the sum of the waiting time standard and the in-vehicle time standard, *i.e.* $\widetilde{J} = \widetilde{W} + \widetilde{V}$.

Without loss of generality, consider an origin station ("station 1") on a rail line, a randomly selected passenger traveling from that station to a destination station ("station 2") on the same line, a set of trains that passenger is willing to board, including a pair of those consecutive trains scheduled to depart from station 1 towards station 2 with the first train scheduled to depart at time $t = 0$. Figure 7-1 uses a time-distance diagram to illustrate the following additional quantities relevant to this analysis.

d = the scheduled departure time from station 1 of the next train that the passenger in question is willing to board.

h = the scheduled headway between the prior scheduled departure and the next scheduled departure.

a = the scheduled arrival time at station 2 of the train departing station 1 at time d.

v = the scheduled running time from station 1 to station 2 of the train departing at time d.

a' = the actual arrival time at station 2 of the train carrying the selected passenger (whichever train that may be).

l = the difference between the actual arrival time at station 2 of the train carrying the selected passenger and the scheduled arrival time at station 2 of the train scheduled to depart station 1 at d.

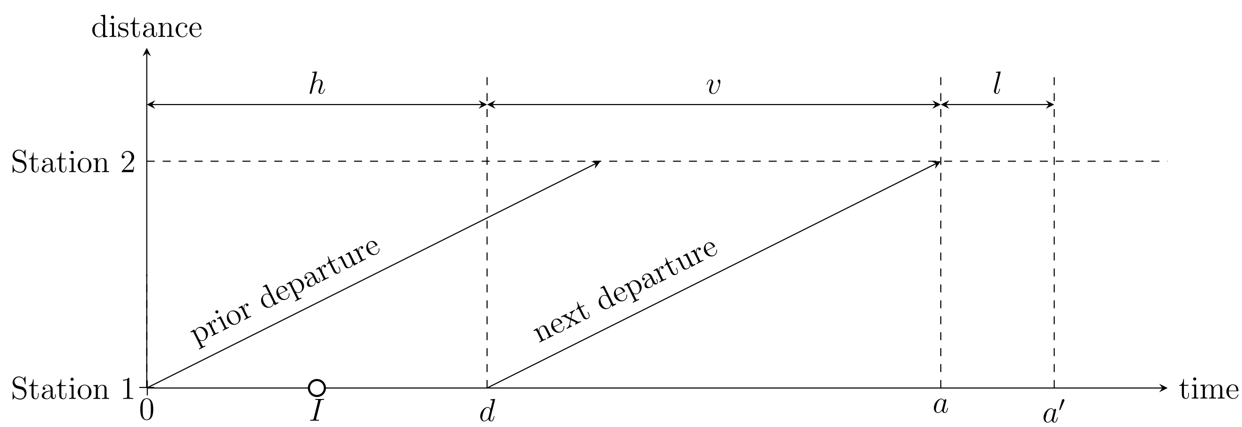


Figure 7-1: Example Time-Distance Graph

The following useful relationships are implied by this diagram and related definitions:

$$h = d - 0 = d \tag{7-4}$$
$$a = d + v \tag{7-5}$$
$$a' = a + l \tag{7-6}$$
$$J = a' - I. \tag{7-7}$$

Some aspects of this framework are worth noting. Firstly, and without loss of generality, it is conditioned upon the given passenger being incident during the specific (but arbitrary) headway $[0, h]$. Secondly, there is no requirement of correspondence between the identities of the train scheduled to arrive station 2 at a and the train actually arriving at a'. That is, a specific train in the timetable is scheduled to arrive at a, but the passenger arriving at a' may or may not be on that specific train. Finally, neither the *actual* waiting nor in-vehicle time experienced by the passenger is used within this framework. The data required to apply this framework consists only of the timetable and measurements of incidence and arrival times for individual passengers. This framework, illustration, and notation are designed to reflect the issues arising in measuring EJT with AFC (*e.g.* Oyster smartcard) data.

7.2.2 Passenger Incidence and Behavioral Assumptions

Let $f_I(i), i \in [0, d]$ be the probability density function for passenger incidence times during the headway in question. This function is assumed to be continuous, representing a smoothed description of behavior during the given headway on an average day.

It is assumed that all passengers belong to one of two behavioral classes of individuals, each with its own method for setting journey time standards. These two classes will be referred to here as *scheduled incidence* and *random incidence*. Mathematical terms for these classes of passengers will be superscripted with S and R, respectively, as in $f_I^S(i)$ or X^R. These two classes of passengers correspond to the two broad categories of passengers identified in studies of passenger incidence as discussed in Chapter 6.

The following sections describe the appropriate journey time standards to use for each class of passengers. They also describe a type of probability density function for incidence times that would be observed in the presence of each class. It will be shown later that it is not necessary to assume that the entire set of passengers at a given station, or even on a given origin-destination (OD) pair, come from *only* one of these two classes.

Scheduled Incidence

Scheduled incidence passengers are passengers whose incidence is timetable-dependent – their behavior cannot be considered entirely random with respect to scheduled departure times. They are assumed to have knowledge of scheduled departure times and scheduled running times, which they use both to time their incidence and to set waiting and in-vehicle time standards.[10] It is assumed that their standard for waiting time is exactly the time between

[10] This approach neglects the possibility of passengers who set their standards based on their experience of actual train departure and running times. Such standards would result in a measure of absolute service

incidence and the next scheduled departure (*i.e.* the time they would expect to wait, given their time of incidence, if they expected the next train to depart as per the timetable), and that their standard for in-vehicle time is as per the timetable. In the context of the analytical framework, this implies

$$\widetilde{W}^S = d - I \tag{7-8}$$

$$\widetilde{V}^S = v \tag{7-9}$$

which, along with Equations (7-1), (7-2), and (7-5), imply:

$$\widetilde{A}^S = I + (d - I) + v = a \tag{7-10}$$

$$\widetilde{J}^S = a - I. \tag{7-11}$$

These results correspond with the simple intuition that if a passenger has knowledge of the timetable, her standards for a given journey depend on her time of incidence only insofar as it determines the next scheduled departure. Her standard for arrival depends only on the timetable for that departure. These equations, along with Equations (7-6) - (7-7), substituted into Equation (7-3) yield the similarly intuitive result that

$$X^S = l. \tag{7-12}$$

Consequently, conditioned upon the passenger being incident on the given headway and arriving at time a', EJT is independent of I and thus is *not* a random quantity.

Because this class of passengers are assumed to be aware of the schedule, all that is assumed regarding the distribution of their incidence times over a given headway h is that it is *not* uniform (*i.e.* completely random). Specifically speaking, a continuous function $f_I^S(i)$ is taken to be a probability density function for the incidence times of scheduled incidence passengers if it meets the following conditions:

$$f_I^S(i) \geq 0, i \in [0, h] \tag{7-13}$$

$$\exists\, i \in [0, h] : f^S(i) \neq \frac{1}{h} \tag{7-14}$$

$$\int_0^h f_I^S(i) di = 1. \tag{7-15}$$

Figure 7-2 shows an example of such a distribution, where the rate of passenger incidence increases linearly as the departure time approaches, though the conditions as stated are much less restrictive than this specific example. As discussed in Chapter 6, if such a distribution were observed in practice, one could argue that it would then be reasonable to assume that passengers somehow schedule their incidence, and thus it would be reasonable to use the journey time standards in this section.

---

quality similar to RBT and thus contrary to the nature of EJT as a measure of service quality explicitly relative to the timetable.

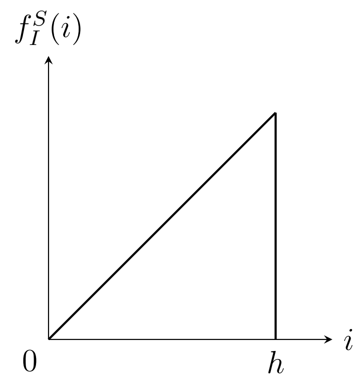


Figure 7-2: Example probability density function of incidence time for scheduled incidence passengers

Random Incidence

Random incidence passengers are passengers whose incidence behavior is completely independent of scheduled departure times. They are assumed to have knowledge of scheduled running times and headways but not to have or not use any knowledge of scheduled train departure times. These passengers are assumed to set standards for waiting time based on knowledge of scheduled train headways and to set standards for in-vehicle time based on knowledge of scheduled train running times. Specifically, it is assumed that their standard for waiting time is exactly half the scheduled headway in which they are incident, and that their standard for in-vehicle time is as per the timetable. In the context of the analytical framework, this implies[11]

$$\widetilde{W}^R = \frac{h}{2} \tag{7-16}$$

$$\widetilde{V}^R = v \tag{7-17}$$

which, along with Equations (7-1), (7-2), and (7-5), imply

$$\widetilde{A}^R = I + \frac{h}{2} + v \tag{7-18}$$

$$\widetilde{J}^R = \frac{h}{2} + v. \tag{7-19}$$

These results correspond with the intuition that if a passenger has no knowledge of specific departure times, his standard for arrival time *will* depend on his time of incidence, but that his *a priori* standard for total journey time is independent of his time of incidence. These equations, substituted into Equation (7-3) and with help from Equations (7-4) - (7-7),

[11] Equation (7-16) does not include the effect on waiting time of variability of scheduled headways, as shown in Equation (6-1), because the analysis is conditioned on the passenger being incident on a specific headway; this is the same conditioning used by Osuna and Newell (1972) to arrive at Equation (7-16).

yield

$$X^R = l + \frac{h}{2} - I. \tag{7-20}$$

EJT for random incidence passengers is, unlike for scheduled incidence passengers, a random variable, even when conditioned upon being incident in the given headway and arriving at time a'. This result is also intuitive, indicating that the EJT for a given randomly incident passenger depends on luck with respect to how close his time of incidence is to subsequent departures. This is further discussed in the following section.

For random incidence passengers, conditional upon being incident at a given station during a given scheduled headway, their specific times of incidence are assumed to be uniformly random. In precise terms, for a passenger incident during a given headway h, the classical assumption (*e.g.* Osuna and Newell, 1972) is made that

$$f_I^R(i) = \begin{cases} \frac{1}{h}\,,\ i \in [0, h] \\ 0\,,\ otherwise \end{cases} \tag{7-21}$$

which is shown in Figure 7-3. As discussed in Chapter 6, if such a distribution were observed in practice, one could argue that it would then be reasonable to assume that passengers are randomly incident, and thus it would reasonable to use the journey time standards in this section.

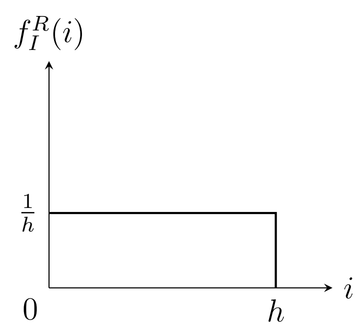


Figure 7-3: Example probability density function of incidence time for random incidence passengers

7.2.3 Framework Intuitions

Under this framework, it should be intuitively clear that for a given individual passenger journey, the wrong assumption about incidence behavior and journey time standards would result in a biased estimation of EJT. For example, consider a passenger who is incident just before the scheduled departure time (*i.e.* $I = d - \epsilon, \epsilon \ll h$) on a day in which all trains run perfectly to schedule and capacity is not binding (*i.e.* $a' = a$ so $l = 0$). Under the above framework, if it is assumed that this was in fact a randomly (and luckily) incident passenger, it would be estimated that $X^R = -\frac{h}{2} + \epsilon$. On the other hand, if it is assumed this was a scheduled incidence passenger, it would be estimated that $X^S = 0$. In this sense, it is clear that at the level of an individual journey, correct behavioral assumptions are crucially important to unbiased estimation of EJT. It should be noted that the bias introduced by the wrong incidence assumption is at most one half headway.

The following sections prove analytically that, contrary to the intuition developed here, the assumption that all passengers are *scheduled* incidence passengers yields an unbiased estimator of EJT regardless of the *actual* passenger incidence behavior. To help see why this would be, consider the set of all passengers incident in the same headway of the lucky passenger in the above example. If they all were random incidence passengers, there would also be, probabilistically speaking, an *unlucky* passenger who just missed the prior train (*i.e.* incident at $I = \epsilon$). For this new passenger, under the random incidence assumption, it would be estimated that $X^R = \frac{h}{2} - \epsilon$. Averaged with the EJT for the lucky passenger, this would yield an aggregate EJT of 0, the same value as would be estimated for both of these passengers under the assumption of scheduled incidence.

7.3 A Unified Unbiased Estimator for Aggregate Excess Journey Time

Figure 7-1 and Equations (7-4) - (7-7) describe the journey of a single passenger who is incident on the headway $[0, h]$ and arrives at his or her destination at time a'. The end of the previous section constructed a trivial example in which the wrong assumption regarding the class of this passenger biased the estimation of EJT for this single journey. In practice there is no intent to report EJT (*i.e.* estimate X) at the level of an individual passenger. Rather, EJT should be aggregated over many passengers to indicate the performance of all or part of the network in question over a period of time. Of interest is an estimate of the probabilistic expectation (*i.e.* the mean) of X, $\mathrm{E}[X]$ for *all* passengers incident on the headway $[0, h]$.

This section first extends the analytical framework to include multiple passengers with multiple arrival times. It then shows that the estimator for aggregate EJT under the scheduled incidence assumptions is also an unbiased estimator for aggregate EJT under the random incidence assumptions if passengers are in fact randomly incident. Finally, it shows that the same estimator is unbiased under a blend of random and scheduled incidence behavior.

7.3.1 Framework for Aggregate EJT

If EJT is to be aggregated over multiple journeys, the analytical framework is insufficient as currently constructed. Different passengers traveling between the same two stations and incident on the same scheduled headway can have different arrival times depending on the *actual* departure times of trains from the origin station. For example, if, on a given day, trains departed station 1 at times 0, $\frac{h}{3}$, and h, and some passengers were incident on $[0, \frac{h}{3}]$ and others were incident on $[\frac{h}{3}, h]$, then it is highly unlikely that all passengers incident on $[0, h]$ could have the same arrival time a'.

To account for this, the framework is generalized. Rather than a single train arriving at station 2 at time a', instead consider K discrete trains arriving at station 2. For the k^{th} train, $k \in 1..K$, let

a'_k = the arrival time at station 2 of train k,

l_k = the difference between the arrival time of train k and a,

Y_k = an indicator random variable, for each passenger, which is 1 if the passenger arrived on train k, 0 otherwise,

α_k = the fraction of all passengers incident at station 1 on $[0, h]$ who arrived at station 2 at a'_k, trivially equal to $\mathrm{E}[Y_k]$,

$g_I^k(i)$ be a probability density function, defined over $[0, h]$, describing the distribution of the incidence time of the passengers who were incident on $[0, h]$ and traveled from station 1 to station 2 aboard train k.

It is appropriate to model the arrival times of passengers discretely since train arrivals are a discrete phenomena, at least as compared to passenger incidence. The set of K trains is exhaustive in that it includes all trains used by passengers incident on $[0, h]$ and traveling from station 1 to station 2. This is sufficient to write that

$$\sum_{k=1}^{K} \alpha_k = 1 \tag{7-22}$$

$$\sum_{k=1}^{K} \alpha_k g_I^k(i) = f_I(i). \tag{7-23}$$

It will also be useful to use the law of total expectation to decompose $\mathrm{E}[X]$ as a function of the respective probabilities and conditional expectations of X for passengers arriving on each of the K trains as

$$\mathrm{E}[X] = \sum_{k=1}^{K} \alpha_k \, \mathrm{E}[X|Y_k = 1]. \tag{7-24}$$

7.3.2 Equivalence of Random and Scheduled Incidence Assumptions for Aggregate EJT of Random Incidence Passengers

In Equation (7-12) of Section (7.2.2) it was shown that under the assumption of scheduled incidence, for a given journey incident at station 1 on $[0, h]$ and arriving at station 2 at time a', EJT is not a random variable but rather equal to l, independent of time of incidence I. Because the extended framework uses the indicator random variable Y_k, X^S is a random variable. However, conditional upon a given passenger being on train k, EJT for that passenger is no longer random and is known to be l_k, which implies that

$$\mathrm{E}[X^S|Y_k = 1] = l_k. \tag{7-25}$$

Substituting this into Equation (7-24) yields, quite intuitively, that under the assumption of scheduled incidence the estimator for aggregate EJT is

$$\mathrm{E}[X^S] = \sum_{k=1}^{K} \alpha_k l_k. \tag{7-26}$$

Under the random incidence assumption, it was seen in Equation (7-20) that EJT for a given journey *does* depend on time of incidence. However, conditioned on a specific incidence time $I = i$, X^R is a deterministic quantity. In the notation of the extended framework, the law of total expectation then can be used to write that

$$\mathrm{E}[X^R | Y_k = 1] = \int_0^h g_I^k(i)\, \mathrm{E}[X^R | I = i, Y^k = 1]\, di = \int_0^h g_I^k(i)(l_k + \frac{h}{2} - i)\, di. \tag{7-27}$$

Substituting this result into Equation (7-24), interchanging sums with integrals, and rearranging terms, it is found that the estimator for aggregate EJT under random incidence assumptions is

$$\mathrm{E}[X^R] = \int_0^h \left(\frac{h}{2} - i\right) \sum_{k=1}^{K} \alpha_k g_I^k(i)\, di + \sum_{k=1}^{K} \alpha_k l_k \int_0^h g_I^k(i)\, di. \tag{7-28}$$

If it is assumed that the passengers in question *are* in fact randomly incident, Equations (7-21) and (7-23) can be used to write that

$$\frac{1}{h} = \sum_{k=1}^{K} \alpha_k g_I^k(i), i \in [0, h]. \tag{7-29}$$

This along with the fact that the integral of any probability density function over its entire domain equals unity simplifies Equation (7-28) to

$$\mathrm{E}[X^R] = \int_0^h \frac{1}{h} \left(\frac{h}{2} - i\right) di + \sum_{k=1}^{K} \alpha_k l_k \tag{7-30}$$

which simplifies further to

$$\mathrm{E}[X^R] = \sum_{k=1}^{K} \alpha_k l_k \tag{7-31}$$

which is the same result as found for scheduled incidence in Equation (7-26).

The estimator for aggregate EJT under scheduled incidence assumptions is thus shown to be equal to the estimator for aggregate EJT under random incidence assumptions if passengers *are in fact* randomly incident. This implies that using the scheduled incidence estimator for aggregate EJT is appropriate if *all* passengers are scheduled incidence passengers or *all* passengers are random incidence passengers. The next section extends this result to the case when passengers come from both classes.

This analysis was based on a conditioning of the waiting time standard for random incidence passengers on the headway in which each passenger was incident. Passengers typically arriving in a period of time that spans multiple headways could face some variability in scheduled headways, and thus could set their standards based on the results of Equation (6-1). The derivation of Equation (6-1) depends on the realization that, under random incidence, more passengers will be incident in longer headways and thus have longer average waiting times. It is intuitively felt that the analysis presented here accounts for the same phenomena – when random incidence passengers are treated with the scheduled incidence assumptions, more of them will be incident in longer headways and thus have longer scheduled waiting times (with respect to the timetable) and longer journey time standards.

7.3.3 Blended Passenger Incidence Behavior

In practice, as found in Chapter 6, it will often be the case that *some* passengers are randomly incident while others clearly make use of the timetable. This would be indicated by a distribution of passenger incidence times over a given headway that is clearly a superposition of two different incidence distributions, one meeting the scheduled incidence conditions of Equations (7-13) - (7-15) (*e.g.* Figure 7-2), and one meeting the random incidence conditions of Equation (7-21) (*i.e.* Figure 7-3).

This can be described as *blended passenger incidence* behavior, and is consistent with the formulations for mixed incidence behavior discussed in Chapter 6. This section derives an estimator for aggregate EJT under such blended behavior and shows that this too is equal to the estimator under scheduled incidence assumptions. Functions and variables for blended behavior are superscripted with B, as in $\widetilde{J}^B$.

Without loss of generality, assume that some fraction γ of passengers incident on $[0,h]$ are random incidence passengers, and so $1-\gamma$ are scheduled incidence passengers.[12] The probability density function for incidence times of all passengers under blended incidence can then be written as the superposition of the respective random and scheduled incidence density functions

$$f_I^B(i) = \gamma f_I^R(i) + (1-\gamma) f_I^S(i)\,,\; i \in [0,h]. \tag{7-32}$$

Figure 7-4 shows an example of such a function, where $f_I^R(i)$ is as shown in Figure 7-3 and $f_I^S(i)$ is as shown in Figure 7-2.

Moreover, assume that for a given passenger it is not known which behavioral class they belong to. Let Z be an indicator random variable that is equal to 1 if a given passenger is a random incidence passenger and 0 otherwise, and $\lambda(i)$ be the probability that $Z=1$ for a passenger having been incident at time $I=i$. As illustrated in Figure 7-4 it can be written formally that

$$\lambda(i) = \Pr(Z=1|I=i) = \frac{\gamma f_I^R(i)}{f_I^B(i)}. \tag{7-33}$$

Because the behavioral class of a given passenger is no longer assumed, the journey time standard $\widetilde{J}^B$ is now a random variable, equal to $\widetilde{J}^R$ or $\widetilde{J}^S$ if the passenger is a random or scheduled incidence passenger, respectively. This can be used to define the random variable

[12] γ, the proportion of random incidence passengers, is the complement to the proportion of timetable-dependent passenger, p, discussed in Chapter 6.

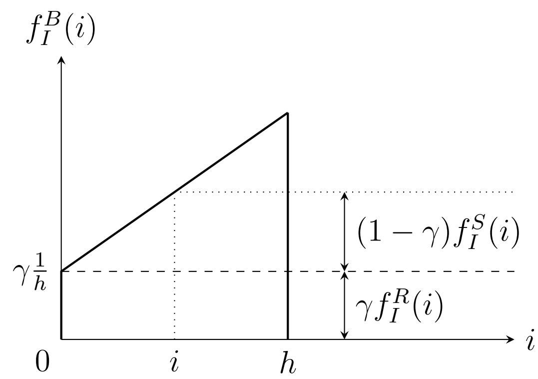


Figure 7-4: Example probability density function of incidence time for blended random and scheduled incidence passengers

for EJT of a single blended incidence passenger journey as

$$X^B = \begin{cases} X^R,\ Z = 1 \\ X^S,\ Z = 0 \end{cases} . \tag{7-34}$$

Thrice applying the law of total expectation, first on Y, next on I, and finally on Z, the estimator for aggregate EJT under blended passenger incidence behavior can be derived by writing

$$\mathrm{E}[X^B] = \sum_{k=1}^{K} \alpha_k \, \mathrm{E}[X^B | Y_k = 1] \tag{7-35}$$

$$= \sum_{k=1}^{K} \alpha_k \int_0^h g_I^k(i) \, \mathrm{E}[X^B | Y_k = 1, I = i] \, di \tag{7-36}$$

$$= \sum_{k=1}^{K} \alpha_k \int_0^h g_I^k(i) \left(\lambda(i) \, \mathrm{E}[X^R | Y_k = 1, I = i] + (1 - \lambda(i)) \, \mathrm{E}[X^S | Y_k = 1, I = i]\right) \, di. \tag{7-37}$$

Fully conditioned on Y and I, the notation defined here and results of the previous section can be used to write

$$\mathrm{E}[X^B] = \sum_{k=1}^{K} \alpha_k \int_0^h g_I^k(i) \left(\lambda(i) \left(l_k + \frac{h}{2} - i \right) + (1 - \lambda(i)) l_k \right) \, di \tag{7-38}$$

which is easily re-arranged to

$$\mathrm{E}[X^B] = \sum_{k=1}^{K} \alpha_k \int_0^h g_I^k(i) \lambda(i) \left(\frac{h}{2} - i \right) di + \sum_{k=1}^{K} \alpha_k l_k \int_0^h g_I^k(i) di. \tag{7-39}$$

Interchanging sums with integrals and rearranging the first term and simplifying the second term yields

$$\int_0^h \lambda(i) \left(\frac{h}{2} - i \right) \sum_{k=1}^{K} \alpha_k g_I^k(i) di + \sum_{k=1}^{K} \alpha_k l_k. \tag{7-40}$$

Substituting Equations (7-23) and (7-33), this result becomes

$$\mathrm{E}[X^B] = \int_0^h \frac{\gamma f_I^R(i)}{f_I^B(i)} \left(\frac{h}{2} - i \right) f_I^B(i) di + \sum_{k=1}^{K} \alpha_k l_k. \tag{7-41}$$

Canceling terms and substituting Equation (7-21), this becomes

$$\mathrm{E}[X^B] = \int_0^h \frac{1}{h} \left(\frac{h}{2} - i \right) di + \sum_{k=1}^{K} \alpha_k l_k \tag{7-42}$$

which is the same as Equation (7-30), again simplifying to

$$\mathrm{E}[X^B] = \sum_{k=1}^{K} \alpha_k l_k. \tag{7-43}$$

The estimator for aggregate EJT in the case of multiple trains carrying blended incidence passengers is thus found to be the same as the estimator for aggregate EJT under scheduled incidence assumptions.

7.3.4 Extension to a Heterogeneous Rail Network with Interchanges

The author's intuition is that this analysis extends readily beyond a single rail line with a single service pattern to a rail network with interchanges and a variety of service patterns. Such an extension would require the model to account for passenger incidence behavior at interchange locations. Without formally extending the model, the following observations should provide an intuitive sense of why the schedule-based estimator for aggregate EJT is appropriate in a network context. Note that in all cases it is expected to measure only the *end-to-end* journey time, which subsumes all interchanges.

- This analysis easily extends to include a network with walking links, such as those between AFC gatelines and station platforms. Such links can be thought of as lines or services with continuously available departure times (*i.e.* infinite frequency, or zero headway), in which case the distinction between scheduled and random incidence is irrelevant.

- On a single line with heterogeneous service patterns, such as short turns or a trunk-and-branch configuration, passengers can be considered, as in Chapter 6 to ignore certain departures that do not improve their overall travel time. This simply changes the timetable applicable to each passenger's journey, not the analysis thereof.

- If passengers are aware of and make plans based on the timetable for the entire network, then clearly the schedule-based estimator is appropriate.

- If passengers are unaware of or do not use the timetable for *any* portion of the network, and if the different services are timetabled independently, then incidence at the interchange location will be random with respect to the departures of the service being interchanged to. In this case, some passengers will be lucky and experience short interchange times while others will experience long ones, and the same probabilistic smoothing seen in Section 7.3.2 should apply.

- If the timetable is designed to facilitate interchanges (i.e. minimize interchange times) between lines, then a schedule-based estimator should be used *regardless* of passenger incidence behavior and standards. Even if passengers are randomly incident at the initial station, their incidence at the interchange station is non-random with respect to the timetable of the subsequent line *by the very nature* of the specially-constructed timetable.

- If passengers are aware of and use the timetable for only a portion of the network, then they either interchange from a service on which they schedule their incidence to a service on which they are randomly incident, or vice versa. In the former case, since they do not know (or care) about the timetable for the second service, their incidence time (and thus journey time standard) for the first service is unaffected.

- The reverse scenario, where passengers are unaware of the timetable on their first service, but have a target departure in mind for the second, is perhaps less straightforward. In this case, it would be reasonable to set a waiting time standard of a full (rather than half) headway for the first service, since this is what an operator would recommend, based on the timetable, to minimize the probability of missing the second, scheduled, departure. A first intuition is that this would bias some of the analyses in this chapter. However, as was found in those analyses, the first intuition with respect to incidence behavior, the timetable, and journey time standards is not always correct. This issue merits further examination.

If these intuitions are to be believed, and the issue in the final observation is resolved, the model and estimators for aggregate EJT developed in this chapter are in fact quite general and should be applicable to a wide variety of contexts.

7.4 Discussion

This section anticipates and discusses some concerns that may arise in the application of the method proposed in this chapter.

7.4.1 Application Considerations

AFC penetration rates may vary across the network for which EJT is measured. In some cases, this may require a weighting of EJT values to account for this variation. For example, if the rate varies significantly across different OD flows on the same line, re-weighting may be needed when analyzing EJT on that line. If penetration is largely consistent for a given line but varies across line, such correction is only necessary if comparing EJT results between those lines. In any case, an OD estimate of the type developed in Chapter 5 should be sufficient to re-weight values across a given network.

Care must be taken if EJT is to be measured for only some portion (*e.g.* the Overground) of a broader network (*e.g.* the entire London railway network), especially if timetables are available only for that portion of the network. The most straightforward way to handle this situation is to select only those OD flows that will use the portion of the network in question with relative certainty. This can be done manually based on judgment, or can take advantage of an assignment model that considers the entire network such as that used in Chapter 5.

7.4.2 Negative EJT

As mentioned, EJT for an individual passenger journey (under the scheduled incidence assumptions) can be negative. This is in and of itself not a cause for concern in terms of the measurement of EJT. However, aggregate EJT that is net negative may indicate certain biases in the EJT estimation process. Negative EJT for individual journeys can occur for several reasons, including the following.

- A passenger uses some service not included in the set of timetables used in setting journey time standards. In this case, the negative EJT can be smaller or larger in magnitude than the headway of the service in question.

- A passenger takes the service on which he or she is scheduled to depart, but that service arrives at that passenger's destination earlier than the timetable indicates. In this case, when the headway is relatively large, the negative EJT should be small in magnitude relative to the headway of the service in question.

- A passenger takes an earlier service than the one which he or she is scheduled to depart (because that earlier service was running late), which naturally arrives at the passenger's destination before the passenger's scheduled arrival time. In this case, the negative EJT can be as large as the headway of the service in question.

The first of these reasons indicates a potential bias the estimation of EJT. If non-scheduled trips were inserted into the timetable by the operator in question as a result of service control decisions, the negative EJT is unbiased in that passengers would not be considered to have expected to use this new service. However, if the service that the passenger used was provided by a different operator (*e.g.* one who shares the same track, or on a different path altogether), the result can be considered a biased EJT in that the service should have been used in setting journey time standards. This reflects a problem with the selection of OD flows for which EJT is measured for a given operator, which may result from biases in an assignment model used to select those OD flows.

The second and third of these reasons do affect the EJT estimate for an individual journey. These negative EJT measurements clearly affect the distribution of EJT, but should not under most circumstances unduly affect the mean. To better understand this phenomena, and its relationship to passenger incidence times, consider the following example. A passenger is incident to a 20 minute headway service at 08:01, one minute after a scheduled departure time of 08:00, and the train scheduled to make that 08:00 departure runs exactly one minute late for its entire trip. The passenger in question will have an EJT of -19 minutes (early), but the other passengers who caught this train and were scheduled to catch this train will have an EJT of 1 minute (late).

In this example, the one large negative EJT will be balanced out by many small positive EJTs. Where this balance (*i.e.* the mean EJT) falls in general depends on the passenger incidence behavior and on the other trains making up the service. If passengers are uniformly and randomly incident and all trains are equally and uniformly late (*i.e.* maintaining an even headway), the balance will be exact (*i.e.* mean EJT will be zero). This is as expected, in that even headways are all that is required to serve random incidence passengers perfectly.

For passengers whose incidence is timetable-dependent and for trains with varying adherence to the timetable, the situation is more complex. This complexity has not been completely analyzed, but it is felt intuitively, especially in light of the analysis of Section 7.3, that mean EJT will appropriately reflect the balance between winners (with negative EJT) and losers (with positive EJT).

An exception to this intuition is the case when a large proportion of passengers do indeed exhibit "late running awareness" as described in Chapter 6. When passengers are incident after a scheduled departure time based on an expectation that trains regularly depart late, and their expectations are correct, they will, individually, have a negative EJT close in magnitude to the headway. If enough passengers exhibit this behavior, EJT can be net negative, even if those passengers arrive at their destinations on time or late compared to the schedule for the train *they* may have expected to take. This does not constitute a bias in EJT measurement as defined here, but would confound the interpretation of EJT results. Such a situation highlights the need to study passenger incidence behavior, as in Chapter 6, before analyzing EJT.

This is an example of the potential peril of applying a measure of relative service quality with standards based on the timetable rather than true passenger expectations. As mentioned, all of these complexities do affect the distribution of EJT, somewhat confounding the interpretation of higher order statistics (*e.g.* percentiles) of EJT. These higher order statistics have important implications for measuring reliability from the passenger's perspective. They are not analyzed here, but are deserving of attention in future research on the subject.

7.4.3 EJT and Longitudinal Analysis

In practice, many tactical planning changes include revisions to the timetable. As discussed in Section 7.1.4, this presents a problem for using EJT in longitudinal analyses. When the timetable is revised, changes in EJT may be driven more by the timetable modification than by any real changes in journey times experienced by passengers. For example, if running times in the timetable are lengthened but passenger journey times remain the same, EJT will decrease.

Furthermore, as discussed in Chapter 6, passengers may adjust their incidence behavior over time as service conditions change. The method proposed here is entirely conditioned on actual incidence behavior, so changes in such behavior will not bias estimates of EJT *per se*. In one sense, this is a positive feature of this method because it absolves the analyst of the need to make any assumptions regarding incidence behavior. However, it also implies that EJT will not capture some of the benefits of improved service reliability. Specifically, it will not reflect the benefits captured by passengers who take advantage of more reliable service by adjusting their incidence behavior to reduce waiting time (likewise for the harm to passengers who react to less reliable service by becoming more randomly incident incidence).

For example, consider a service that has become more reliable over time, perhaps because of improved infrastructure or rolling stock but with no changes to the timetable. If, as a result of this reliability improvement, passengers of this service now arrive at their destinations closer to their respective arrival time standards, such will be reflected in EJT measurements. However, it may also be the case that the journey time standards of some of these passengers has decreased because, as the service has become more reliable, their incidence behavior has become less random (*i.e.* more timetable-dependent, with smaller scheduled waiting times). This would not be reflected in EJT measurements.

These realizations highlight the relative nature of EJT, and suggests that other measures, for example those for absolute service quality, may be necessary for longitudinal analysis. It should also be noted that measures of service quality, including EJT, are not intended to be used for evaluating a timetable on its own merits. They simply speak to the differences between passengers actual journeys and the promise implied by the timetable. Evaluation of a timetable in isolation from passenger journeys is not considered here.

7.5 Conclusions

Excess journey time (EJT), with standards derived from the timetable, is a measure of relative service quality that strikes useful balance between the passenger's and operator's perspectives. It has found lasting application at a number of large urban railways. Actual passenger journey times can now be measured (rather than modeled) directly from automatic data produced by AFC systems such as the Oyster smartcard.

Along with measuring actual passenger journey times, EJT depends on a standard against which to compare those measurements. These standards should be based on the timetable, so as to be as meaningful and useful to operators as possible. Within that constraint, they should reflect passenger concerns as realistically as possible. Most measures of service quality and relative service quality have made the assumption of random incidence, with the implied standard for waiting time of half the scheduled headway. As discussed in Chapter 6 passenger incidence behavior is often, including on the London Overground, much more heterogeneous than that. This heterogeneity of behavior comes with certain implications about what knowledge passengers have of the timetable and how they use that knowledge.

This chapter has established a rigorous framework for analyzing EJT, in particular for reasoning about passenger's journey time standards as implied by varying incidence behaviors. It was found that the wrong assumption about incidence behavior and journey time standards can result in a biased estimate of EJT at the level of an *individual passenger journey*. It was also shown that the estimator for *aggregate EJT* is unbiased, regardless of actual passenger behavior, under the assumption that all passenger incidence and associated journey time standards are dependent on actual departure times in the timetable. This result was proven for a single rail line without interchanges, but intuitively should hold for a rail network.

This is a very useful result in practice. It allows for the estimation of aggregate EJT from only AFC (*e.g.* Oyster smartcard) data and published timetables in a simple unified manner, regardless of service frequencies or passenger behavior that vary across the network or over time. The following chapter applies this result to the London Overground.

Chapter 8

Oyster-Based Excess Journey Time on the London Overground

This chapter presents excess journey time (EJT) results for the London Overground network. It adopts the conceptual approach proposed in Chapter 7 and implements it using published Overground timetables and journey data from the Oyster smartcard ticketing system. It uses some of the results observed on the Overground as consideration points for discussing the properties and merits of EJT as a measure of service quality.

Section 8.1 describes the methodology by which EJT was measured for the Overground network, including implementation details. Section 8.2 validates the results through graphical analysis. Section 8.3 presents EJT results for the Overground network, and compares these results with the network's existing measure of service delivery. Section 8.4 draws some conclusions about the issues that arise in applying EJT to a real-world network such as the Overground. The following chapter uses these and other EJT results to document and evaluate a recent tactical planning exercise on the Overground network.

8.1 Excess Journey Time Methodology for the London Overground

EJT for London Overground journeys is estimated according to the unified methodology described in Sections 7.2 and 7.3 of the previous chapter. This method assumes that the incidence behavior and journey time standards of all passengers are dependent on the timetable. This was shown in Section 7.3 to be unbiased in aggregate, even if some or all passengers are in fact randomly incident.

Under the framework of Section 7.2, for each given journey recorded by the Oyster smartcard ticketing system

- the incidence time, I, is estimated as the timestamp of the entry transaction;
- the actual arrival time, a', is estimated as the timestamp of the exit transaction;
- the scheduled arrival time, a, is estimated from the timetable;

- the total journey time, J, is estimated as $a' - I$ (the difference between the entry and exit transaction times);

- the excess journey time, X, is estimated as $a' - a$ (the difference between the actual and scheduled arrival times).

Figure 8-1 illustrates this method for a passenger traveling from Stratford to Camden Road on the North London Line. In this "time-distance" plot, the X axis represents time and the Y axis represents the distance traveled along the North London Line (NLL) in number of stations. Each line traveling northeast through the plot shows the schedule of one weekday service.

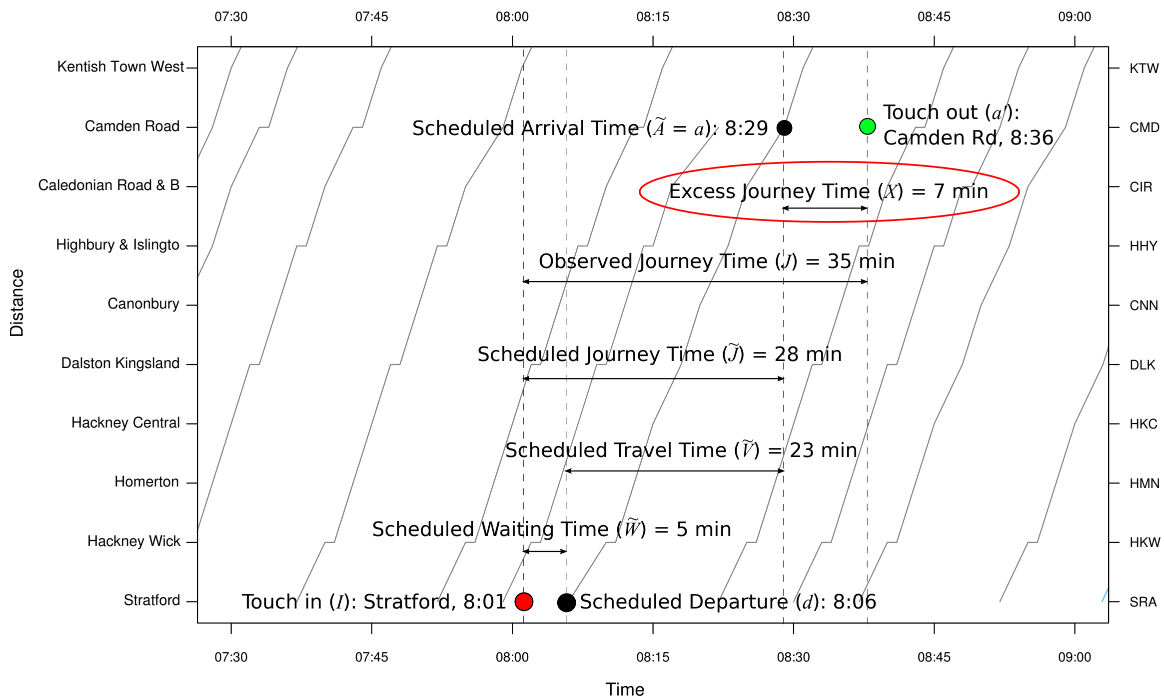


Figure 8-1: Time-distance illustration of EJT estimation for a passenger from Stratford to Camden Road

The estimation of the scheduled arrival time, a, for each passenger is achieved through the same schedule-based assignment process used to analyze passenger incidence behavior in Chapter 6. The schedule-based path returned by the Path() function of Algorithm 6.1 includes information about the scheduled time of arrival at the destination station. This assignment incorporates the following assumptions:

- No time is required to move between the points of Oyster validation (*e.g.* station gate-lines) and train platforms.

- Trains are scheduled to depart at the beginning of the minute indicated in the timetable (*i.e.* 08:00 means 08:00:00 not 08:00:59).

- Entry and exit transactions are uniformly distributed over the minute indicated by the Oyster database (*e.g.* transactions recorded at 08:00:00 actually occur randomly between 08:00:00 and 08:00:59).

These assumptions and their implications for the schedule-based assignment process have already been discussed at length in Chapter 6. The first assumption is broadly justified by the actual characteristics of most Overground stations but is adopted for convenience purposes and could bias EJT estimates for those journeys incident at large stations very shortly before a scheduled departure. The final assumption does not, probabilistically speaking, bias the estimate of total journey time because entry and exit transactions both undergo the same timestamp truncation process.

The same software and data are used here as in Chapter 6. A combination of open source tools and scripts were used to process over 1.6 million Oyster journey records made by over 290 thousand passengers on the 52 weekdays from 31 March, 2008 through 10 June, 2008, inclusive. As in Chapter 6, the data set was filtered to include only those journeys for which it is almost certain (based on the assignment model of Chapter 5) that the passenger in question used only Overground services. In that sense, EJT is estimated through a two-stage assignment process. First, a frequency-based assignment is used to select journeys that are (almost) certain to have used the Overground. Second, a schedule-based assignment is used to determine EJT with respect to the Overground timetable for those journeys.

EJT results were not re-scaled to account for varying Oyster penetration rates across the Overground network. This could have been accomplished using the OD estimate from Chapter 5, but only for the AM Peak results (since only an AM Peak OD matrix was estimated).

8.2 Graphical Validation

This section validates that aggregate Oyster-based EJT measurements accurately reflect events on the ground, including train operations (*i.e.* service delivery) and the passenger's experience (*i.e.* service quality). Because of the complex and dynamic nature of even a single day's rail operations and passenger journeys, a graphical approach is used. An example of the type of plot used is shown in Figure 8-2.

This plot is similar to that shown in Figure 8-1 with the addition of the times, locations, and EJT of actual passenger journeys.[1] The size of the slanted hatch marks represent the number of Oyster journeys (in the Stratford → Richmond direction) that exited a given station on a given minute of the day on Thursday 3 April, 2008. The color of each hatch mark indicates the average EJT for the trips it represents. The more yellow and then red the mark, the more positive (*i.e.* late) the EJT; the greener the mark, the more negative (*i.e.* early) the EJT.

Unfortunately, automatic data on actual train movements was not available for this research (though it does exist for the Overground network). However, the trails of Oyster exits clearly describe the movement of trains over the course of a given trip along the line.

[1] For station abbreviations see Appendix A

In this sense, the plot contains all the information needed to asses whether EJT faithfully captures the delays incurred by passengers of degraded train service.

There is much that can be inferred from this plot about the service delivery and quality on the day in question, for example:

- Trains arriving Richmond (RMD) after 08:00 were less patronized and ran to schedule or a bit early.

- Starting with the 07:07 service from Stratford there are slight delays, which become severe (i.e., bright red) for the 08:06 and 08:22, and, perhaps, also the 08:30 and 08:37 trips.

- The 08:52, 09:03, 09:22, and 09:38 services ran smoothly, at least as far as Willesden Junction (WIJ).

- The 09:31 shuttle to Camden Road (CMD) may not have run at all, as reflected by the late (i.e., red) passengers as far as Camden Road on the 09:38 service.

- The 09:52 from Stratford ran extremely late or not at all.

- By the 11:07 departure from Stratford, the service had largely recovered.

The above hypotheses can be judged against the true record of events found in the Overground's incident logs, which indicate the following:

- No NLL trains departing Stratford before 07:00 on Thursday 3 April, 2008 were more than a minute late arriving at their destinations.

- The trains departing Stratford between 07:00 and 08:00 arrived at their destinations between -2 minutes early and 9 minutes late.

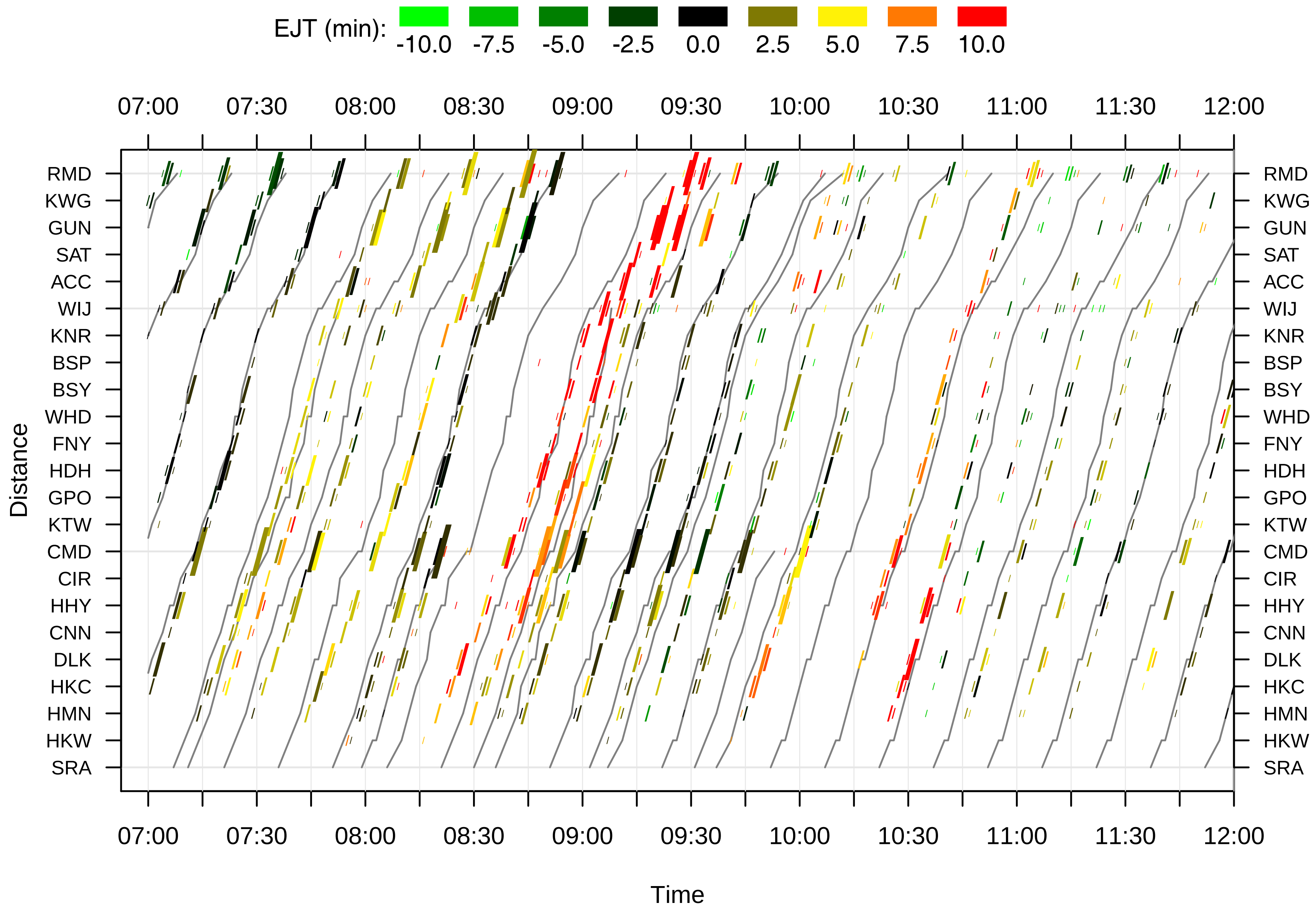
- The 08:06 and 08:22 trains from Stratford arrived at Richmond 18 and 10 minutes late, respectively, because of severe door problems on the 08:06 train.

- The 08:52, 09:03, 09:22, and 09:38 services were 2 minutes early and 5, 2, and 0 minutes late, respectively, at their destinations.

- The 09:31 shuttle was cancelled, because the rolling stock scheduled for this trip was instead used for the 09:38 service, whose assigned rolling stock had been delayed inbound from Richmond by a sick passenger.

- The 09:52 departure from Stratford was short-turned at Hackney Wick (HKW) on its incoming trip, a knock-on effect of the above sick passenger, and arrived 11 minutes late at Richmond.

- All trains departing Stratford between 11:00 and 12:00 were between 2 minutes early and 3 minutes late at their destinations.



Hatch marks indicate by size the number of exits (per minute) and by color the average EJT (red is positive/late, green is negative/early)

Figure 8-2: Time-distance plot of timetable and observed Oyster exits for westbound travel on the North London Line on 3 April, 2008

The consistency between the service inferences derived from the coupling of the Oyster data and the timetable on the one hand and the incident logs on the other is evident. Particularly compelling about this example is the cancellation of the 09:31 train and the short-turn of the 09:52 train. The precise effect of these control actions on passengers, relative to the scheduled service, is captured in the EJT measure, as shown in Figure 8-2

Validation of this sort was performed for each of the 10 weekday mornings from 31 March 2008 to 11 April 2008, inclusive, with similar results. In each case, a similar correspondence was found between the chart and the Overground's incident logs.

8.3 Results

This section presents EJT results for the London Overground network, first in isolation and next in comparison to the existing measure of service delivery. The results in this section are presented as much for the sake of exploring EJT as a measure of the passenger experience as they are for analyzing the performance of the Overground network itself.

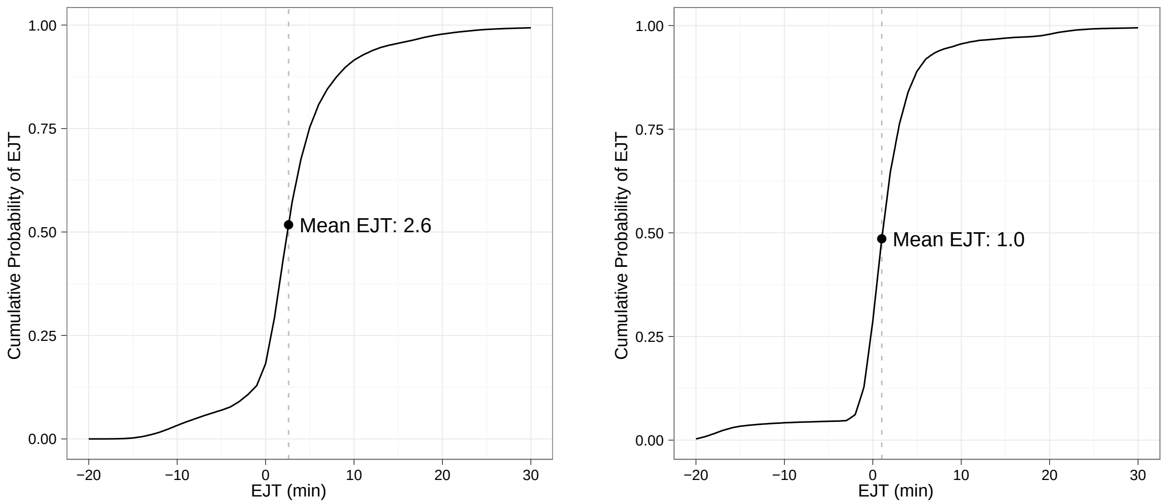
8.3.1 Excess Journey Time on the London Overground

Distributions and Negative EJT

Figure 8-3 plots the cumulative distributions of EJT in the AM Peak period for the NLL and the Gospel Oak to Barking Line (GOB).[2] It is not a surprise that in both cases a substantial fraction (18% and 28%, respectively) of EJT measurements are negative, implying that many passengers arrived at their destination earlier than the timetable indicates they should have. The mass of the distributions, especially on the GOB, between approximately -3 and 0 minutes is likely the result of trains running slightly ahead of schedule. The distributions effectively start at -15 minutes on the NLL and -20 minutes on the GOB, as predicted in the discussion of negative EJT in Chapter 7 – these are the headways of those services during the AM Peak.

Based on these results and the results in Chapter 6, it is not felt that aggregate EJT will be meaningfully biased by the combination of "late running awareness" behavior among a large proportion of Overground passengers coinciding with actual late running by trains.

[2] These plots smooth over the Oyster timestamp truncation by adding a random perturbation uniform on $[0, 1]$ to each EJT measurement (not included in the calculation of the mean EJT).



(a) North London Line (b) Gospel Oak to Barking Line

Figure 8-3: Distributions of EJT in AM Peak periods

Mean and Total EJT by Line and Time Period

Figures 8-4 and 8-5 show two different aggregations of EJT by line[3] and time period. The first of these presents the daily average of the *sum* of EJT for all Oyster journeys. This plot emphasizes the passenger-weighted nature of EJT as a measure of system performance. It is clear that, as a product of the number of passengers and the length of delays experienced by those passengers, the NLL, particularly during the AM and PM Peak periods, is the most problematic part of the Overground network. It is the line most deserving of management and tactical planning attention; the other lines frankly pale in comparison.[4] For the GOB, the West London Line (WLL) and for interchange passengers (INT) the AM and PM Peak periods have the most total passenger delay. The Watford DC Line (WAT) breaks this pattern, with negative total and mean EJT in the Early and AM Peak periods. This net negative EJT is further discussed later in this section.

Figure 8-5 shows the pure mean passenger EJT. It puts all lines and time periods on equal footing by normalizing by the total number of journeys. This plot is primarily useful for comparing the performance of different lines at different times of day from the perspective of the *average* passenger, rather than *all* passengers. Overall mean EJT clearly varies across lines and time periods. After normalizing for the total number of journeys, the AM and PM Peak periods, with EJT of 2.6 and 2.2 minutes, respectively, are still the most problematic periods for the NLL.

[3] Lines, as in Table 2-2, represent single-line passengers. INT refers to passenger journeys interchanging *between* Overground lines.

[4] The numbers in Figure 8-4 are not adjusted to account for varying rates of Oyster penetration, which should bias the relative differences in total EJT between lines. For example, Chapter 5 found in Table 5-2 that Oyster penetration in the AM Peak on the WLL was in rough terms half that of the other three lines. However, even correcting this plot for these differences (*i.e.* doubling the height of the bars in the WLL graph) would not affect the conclusion that the NLL is the source of the lion's share of EJT.

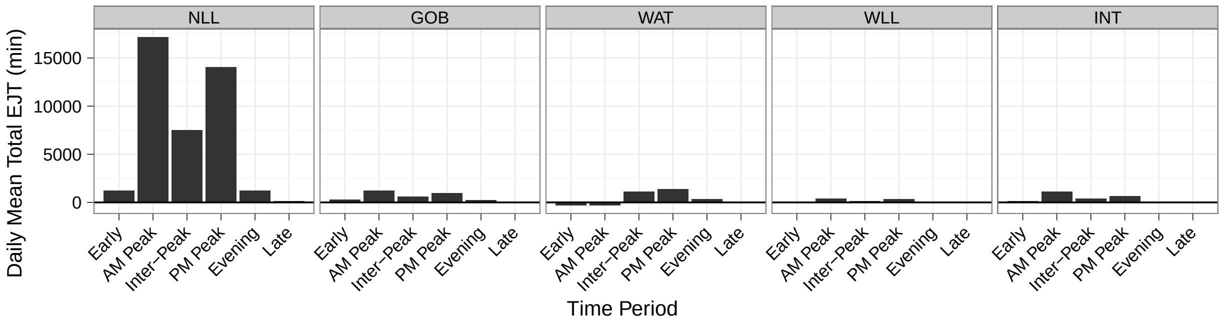


Figure 8-4: Total EJT, by line and time period

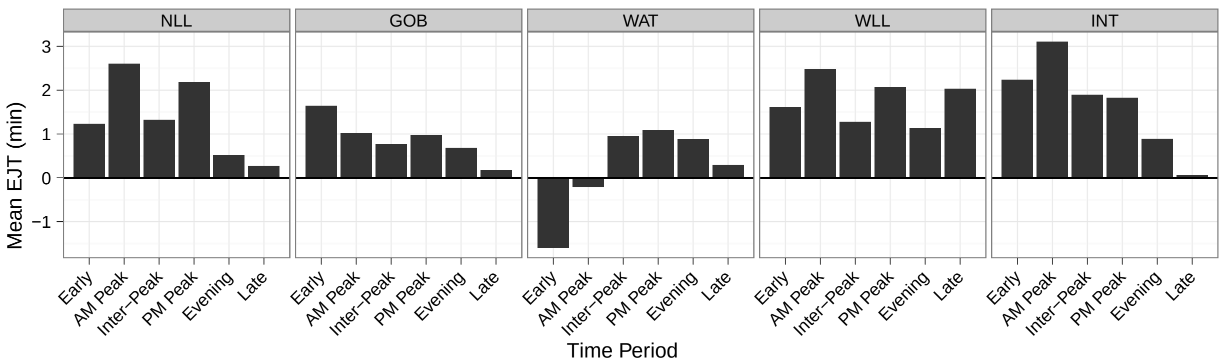


Figure 8-5: Mean EJT, by line and time period

These measurements are higher than all other lines for corresponding time periods except for interchange passengers (INT) in the AM Peak, with an EJT of 3.1 minutes. It is not unexpected that interchange passengers (most of whom likely use the NLL for one leg of their journey) suffer longer delays than single-line passengers. A short delay on the first leg of an interchange journey can cause the passenger to miss the targeted departure of the second leg. This could magnify the small delay on the first leg to an entire headway of the service on the second leg.

The WLL is very close to the NLL in terms of mean EJT, whereas it was dwarfed in terms of total EJT. The implication is that journeys on the WLL are subject to delays of similar (average) magnitudes, but many fewer journeys are affected. Unlike on the NLL, the normalization by total passengers does change the relative picture for the GOB. While total EJT is greatest in the AM and PM Peak periods, the highest average EJT is experienced by passengers during the Early morning period. These relative differences do not necessarily mean that one time period or line is more worthy of attention than the other. Rather, it presents a more nuanced picture of service quality which can be acted upon differently depending on management policies and priorities.

EJT is net negative on the WAT in the Early and AM Peak periods. To understand this further, EJT of WAT passengers was investigated at the level of individual origin-destination (OD) flows. In the AM Peak, 46 out of 236 OD flows on the WAT had net negative EJT, including *all* 15 OD flows into Euston, the line's southern central London terminus. These flows into Euston account for 93% of net negative EJT on the 46 net negative OD flows. Almost half of that net negative EJT into Euston comes from the flows originating at Queens Park and Kensal Green (towards the southern end of the line), both with average EJT of almost -3.0 minutes. This is explained, in consultation with Overground management (Bratton, 2010) by the fact that WAT trains often depart Queens Park on time or slightly late and arrive Euston terminal up to 5 minutes early. In other words, their scheduled running time over the last portion of the line is generous.

Another quarter of the net negative EJT into Euston comes from the OD flows originating between Watford Junction and Harrow & Wealdstone (at the northern end of the line), non-inclusive, which have a mean EJT of -5.1 to -12.4 minutes. This likely represents a problem with the assignment model used to filter non-Overground passengers. Another National Rail service provides twice-hourly express service from Watford Junction and Harrow & Wealdstone to Euston in substantially less time than the Overground. The assignment model correctly assigns passengers from these two stations to that service, but not for passengers who start on the Overground and interchange to this express service, perhaps opportunistically, at Harrow & Wealdstone. Further discussion of these two sources of net negative EJT, and their implications for EJT as a tool for service quality measurement and tactical planning, are postponed until later in this chapter.

In general, these aggregate results are in line with a priori expectations held by the management of the Overground network (*e.g.* Bratton, 2008). The most strongly held belief, confirmed here, is that the NLL carries the largest passenger loads and has the most delays, especially during the peak periods.

Time Series of Mean EJT

Figure 8-6 shows daily mean EJT over time for each London Overground line, for the whole day and for the AM Peak period. On all lines, there is marked day-to-day and week-to-week variability of EJT. As expected, mean EJT exhibits some volatility on a day-to-day basis. This is particularly the case as sample sizes decrease, namely in the AM Peak compared to the whole day, and for the WLL and INT compared to the other lines. There is not a clear up or down trend over time in this dataset suggesting that overall relative service quality on the Overground was steady over this period.

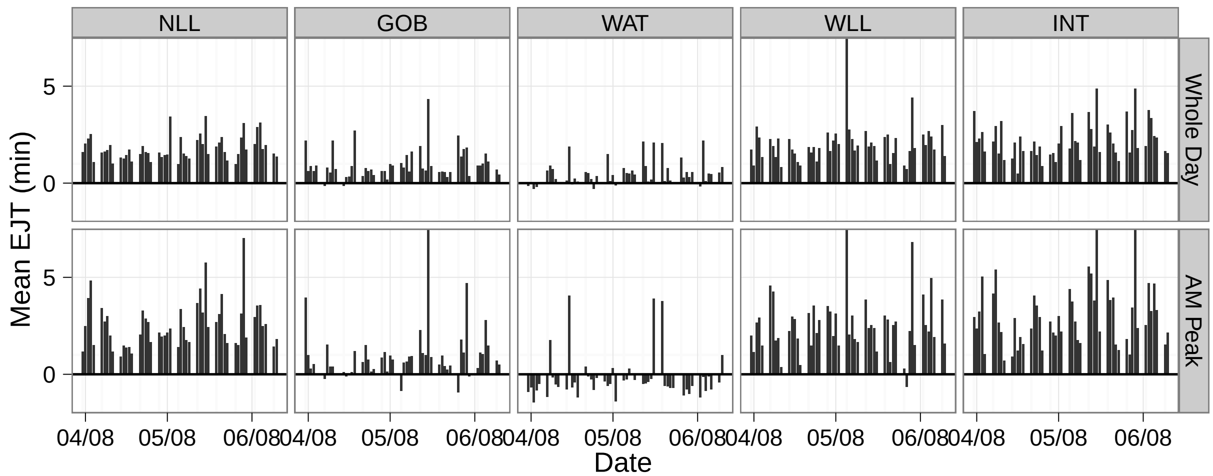


Figure 8-6: Daily Mean EJT

Mean and Total EJT by Time Period and Direction (NLL)

Figures 8-7 and 8-8 further disaggregate EJT results for the NLL by direction of travel. This aggregation is important because of the unbalanced nature of passenger demand on the NLL (and indeed on many railways) in different periods of the day. Figure 8-7 shows total EJT to be substantially worse in the westbound direction than in the eastbound direction in the AM Peak period. Figure 8-8 shows mean EJT to be similarly unbalanced, though somewhat less so than total EJT, in the same period. This indicates that, in the AM Peak period, there are more passengers suffering longer delays in the westbound than the eastbound direction. Similar results can be seen in the PM Peak period with the directions reversed, though the unbalance is not nearly as severe as in the AM Peak.

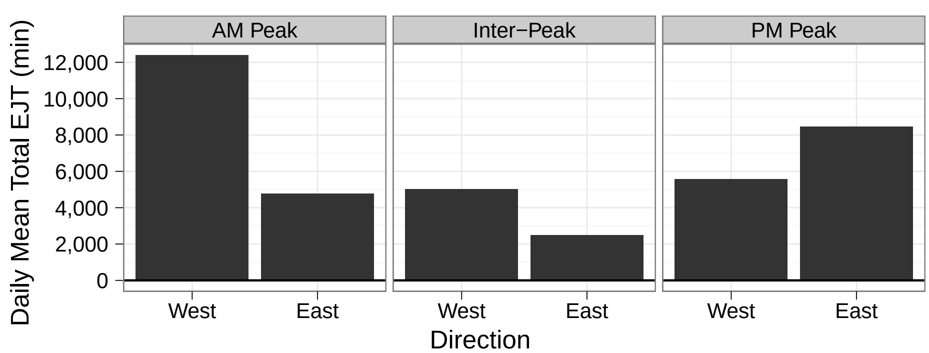


Figure 8-7: Total EJT on the NLL, by time period and direction

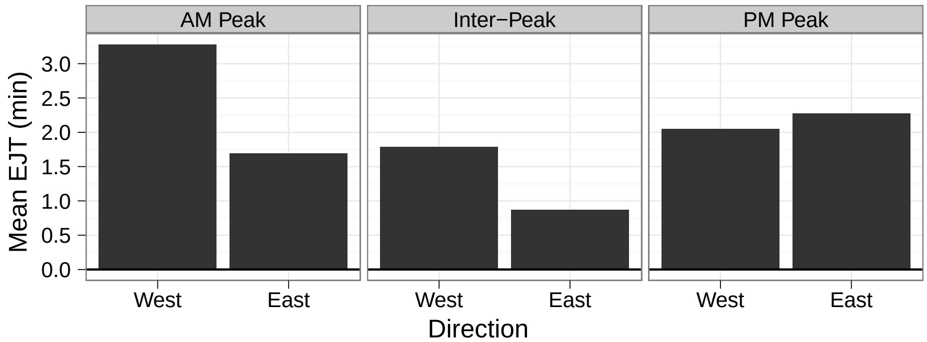


Figure 8-8: Mean EJT on the NLL, by time period and direction

Mean and Total EJT by Scheduled Service (NLL AM Peak)

One advantage of the approach proposed here is that, with such a large and detailed data set, it is possible to probe deeper into the specifics of delays and their effects on passengers. To estimate EJT, each passenger journey was assigned to a specific scheduled service. This assignment indicates only which train a given journey would likely have taken under right-time service delivery, not which train the passenger actually rode. In that sense, each scheduled service defines a specific market over time and space, and the assignment places journeys into these markets.

Figures 8-9 and 8-10 aggregate EJT to the level of these markets. They show total and mean EJT, respectively, for westbound passenger journeys on the NLL between Stratford and Willesden Junction.[5] These are the peak London Overground markets – the peak direction (from Stratford towards Richmond) at the peak time of day on the busiest line – which were shown in Figures 8-4 and 8-5 to have the most severe EJT problems on the whole Overground network. The bars in these plots are spaced according to the actual headway (at Stratford) of each service.

Figure 8-9 clearly shows that the 07:52 and 08:22 trains are the most problematic services in terms of total passenger delay. It also shows how unbalanced the headways in the timetable are, especially between Stratford (SRA) and Camden Road (CMD).[6] Between 07:00 and 09:00, the services to Richmond with full 15 minute headways have the highest total EJT relative to their shorter-headway leaders and/or followers. This observation, along with the finding in Chapter 6 that passengers on the NLL display incidence behavior that is largely random, suggests a causal relationship. Because of (mostly) uniform passenger arrival rates, the respective markets of services with longer leading headways will contain relatively more passenger journeys. Unbalanced market sizes can translate into unbalanced passenger loadings and overcrowding on some trains, with potentially serious implications. Overcrowded trains can extend train dwell (and thus running) times at stations and also impact passengers' ability to board their desired trains. This dynamic, and its implications for tactical planning, is further explored in the following chapter.

[5] Journeys bound for west of Willesden Junction are excluded because, while most NLL services go to Richmond, one per hour in the AM Peak divert at Willesden towards Clapham Junction.

[6] These unbalanced scheduled headways are also visible in the time-distance plots in Figures 8-1 and 8-2.

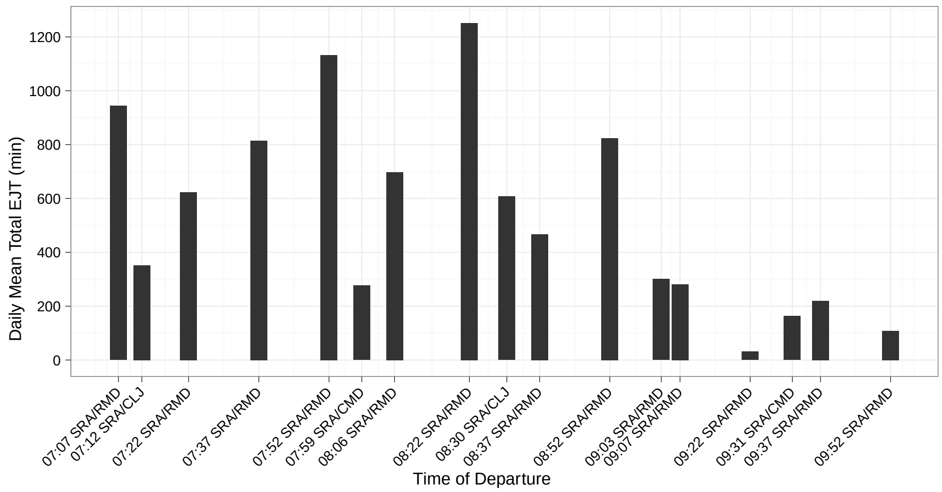


Figure 8-9: Total EJT by scheduled service, westbound

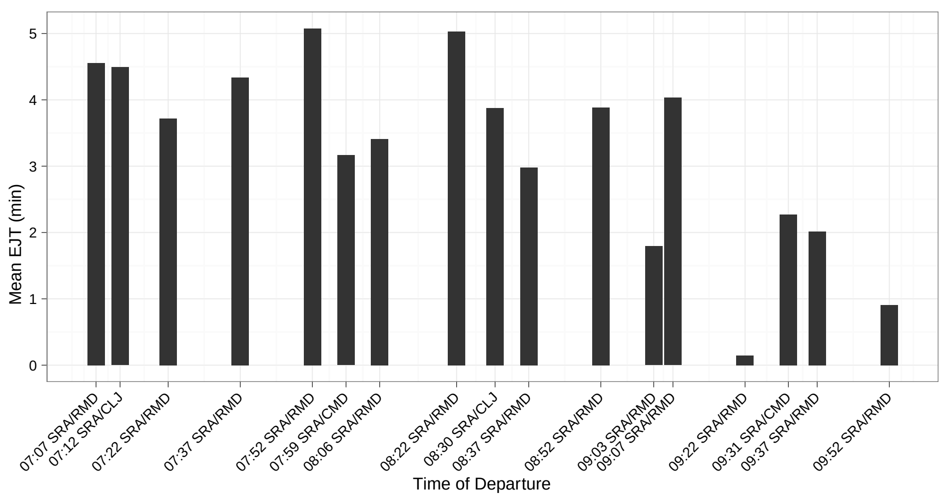


Figure 8-10: Mean EJT by scheduled service, westbound

The mean EJT results in Figure 8-10, normalized by the total number of journeys in each market, show similar results. The most substantial relative differences are for the short headway services at 07:12 and 09:07. Their mean EJT is much higher compared to other services than was their total EJT. This stands to reason, in that with short headways they should have fewer journeys and thus less total EJT.

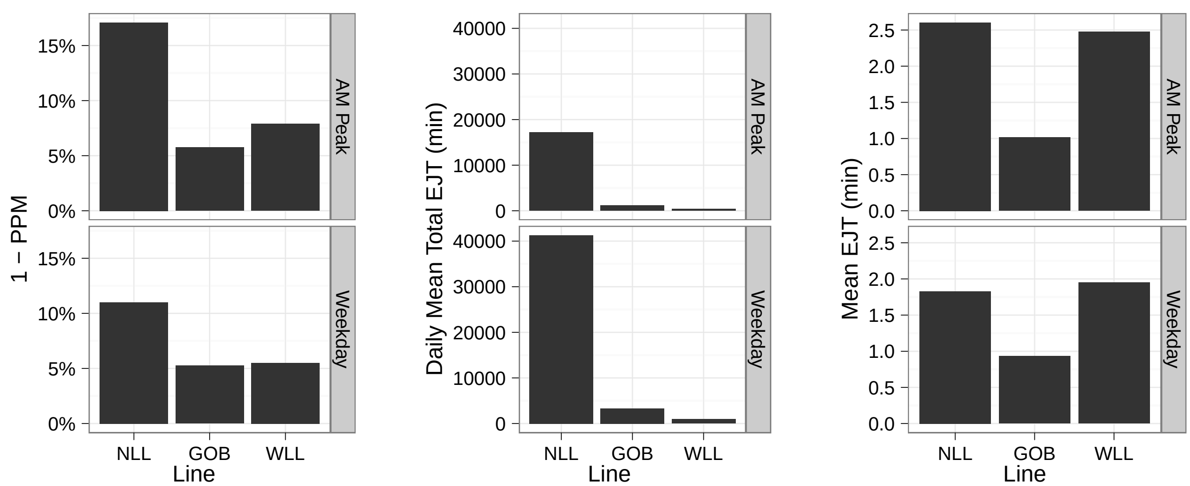
8.3.2 Comparison with Existing Performance Metrics

This section compares EJT measures with corresponding on-time performance (OTP) results from the existing London Overground performance regime, the Public Performance Measure (PPM). These comparisons are presented as much to explore the differences between EJT and OTP as public transport performance measures as to characterize performance on the

Overground.

Figure 8-11 plots PPM and total and mean EJT by line,[7] for the AM Peak and for the whole day, over the entire study period. The plot shows the complement of PPM so that the measures are directionally aligned (*i.e.* a higher number indicates worse performance). PPM and total EJT correspond in that the NLL is by far the worst performing line. The difference between the NLL and other lines is even more pronounced in terms of total EJT than in terms of PPM. This reflects the difference in passenger volumes between the different Overground lines.

The WLL appears much worse than the other lines in terms of mean EJT than it is in terms of PPM. This could be because the WLL has the lowest frequency of the Overground lines. Consequently, each delayed train (counting against PPM) may have a greater proportionate effect on the line's passengers. It could also be that the manner in which WLL trains are delayed has worse effects on passengers than on the other lines.

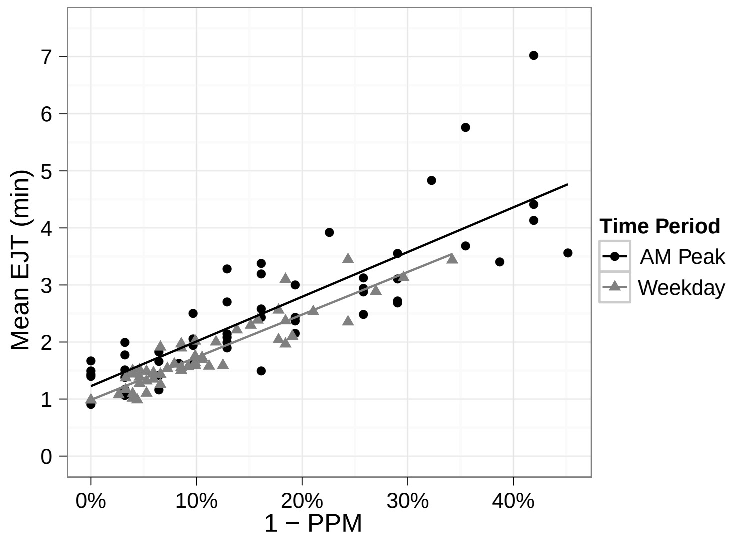


(a) PPM (b) Total EJT (c) Mean EJT

Figure 8-11: EJT and PPM, by line

Figure 8-12 plots mean EJT against (the complement of) PPM for the North London Line, for the whole day and for the AM Peak, for the 52 weekdays in the study data set. As expected, the measures are substantially but not perfectly correlated. Clearly the OTP of NLL trains will affect the EJT of their passengers, but it is expected that EJT will capture additional information about passenger journeys that OTP does not. In both the whole day and the AM Peak cases there is clear variation of EJT around the linear regression fit with PPM (adjusted R^2 of .84 and .69, respectively), indicating that there is additional information captured in EJT measurements. The variation is greater in the more disaggregate AM Peak calculations, and tends to increase as PPM decreases. This pattern suggests that PPM reflects the passenger experience less effectively as conditions worsen.

[7] The WAT is excluded here because of the problems with the assignment model discussed in the previous section.



Lines are simple linear regression fits. Adjusted R^2 for Weekday and AM Peak are 0.84 and 0.69, respectively

Figure 8-12: Mean EJT vs PPM, for NLL

8.4 Conclusions

This chapter presented an analysis of aggregate and disaggregate EJT results for the London Overground, both in isolation and in comparison to the Overground's existing measure of service delivery. This section presents some conclusions drawn from this analysis, first regarding EJT as a measure of service quality in general, and second regarding EJT on the Overground network itself.

As a Measure of Relative Service Quality

EJT for individual passenger journeys can be estimated through the combination of journey data from AFC systems with published timetables by means of schedule-based assignment. This method takes into account the scheduled and actual service delivery on the network over the time and space of each passenger journey. The requisite schedule-based assignment can be achieved efficiently through the use of free software tools and timetables in open standard formats.

EJT for individual passenger journeys on a given service was found to range from negative (*i.e.* early) by up to one headway to positive (*i.e.* late) by substantially more than one headway. A negative EJT for a single journey does not in and of itself represent a problem or inconsistency in the measurement scheme. However, it is difficult to interpret EJT for individual journeys, in part because of the ambiguity with respect to passenger's standards and incidence behavior discussed in Chapter 7. Consequently, EJT is not a particularly useful measure to analyze individual passenger journeys.

Aggregate EJT, on the other hand, is a measure of relative service quality with clear meaning. It expresses the average passenger's experience in terms of total journey time compared to what the timetable would imply, for a wide range of passenger incidence behaviors. Individual EJT measurements are, by nature of the assignment process by which they are estimated, easily aggregated both spatially and temporally. Depending on the analytical need, aggregate EJT can be estimated at the level of line, origin-destination flow, scheduled service, time period (*e.g.* AM Peak), day, week, etc.

When aggregated, EJT is by its very nature passenger-weighted. Different aggregation functions (*e.g.* mean or sum) can be used to present different types of passenger orientation. Total EJT accounts for relative volumes of passengers between different aggregation variables (*e.g.* lines, times of day, etc). It can be used to assess where and when management and planning attention could make the most overall progress towards aligning service delivery and service quality for *all* passengers. Mean EJT on the other hand normalizes by number of passengers. It can be used to assess relative service quality for the *average* passenger for each aggregation variable. Neither total nor mean EJT is a better analysis tool *per se*; they simply represent different analytical goals and priorities. Both aggregations capture certain information about the passenger's experience that a service delivery measure such as on-time performance (OTP) does not.

Aggregate EJT was observed here to be net negative as a result of two of the three causes discussed in Chapter 7. Firstly, trains can regularly run ahead of schedule. In this case, aggregate EJT indicates that there is extra running time in the timetable. This illustrates a difference between EJT and OTP – in the case of EJT, less is not always more. A negative EJT can be as valuable an indicator as a positive EJT that tactical planning attention is required. OTP, on the other hand, can never go beyond 100%, and so will not explicitly indicate excessive running time in the timetable.

Secondly, aggregate EJT can be net negative because of problems with the assignment models used to estimate EJT for individual journeys. Incorrect modeling assumptions or missing timetables, for example, can result in passengers being able to reach their destinations by a different (and substantially faster) route than the assignment process implies. This is an error, or bias, in the estimation process. It highlights the sensitivity of EJT to certain behavioral aspects of the assignment model.

No evidence was found of the final cause of net negative EJT – an outsized proportion of passengers being incident shortly after a scheduled departure time to catch a train that they expect to, and that does, depart late. Such could still occur in practice.

This chapter has not adequately explored higher order statistics of distributions of EJT. These statistics have important implications for measuring reliability as one aspect of relative service quality, and deserve further attention in future work.

On the London Overground

Aggregate EJT was found to vary substantially across the different London Overground lines and across time periods of weekday service. Total EJT is greatest on the North London Line in the AM and PM Peak periods, which also had among the highest estimates of mean EJT. Passengers on this line during the peak periods suffer the most total EJT because there are many more of them and because, on average, their journeys are the most delayed relative to the timetable. The North London Line in the peak periods is, by this measure, especially deserving of management and tactical planning attention. This is consistent with

PPM (*i.e.* OTP) results.

Further disaggregating EJT by direction, it was found that total and mean EJT on the NLL in the AM Peak were substantially worse in the westbound direction (*i.e.* from Stratford), driving the overall NLL AM Peak results. EJT was disaggregated again to the level of individual scheduled services (in the westbound direction), where it was found that total EJT was generally higher on those services with full 15 minute headways than on their shorter headway leaders and followers.

Interchange passengers were found to have higher mean EJT than passengers on any single line, particularly during the AM Peak. This is an indication that smaller delays on individual lines lead to missed interchanges and thus longer delays for interchange passengers. However, there are relatively few passengers interchanging between Overground lines, so the total EJT attributed to them is small. The experience of interchange passengers is not explicitly captured by PPM measurements.

The West London Line has mean EJT levels comparable to the North London Line, despite having a substantially better PPM. The Gospel Oak to Barking Line is similar to the other lines under both PPM and mean EJT. Line-level EJT results for the Watford DC Line are not reliable because of the aforementioned problems with the assignment model. Analyzing EJT for certain origin-destination flows on the WAT (those with little room for assignment error) does indicate that this line has excessive running time between some of its southernmost stations and Euston terminal.

EJT has the potential to be a useful input into the analytical tactical planning process, as is explored in the following chapter. However, it is not possible to conclude at this point that EJT is appropriate for use in contract or performance management on the Overground or on other networks.

Chapter 9

Tactical Planning Case Study on the London Overground

During the time that the research for this thesis was being conducted, London Overground management researched, designed, and implemented a new tactical plan for AM and PM Peak service on the North London Line. This chapter documents that tactical planning intervention, which was influenced by some of the preliminary results of this thesis, and evaluates its outcome in terms of certain aspects of service delivery and service quality.

This chapter is presented as a case study in that most of the work it describes was conducted by other analysts and professionals. It depends heavily on in-person and e-mail interviews with key Overground managers and on research conducted for those managers by an industry consultant. The goal here is to illustrate how the methods developed in this thesis for using automatic data can contribute to real-world tactical planning processes considering a range of decision factors and variables. It is descriptive in nature, rather than prescriptive.

Section 9.1 describes some aspects of the service plan, passenger demand, and operating performance on the North London Line at the time of the tactical planning exercise. Section 9.2 describes how understanding of and relationships between these factors was synthesized to guide the development of a revised tactical plan. Section 9.3 evaluates the outcomes of the implementation of this plan using longitudinal before-and-after analysis. Section 9.4 draws some conclusions regarding this evaluation, including its use of service delivery and service quality measurements.

9.1 The North London Line: Spring 2008

This section describes some relevant information about the North London Line as of the Spring of 2008, first in terms of the existing service plan, next in terms of passenger demand, and last in terms of operating performance as expressed by different measures of service delivery and service quality.

9.1.1 The Service Plan

The North London Line (NLL) is the backbone of the London Overground network. It serves 23 stations running 28 kilometers circumferentially around central London from Stratford in the northeast to Richmond in the southwest. It connects to the Gospel Oak to Barking Line (GOB) at Gospel Oak station (GPO) and to the Watford DC Line (WAT) and West London Line (WLL) at Willesden Junction station (WIJ). Figure 9-1 schematically illustrates the AM Peak service patterns and frequencies on the NLL (and other Overground lines) in Spring 2008.

Figure 9-2 uses a time-distance plot to show the published AM Peak timetable for the NLL in Spring 2008.[1] This plot shows a regular 15 minute headway (4tph) service making all stops westbound from Stratford to Richmond and eastbound from Richmond to Stratford. This regular service is augmented by occasional irregular additional services that split some of the 15 minute headways into two smaller headways (of 7 and 8 minutes, 9 and 6 minutes, etc). These irregular services include

- "Camden shuttles" that run approximately hourly between Stratford (SRA) and Camden Road (CMD) (departing Stratford at 07:59 and 09:31);
- "Clapham specials" that run approximately hourly between Stratford and Clapham Junction (CLJ), diverting from the NLL to the WLL (not shown in plot) at Willesden Junction (departing Stratford at 07:11 and 08:30);
- one full NLL service from Stratford to Richmond (RMD) at 09:02.

This timetable was developed before TfL and LOROL controlled the Overground network, when it was a standard National Rail concession operated by the Silverlink TOC. David Warner (2010), a planner in TfL London Rail, noted that the Camden shuttles were added in 2004, and the Clapham shuttles in 2006. Oliver Bratton (2009), Head of Performance and Planning for LOROL, noted that "TfL was getting concerned about the overcrowding on the NLL (even though it was a DfT franchise). It therefore agreed to 'buy' additional services from Silverlink for the peaks." To add the additional trips, Silverlink planners "put them into the existing schedule ... amongst the 15 minute service when appropriate."

Describing the origins of this timetable, Warner (2010) noted that "the service was entirely driven by the rolling stock available, and the incremental nature in which additional trains were added to the timetable. An overall view was not taken." Bratton (2009) discussed the timetable development process for the National Rail network more generally, noting that "typically, a timetable evolves. As more and more trains run, the timetable tends to ossify, becoming harder and harder to alter." It appears from these comments that the highly irregular NLL peak period timetable was in place largely as a historical artifact. It was not the product of an analytical or data-driven tactical planning process.

[1] Appendix A contains a list of station abbreviations.

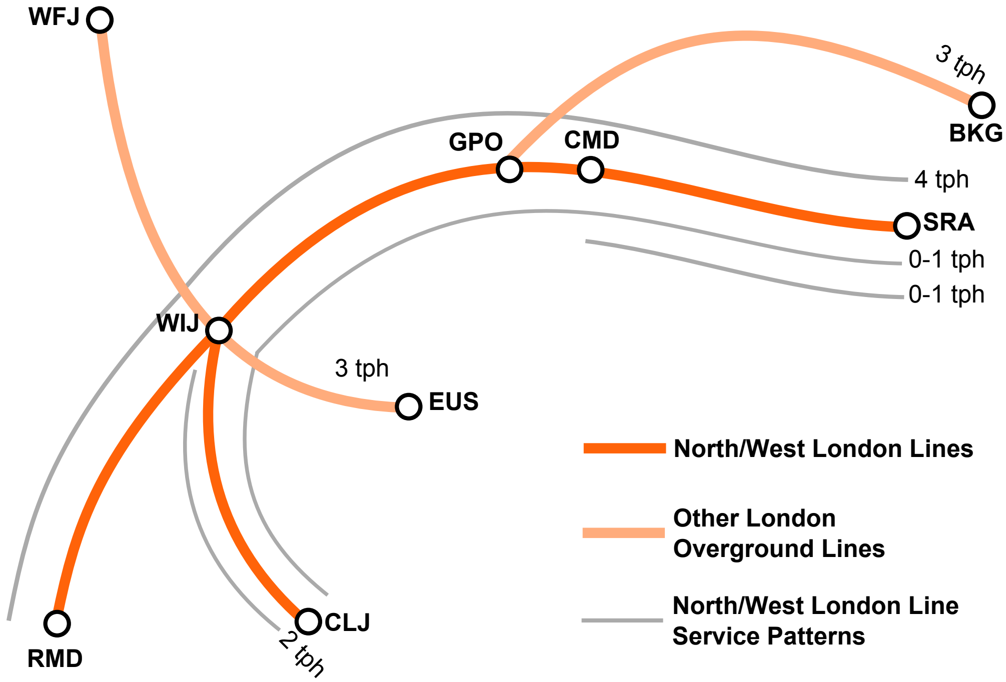


Figure 9-1: London Overground Spring 2008 AM Peak service patterns and frequencies

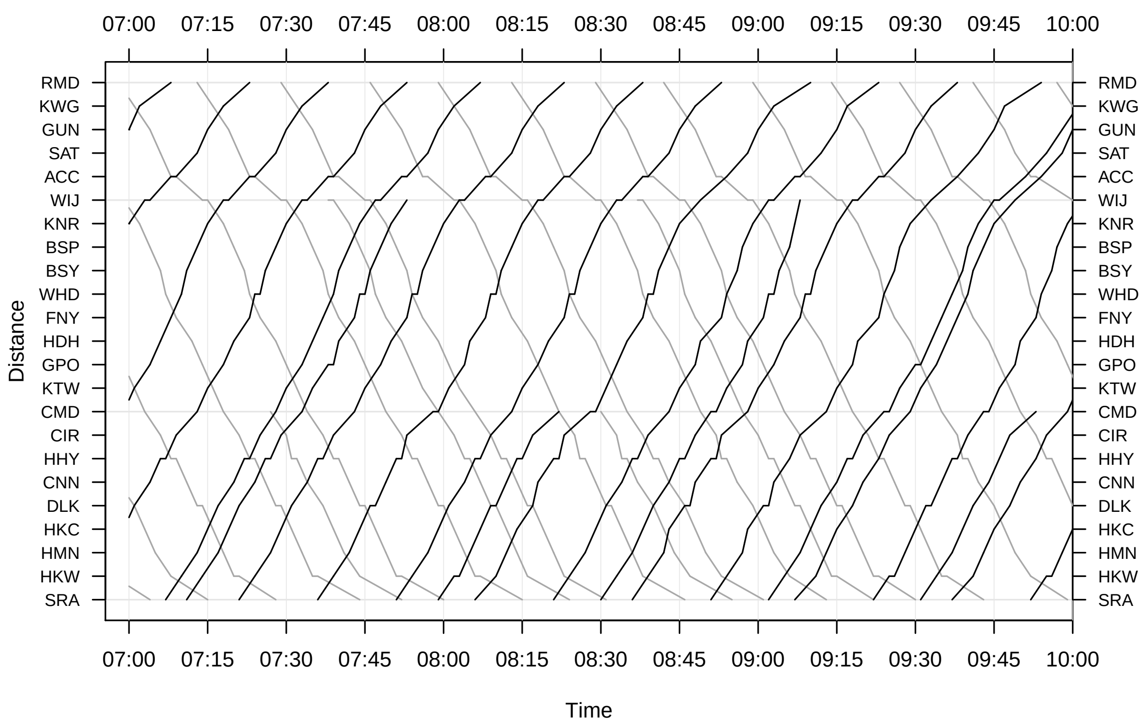


Figure 9-2: North London Line Spring 2008 AM Peak timetable

9.1.2 Passenger Demand

NLL AM Peak Origin-Destination Matrix

Table 9-1 shows a segment-level[2] AM Peak origin-destination (OD) matrix for the NLL. It is a summary of the results of Chapter 5, which were estimated from passenger journey data from the Oyster smartcard ticketing system, automatic entrance counts from selected station gatelines, and manual on-board passenger counts.[3] This matrix includes journeys that interchange with the NLL, but it "clamps" those journeys to the point at which they would make that interchange (*i.e.* at Gospel Oak or Willesden Junction).

Origin	Destination Segment					
Segment	NLLE	GPO*	NLLC	WIJ*	NLLW	Total
NLLE	9,309	369	1,848	392	517	12,435
GPO*	158		356	94	145	753
NLLC	1,362	268	778	596	1,019	4,023
WIJ*	524	83	465		808	1,880
NLLW	687	63	678	359	1,234	3,021
Total	12,040	783	4,125	1,441	3,723	22,112

Cells highlighted in grey depend only on NLL service West of Willesden Junction
* Includes flows interchanging between NLL and other Overground lines at these interchange stations.

Table 9-1: Segment level NLL AM Peak origin-destination matrix

Section 5.5 of Chapter 5 estimated that a total of 37,124 passengers use the London Overground on an average weekday AM Peak period. The NLL OD matrix in Table 9-1 shows that 22,112, or 60%, of all AM Peak Overground passengers use the NLL for some portion of their journey. The cells of Table 9-1 highlighted in grey indicate passenger flows which use the NLL only between Willesden Junction and Richmond (RMD), inclusive. They total 5,510 AM Peak passengers, or 25% of total NLL AM Peak patronage. In other words, 75% of all AM Peak NLL passengers use the NLL only between Stratford and Willesden Junction, inclusive.

This OD matrix was estimated on data from Spring 2009, after the tactical planning change described in this chapter was analyzed (but before it was implemented). It is included here because (i) it is representative of the rough Oyster-only analysis that was conducted at the time; (ii) the same OD estimation *could* have been conducted on 2008 data; (iii) such an estimation would likely have produced very similar results for the Spring of 2008, especially in terms of OD distributions; (iv) this chapter is intended to illustrate how the methods developed in this thesis *can* be used in practice.

[2] Appendix B describes the line and segment abbreviations used here.

[3] As described in Chapter 5, the manual on-board counts are intended to be replaced with automatic estimates from loadweigh data.

Aggregate Load Profiles (NLL AM Peak)

Figure 9-3 plots the aggregate load profile for the NLL (in each direction) as measured by the manual on-board passenger counts that were used to estimate the above OD matrix. The most salient observation to be drawn from these plots is that by far the largest aggregate link loads on the NLL during the AM Peak are westbound between Stratford and Highbury & Islington (HHY). The aggregate load starts at over 4,000 total passengers out of Stratford and grows at each successive station, peaking at close to 6,000 total passengers between Canonbury (CNN) and Highbury & Islington.

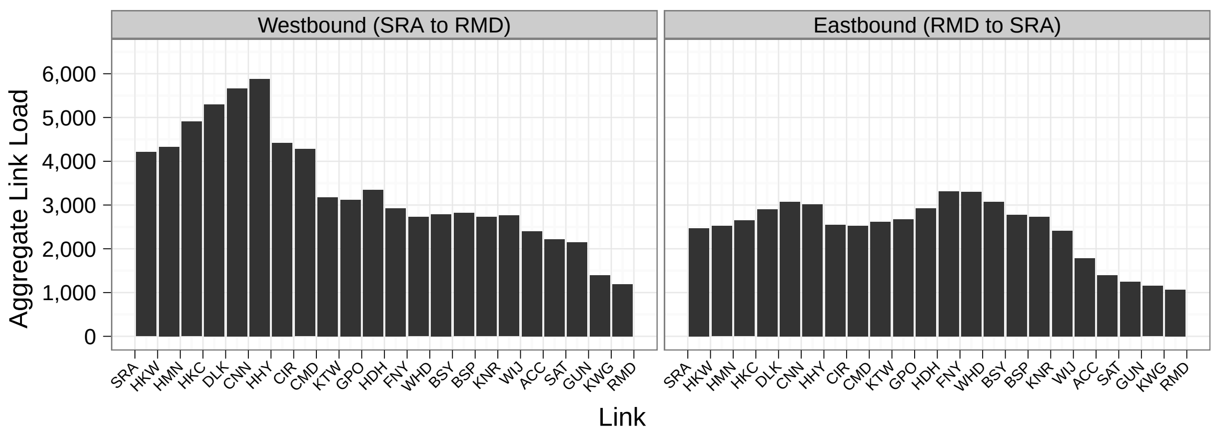


Figure 9-3: NLL AM Peak load profile

Passenger Incidence Behavior

Section 6.3.3 of Chapter 6 found that the temporal distribution of passenger demand on the North London Line was much less timetable-dependent than on the other London Overground lines. That is, passenger incidence behavior was more random on the NLL and less a function of the timetable. Figure 9-4 plots a small selection of the passenger incidence results from Figure 6-4 illustrating the difference between the NLL and the GOB in terms of the respective distributions of passenger incidence times (over a given headway) in the AM Peak.

Section 6.3.3 also quantified the implications of these distributions in terms of their effects on passenger waiting time (with respect to the timetable). It found that the dependence on the timetable observed on the GOB in the AM Peak reduced waiting time (with respect to the timetable) by 29% compared with purely random passenger incidence. On the NLL the comparable reduction was only 7.2%.

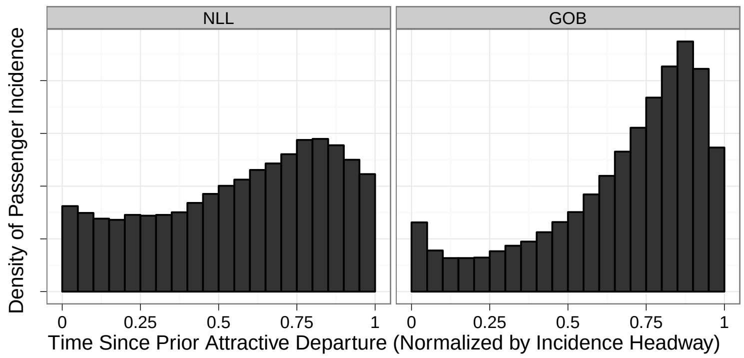


Figure 9-4: AM Peak passenger incidence on NLL and GOB

9.1.3 Operating Performance: Service Delivery and Quality

PPM

Service delivery on the London Overground network is currently characterized in terms of PPM – the fraction of trips making all scheduled stops and arriving at their terminal no more than 5 minutes late. Section 8.3.2 of the previous chapter presented PPM results for the 52 weekdays from 31 March, 2008 through 10 June, 2008, inclusive. PPM for the NLL over this range of dates was 83% for the AM Peak period and 89% for the whole weekday. This was the worst of all of the Overground lines.

Running and Dwell Times

ACT, A British railway consultancy, was retained by LOROL to study operations on the NLL (ACT, 2008). They used automatic train movement data from the network's signaling and control system to analyze running and dwell times[4] from April, 2007 through March, 2008, inclusive. As TfL and LOROL took control of the Overground network in November of 2007, this study included a period of substantial institutional and branding change on the NLL. ACT's observations most relevant to this chapter can be summarized as follows.

- Over the course of the study period, increases in terminal-to-terminal running times were observed in the AM and PM Peak periods in both directions. The largest increase of 4% was observed in the AM Peak in the westbound direction (*i.e.* from Stratford).

- For both directions in the peak periods, the 80^{th} percentile running time is between the running time in the timetable and the PPM threshold (*i.e.* timetable plus five minutes); the 90^{th} percentile running time is above the PPM threshold.

- At the 80^{th} and 90^{th} percentiles, the peak period running time westbound (*i.e.* from Stratford) exceeds the peak period eastbound running time by just over two minutes.

[4] Dwell time, the time at which a train is stopped at a station, can be thought of as the segment running time between the arrival and departure at that station.

- Dwell times increased (by an unspecified amount) over the course of the study period, especially during the peak periods.

- As measured during only the first quarter (*i.e.* January through March) of 2008, there was substantial "dwell time loss" – station dwell times in excess of those specified in the working timetables. Nine out of the top ten westbound (*i.e.* from Stratford) scheduled services in terms of average dwell time loss were in the AM Peak period. Likewise, nine out of the top ten eastbound (*i.e.* from Richmond) scheduled services in terms of average dwell time loss were in the PM Peak period.

- The worst five westbound services had average dwell time losses of 311 - 459 seconds. They are, in order of decreasing least dwell time loss, the 07:52, 08:06, 08:22, 07:37, 08:52 trains from Stratford. For these services, the largest single-station dwell time loss was experienced at Canonbury, which is the station with the peak departing load (as seen in Figure 9-3) on the NLL.

- A statistical correlation was found, for individual station stops by individual scheduled services, between the length of the leading headway and the station dwell time.

From these observations, the consultants drew two important conclusions regarding service delivery on the NLL. Firstly, that running times in the timetable are insufficient, particularly for the AM and PM Peak periods. Secondly, that dwell times, and corresponding dwell time losses, were driven by passenger demand. The consultants provided only this analysis of current conditions and their conclusions as to what may have been causing those conditions. They did not offer any explicit recommendations on what actions should be taken to improve those conditions.

Train Congestion

Two types of train congestion[5] were also observed on the London Overground network – that between Overground trains, and that between Overground trains and freight trains on the NLL. Bratton (2009) noted that "turning trains at Camden was causing congestion on the network" between Overground services. This congestion has not been studied or quantified directly, only reported anecdotally by Overground management and operational staff. However, based on examination of the timetable in Figure 9-2, it is not hard to believe that such congestion was occurring. The two Camden shuttles that depart Stratford at 07:59 and at 09:31 both arrive at Camden Road within a few minutes of an eastbound train from Richmond. Under these circumstances, even slight deviations from schedule could cause congestion and delays to the trains from Richmond, to the Camden shuttle on its return trip to Stratford, or to the subsequent westbound train from Stratford.

It is of course also possible that the NLL suffered from additional types and instances of train congestion, especially at other junctions and at terminals. Congestion between Overground trains at Camden Road and other locations could have been identified and studied in further detail through the use of time-distance plots (as in Rahbee, 1999, 2006; Vescovacci,

[5] Train congestion can also be considered an element of service delivery; it is simply the extension of certain segment running times with a specific causal explanation.

2001; Lee, 2002, for the Boston and Chicago metros). Such plots can be generated from automatic train movement data recorded by the Overground network's signaling and control systems (that were used to study running and dwell times), rather than from timetables as in Figure 9-2.

The study by ACT (2008) also examined freight services running on the NLL. These services join the NLL at one of several "boundary" junctions between the NLL and the larger National Rail network, run on the NLL for some length, and depart the NLL at another boundary. The boundaries most commonly used by freight services to join or depart the NLL are, in descending order, at Stratford, west of Camden Road, west of Gospel Oak, and east of Acton Central. ACT found that, between 06:00 and 23:00, over 55% of freight services were more than 5 minutes late or early at their boundary with the NLL.

In the same study, ACT also used data from the National Rail train incident and delay tracking system to assess the impact of non-punctual freight trains on passenger services. Their findings indicated that freight trains joining the NLL with larger deviations from their respective timetables had a larger chance of causing a delay to Overground NLL trains. Bratton (2009) reported anecdotally that "there was no ability to absorb late running freights" during the shorter headways in the timetable.

Excess Journey Time

Excess journey time (EJT) for the NLL as of the Spring of 2008 was examined in detail in Section 8.3 of the previous chapter. The relevant results can be summarized as follows.

- The NLL in the AM Peak had by far the most total EJT of any Overground line in any time period, followed by the NLL in the PM Peak.

- In terms of mean EJT (*i.e.* after normalizing by number of passengers) the AM Peak was the worst time period for the NLL.

- In terms of total and mean EJT, the NLL in the AM Peak had over twice the EJT in the westbound direction (*i.e.* Stratford to Richmond) as in the eastbound direction.

- Total EJT was severely unbalanced among individual westbound AM Peak scheduled services on the NLL. The five scheduled services with the highest total EJT were, in descending order, the 08:22, 07:52, 07:07, 08:52, and 07:37 trains from Stratford. These services had full 15 minute headways (at Stratford) and had substantially more total EJT than their respective shorter-headway leaders and/or followers.

- Mean EJT on the NLL was somewhat linearly correlated with PPM. However, the relationship was weaker in the AM Peak than for the whole day, and in the AM Peak for lower PPM values.

9.2 Tactical Planning Intervention: The Case for Even Intervals

This section synthesizes the results of the previous section to explain the thinking behind a specific tactical planning intervention that was implemented on the NLL. It does so primarily from the perspective of the London Overground manager who was the driving force behind this change. This manager was influenced by the results of some of the early research from this thesis, and used some of these results to make the case to his stakeholders.

Certain factors in the decision to implement this change, such as complete costs and benefits, were not available for this case study. In addition, the manager's perspective is supplemented here by analysis drawn from the final results of this thesis. In that sense, the case this section makes for the tactical planning change is not precisely the case that was made in practice. Nevertheless, it provides a context in which to illustrate the value of the methods developed in this thesis for using automatic data to support the tactical planning of an urban railway.

Of the results and analyses discussed in the previous sections of this chapter, the key points that influenced the tactical planning intervention on the NLL are as follows. They are focused on the AM Peak period, in which, as of Spring 2008, the NLL had the most problems with performance.

- NLL trains were routinely delayed en route, as reflected by running times substantially in excess of the timetable and by low PPM on-time performance scores.
- Excessive dwell times were found to be a major cause of en route train delays. They were particularly problematic for westbound services from Stratford, for those services with full 15 minute scheduled leading headways, and at those stations with the highest departing passenger loads.
- Evidence existed to support the judgment that these dwell times were primarily a function of passenger volumes.
- Near-random passenger incidence behavior suggested an explanation for uneven passenger volumes and resultant uneven dwell times – when passengers arrive (even approximately) randomly, services with longer headways will serve proportionally more passengers.
- The confluence of these analyses is confirmed by the corresponding EJT results:
  - In the AM Peak, mean and total EJT were substantially higher for passengers traveling westbound than for those traveling eastbound.
  - Moreover, total EJT for passengers in the market for the services with full 15 minute leading headways was substantially higher than for the services with shorter leading headways. This was a product of the volumes of passengers in those markets and the average delay to each of those passengers.
  - The set of scheduled services with the highest total EJT in the AM Peak was nearly congruent with the set of services with the longest dwell times in the same time period.

- These results also reflected the delays to any passengers who may have been unable to board crowded trains.

- Additionally, there was congestion caused by the reversing of shuttle services from Stratford at Camden Road and by freight trains arriving at the NLL off-schedule during short headways in passenger service.

As described in Section 9.1.1, the typical approach to tactical planning on this network was to update the timetable incrementally. It became evident that more drastic measures were needed in this case. Specifically, that the timetable should be revised wholesale to provide *as even headways as possible*. Under the circumstances, it was proposed to achieve this by combining the NLL and WLL during the AM and PM Peak periods into an even 10 minute headway (6tph) service between Stratford and Willesden Junction. From Willesden, alternating trains would go on the NLL to Richmond and the WLL to Clapham Junction. This was referred to as the "3 + 3" service.

The core idea behind this strategy was to balance passenger volumes across trains, thus reducing dwell times and train and passenger delays. It was expected that passenger volumes would increase on some trains (*i.e.* those with longer headways than in the current plan) and decrease on others (*i.e.* those with shorter headways). While the change was not expected to materially affect the total volume of passengers, the outcome was expected to be positive on balance. The reasons for this are twofold.

Firstly, dwell times have been found to have a non-linear relationship to passenger volumes (*e.g.* Wilson and Lin, 1993; Puong, 2000, which found dwell time to respond as the square or cube of the number of standing passengers). This implies that, holding the total volume of passengers constant, the decreases in dwell times on trains losing passengers (*i.e.* the existing 15 minute headway services) would be larger than the increases in dwell times on trains gaining passengers (*i.e.* the shorter headway services). Consequently, overall delays should decrease. While the managers of the Overground may not have been aware of the literature and mathematical modeling of dwell time. It is likely that they were intuitively aware of its dynamics.

Secondly, consistent with the hypothesis of unbalanced on-train loads, it was anecdotally reported that some passengers were unable to board overcrowded trains were be left behind on the platform. Under the new timetable, some trains would be more crowded than under the existing timetable but still below capacity. Others would be less crowded and allow more (or all) passengers to board. Passengers currently denied boarding by overcrowded trains would thus benefit substantially while other passengers would benefit (from less crowded trains) or suffer (on more crowded trains) to a lesser degree.

Additionally, with no change expected in the punctuality of freight services using the NLL, the even interval strategy would minimize the chance that a freight service would have to be fit in between two passenger services separated by only 6 or 7 minutes.

While the decision to move to the "3 + 3" service is presented retrospectively in a logical and linear fashion, the strategy was in fact realized somewhat serendipitously. Bratton (2009) described the process by saying that "although it wasn't obvious at the time, the only solution from the above was a regular interval service which ran for as long [*i.e.* far] as possible."

9.2.1 "3 + 3" Service on the North London Line

Headway Adjustment and Frequency Reallocation

The new tactical plan resulting from the above analysis was referred to as the "3 + 3" service because it integrated the NLL and WLL into a single trunk-and-branch service for the AM and PM Peak periods. It is effectively an even 20 minute headway (3tph) service between Stratford and Richmond superimposed with an identically spaced service between Stratford and Clapham Junction. The two services are offset by 10 minutes, yielding an even 10 minute headway (6tph) trunk service between Stratford and Willesden Junction.

Figure 9-5 illustrates the "3 + 3" service pattern schematically. Figure 9-6 uses a time-distance plot to show the corresponding published AM Peak timetable for the NLL. Table 9-2 summarizes the changes in the evenness and frequency [6] of service for each segment of the NLL and WLL.[7]

	Spring 2008 Service (tph)			"3 + 3" Service (tph)		
Segment Between	Core (even)	Add'l (uneven)	Total	Core (even)	Add'l (uneven)	Total
North London Line						
SRA ⇔ CMD	4	1-2	5-6 uneven	6		6 even
CMD ⇔ WIJ	4	0-1	4-5 uneven	6		6 even
WIJ ⇔ RMD	4		4 even	3	0-1	3-4 uneven
West London Line						
WIJ ⇔ CLJ	2	0-1	2-3 uneven	3		3 even

Table 9-2: NLL and WLL service under the Spring 2008 and the "3+3" tactical plans

The only part of the WLL and NLL to lose any service frequency under this reallocation was the western end of the NLL between Willesden Junction and Richmond. One additional shuttle trip over this segment was added for the entire peak period, but overall the headway increased from 15 minutes to 20 minutes (4tph to 3tph). Not surprisingly, this was the most contentious aspect of this plan. However, the OD matrix estimated in Chapter 5 and summarized in Section 9.1.2 shows that only 25% of the NLL passengers using the NLL in the AM Peak used this segment of the line. The "3 + 3" tactical plan thus reallocated service such that more passengers gained service than lost it. In doing this, it was able to establish a service pattern with even headways throughout the network that was easier for customers and operators to understand, remember, and use.

Running Time Adjustment

In addition to the headway and frequency changes, timetable running times were also adjusted. Bratton (2009) described conventional practice in National Rail timetable develop-

[6] For even service, headway (in minutes) is $\frac{60}{tph}$.

[7] The segments used here are slightly different than the segments used elsewhere in this thesis. These segments are defined on the basis of service frequencies, whereas the other segments are defined based on the location of interchange stations.

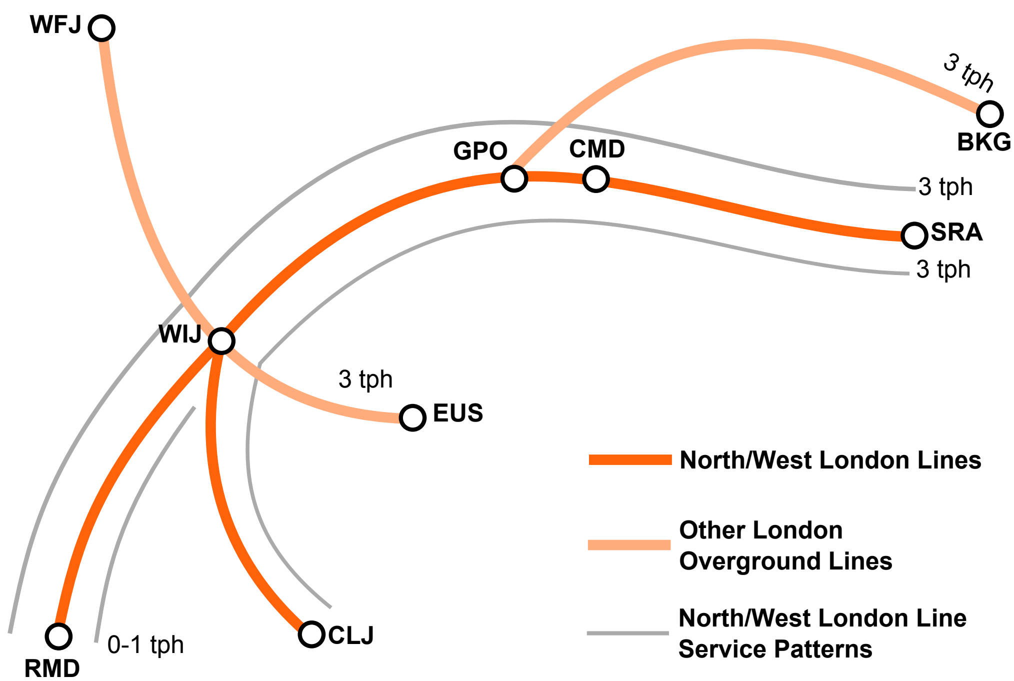
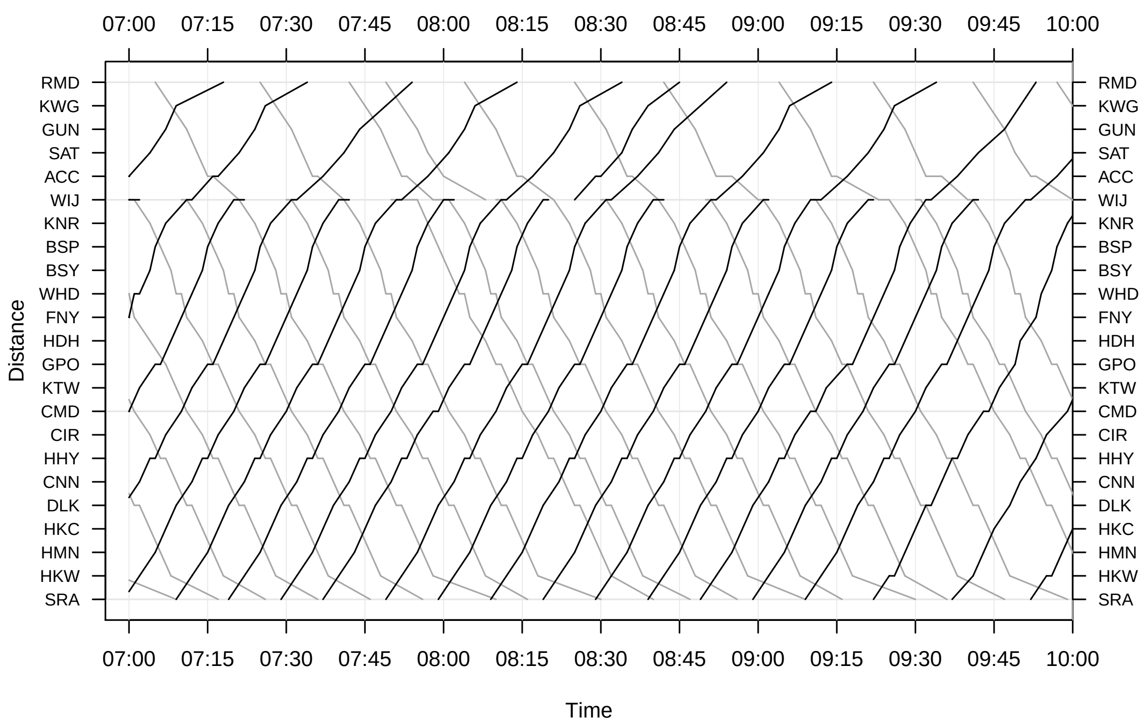


Figure 9-5: London Overground "3 + 3" service patterns and frequencies



Services starting or ending at Willesden Junction (WIJ) on this diagram start from or continue on to Clapham Junction over the West London Line

Figure 9-6: North London Line "3 + 3" timetable ment being to add running time between the penultimate and final stations on a line. This increases PPM scores (*i.e.* on-time performance) and increases the chance that a train will be in place for the beginning of its next trip, but has little effect on the fidelity of the timetable to actual operating conditions on most of the line.

In this case, Bratton (2009) stated that his goal in designing the new timetable was to bring it into alignment with actual operations so that customers traveling anywhere on the line would not be "late" at their destinations. He was motivated to approach timetable design this way by the concepts exposed in the development of the Oyster-based EJT measure. Specifically, that lateness can be defined at the level of passenger journeys or OD flows rather than terminal-to-terminal train trips.

In the new timetable, the running time between Stratford and Camden Road was shortened by 1-2 minutes on the basis that evening the headways would drastically reduce dwell times. The running time between Camden Road and Richmond was lengthened by 3-4 minutes to account for discrepancies found in the study by ACT (2008). Both changes were effected through 1 minute adjustments en-route rather than in large blocks of time at the end of the line or segment of the line. The total running time between Stratford and Richmond was lengthened by approximately 3 minutes on average over the AM Peak period.

Costs

Bratton (2010) noted that the changes in service described here were for all intents and purposes cost-neutral. The reallocation of service frequencies and adjustments in running times were such that the "3 + 3" service could be operated at approximately the same costs as the existing timetable. No new rolling stock was required and only two additional crew members – conductors at an annual salary of £23 thousand each – were needed to fully staff the new timetable.

9.3 Evaluation

"3 + 3" service went into effect on the NLL and WLL on Monday, 20 April, 2009. This section evaluates the outcomes of this tactical planning intervention, primarily through longitudinal analysis of measures of service delivery and service quality. These measures are taken for a period directly after the implementation, and compared to two periods before the implementation – one directly before and one a year earlier. The goal of this evaluation is to assess, to the degree possible, the causal effects on passengers and on the operation of changing the tactical plan. However, despite the large amount of data available, any number of uncontrolled factors may confound this analysis. Final conclusions on the effects of the "3 + 3" plan will be a matter of judgment on the part of the reader.

Because this evaluation is of a major *change* to the timetable, it is important to evaluate service quality in absolute as well as relative terms. Consequently, total observed journey time (OJT) will be analyzed along with EJT. All else being equal, any change in OJT should be reflected in an equal change in EJT. However, any adjustment to the timetable will break this link.

Service delivery is evaluated primarily in terms of PPM (*i.e.* on-time performance) and dwell times, the latter of which was analyzed in a follow-up study by ACT (2009) (the same consultancy whose analysis contributed to the formulation of the "3 + 3" plan). Interviews with London Overground management are also considered.

This evaluation is focused on the NLL, and uses the GOB as a control since there were no substantial changes to the GOB timetable over this period. The GOB is not a perfect control in that it has different service and demand characteristics from the NLL, but it should be subject to similar universal influences such as weather, overall economic conditions, etc. The NLL is evaluated as a whole (including passengers in both directions) and also for the core market passengers traveling between Stratford and Camden Road in the westbound direction (*i.e.* towards Camden Road). This "NLL Core" market is analyzed separately because it is the only section of the line for which there was only a change in the evenness of headways in the peak hours and not a change in the overall frequency of service.

Unfortunately, Overground services were moved to a different platform at Stratford station on the same day that "3 + 3" service started. This new platform was substantially further from the ticket gatelines than was the original platform (which was close to the station entrance). The additional walking time was measured by this author to be at least one minute at a brisk pace. This has the potential to seriously confound a longitudinal analysis of journey times measured by the Oyster system, so all journeys to or from Stratford are excluded. This is far from ideal, since Stratford is one of the most heavily-used stations on the NLL. The resulting loss of data is just over 50% for the core market and just under 25% for the entire NLL. While these numbers are large, they do leave a substantial amount of data remaining for analysis, so the losses are considered acceptable.

9.3.1 Evaluation Data

The following three study periods are analyzed to evaluate the effects of introducing "3 + 3" service. They are determined in part by data availability.

$After2009$**:** Weekdays from 20 April through 15 May and 1 June through 5 June, 2009, inclusive. This is 5 out of the first 7 weeks directly following the introduction of "3 + 3" service.

$Before2009$**:** Weekdays from 2 March through 13 March, 2009. This is a period of two weeks shortly before the introduction of "3 + 3" service.

$Spring2008$**:** Weekdays from 21 April through 16 May and 2 June through 6 June, 2008, inclusive. These are the weeks in 2008 corresponding to the weeks in the $After2009$ period.

Complete PPM and timetable data were available for these study periods. Observed and excess journey times (*i.e.* OJT and EJT) are measured from Oyster journey data, the volumes of which[8] (after exclusion of journeys to or from Stratford) are shown in Table 9-3. At first glance the numbers in this table indicate increasing weekly ridership. However, this interpretation does not account for changes in the Oyster penetration rate among Overground riders. An increasing penetration rate would result in increasing volumes of Oyster data

[8] "Passengers" are unique passengers as indicated by unique Oyster card IDs

despite static volumes of overall ridership. This evaluation does not explicitly analyze total ridership on the Overground network or lines in question.

Study Period	Length (weeks)	GOB Journeys	GOB Pax.	NLL Journeys	NLL Pax.	NLL Core (SRA→CMD) Journeys	NLL Core (SRA→CMD) Pax.
$Spring2008$	5	30,505	6,566	133,030	28,222	17,226	4,050
$Before2009$	2	13,428	4,351	51,862	16,573	7,280	2,487
$After2009$	5	34,408	9,248	136,655	36,738	19,217	5,755

Table 9-3: Volumes of Oyster data in evaluation study periods

Line and segment running time analysis is drawn from the report by ACT (2009).[9] Unfortunately, ACT reported on changes in median rather than mean dwell times. Median values mask the effects of large outliers, which for a nonlinear phenomena such as dwell time are expected to be quite important, so changes in dwell time are not directly analyzed here.

9.3.2 Evaluation Results

Table 9-4 shows PPM, EJT, and OJT results for the three network segments for the three study periods. It shows the difference from $Spring2008$ to $Before2009$ to indicate the changes in performance between the time the tactical planning analysis was done and just before "3 + 3" was implemented. It shows the difference from $Before2009$ to $After2009$ with the hope of isolating the effects of introducing the "3 + 3" service.

The differences in EJT and OJT were tested in a single-sided difference of means t-test. All differences on the NLL and NLL Core were statistically significant at the 1% level. Differences on the GOB between $After2009$ and $Spring2008$ were significant at the 5% level, but between intermediate periods were not significant even at the 10% level.

	GOB			NLL			NLL Core	
Study Period	PPM (%)	EJT	OJT	PPM (%)	EJT	OJT	EJT	OJT
$Spring2008$	95.5	1.27	25.32	85.9	2.77	26.50	2.33	17.97
$Before2009$	96.3	1.21	25.25	79.7	2.29	25.69	1.39	17.42
$After2009$	95.5	1.19	25.14	92.4	1.68	25.51	1.75	17.06
$Bef09 - Spr08$	0.8	-0.06	-0.07	-6.2	-0.48	-0.81	-0.94	-0.55
$Aft09 - Bef09$	-0.8	-0.02	-0.11	12.7	-0.61	-0.18	0.36	-0.36
$Aft09 - Spr08$	0.0	-0.08	-0.18	6.6	-1.09	-0.99	-0.57	-0.91

Table 9-4: PPM and passenger journey time results and comparisons for "3 + 3" implementation

[9] ACT did not use exactly the same study periods for its analysis of train operations data as was used for the analysis of PPM and Oyster journey time data. However, the periods are very similar – two weeks in May 2008, two weeks in March 2009, and two weeks following the "3 + 3" introduction – so the results are considered to be applicable here.

On the GOB, the changes in all three measures were small – PPM fluctuated by 0.8% and returned to its original value of 95.5%, while OJT and EJT decreased by 0.08 minutes (6.3%) and 0.18 minutes (0.7%), respectively, between the initial and final study period. It is not unexpected that OJT and EJT did not vary by the exact same amount. While the changes to the GOB timetable were minor, OJT could be affected by slight shifts in (i) passenger incidence behavior, since EJT is calculated against scheduled waiting time; or (ii) the temporal distribution of ridership over the AM Peak, since running times in the timetable do vary slightly over the AM Peak. This illustrates some of the factors that may confound the longitudinal analysis for the NLL, if only to a small degree.

Interpretation: Changes on the NLL Before the "3 + 3" Introduction

On the NLL as a whole, changes may be observed in all the calculated measures. PPM increased (*i.e.* worsened) between the $Spring2008$ and $Before2009$ study periods by nearly 6 percentage points. ACT (2009) found that over this time average train journey time from Stratford to Richmond increased by about 30 seconds.

OJT and EJT decreased (*i.e.* improved), by 0.81 minutes (3.1%) and 0.48 minutes (17.3%), respectively, over the 9 months between the first two study periods. For the NLL Core passengers, OJT and EJT decreased by 0.55 minutes (3.1%) and 0.94 minutes (40.3%), respectively. It is interesting to note in these cases that the changes in EJT and OJT, measures of relative and absolute service quality, were directionally opposite of the changes in PPM,[10] a measure of service delivery.

It appears that there were substantial improvements to absolute and relative service quality on the NLL as experienced by passengers *before* the implementation of the "3 + 3" service (*i.e.* NLL over the 9 months between $Spring2008$ and $Before2009$). Bratton (2010) attributes this primarily to "higher performing Network Rail infrastructure." The 2008 TfL Investment Programme (Transport for London, 2007*b*) indicates that £56.9 million of infrastructure (*i.e.* track, switch, and signal) upgrades were planned during this period. Much of this investment was in support of capacity on the NLL to an eventual 12tph. It is difficult to separate these capacity upgrades from investments that would improve infrastructure performance at the same level of throughput, but this figure indicates the intensity of the work that was done.

Interpretation: "3 + 3" Effects on the NLL

For the NLL, the comparison between the $Before2009$ and $After2009$ study periods should give the clearest insight into the direct effects of introducing the "3 + 3" service. PPM increased by over 12 percentage points while the average running time from Stratford to Richmond increased by just under 20 seconds. Unfortunately, ACT (2009) did not report on the distributions of running times, so it is impossible to say how much the change in PPM is a result of improved service delivery as compared to the more generous standard set by lengthening the running time in the timetable. During this period, the average train running time from Stratford to Camden Road decreased by 50 seconds.

[10] Directionally in the sense of improvement or worsening, not in terms of the numeric sign of the delta.

OJT decreased by an additional 0.18 minutes (0.7%) for the NLL as a whole and 0.36 minutes (2.1%) for the NLL Core, indicating improved passenger journey times. The changes in EJT, like those in PPM, are more difficult to interpret. For the NLL Core, EJT worsened (*i.e.* increased away from zero) while PPM and OJT both improved. This is not a surprise, as the running time between Stratford and Camden Road was shortened in the new timetable. OJT decreased but the scheduled journey times (SJT) decreased *even more*, so EJT increased. This illustrates one of the disadvantages of timetable-based measures of service quality such as EJT, which is further discussed in the following chapter.

Given the substantial changes to the timetable, the effects of introducing "3 + 3" service may best be judged in terms of absolute service quality. The decreases in OJT suggest that the tactical planning intervention improved the experience of NLL passengers. The changes in OJT are 22.2% and 65.4% of the changes observed on the NLL and NLL Core, respectively, between the $Spring2008$ and $Before2009$ study periods.

Figure 9-7 plots total EJT by scheduled service for westbound passenger journeys on the NLL between Stratford and Willesden Junction in the $After2009$ period. Comparing it to Figure 8-9 (the same plot but for the $Spring2008$ period) shows a more even distribution of EJT across scheduled services during the height of the peak period. For example, in $After2009$ the differences between the services with the highest total EJT (the 08:09 and 08:39 trains from Stratford) and their respective leaders and followers is smaller, even in relative terms, than the same differences for the services with the highest EJT in $Spring2008$ (the 07:52 and 08:22 trains from Stratford).

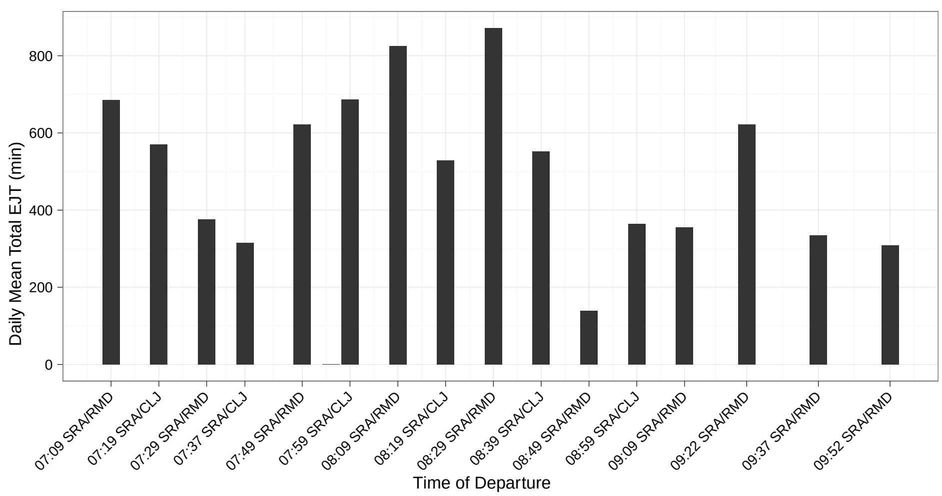


Figure 9-7: Total EJT by scheduled service, westbound, after "3 + 3" implementation

Because the timetable and headways changed between these two study periods, incidence behavior was also examined. Plots similar to Figure 9-4 did not show any noticeable change in overall incidence behavior. Mean scheduled waiting time (which is part of the scheduled journey time against which EJT is measured) decreased by 0.1 minutes (6 seconds) on the NLL Core and increased by approximately the same amount over the entire NLL. These changes appear to be a function primarily of the changes to the timetable rather than changes in incidence behavior (*i.e.* passengers becoming more or less attuned to the timetable as discussed in Section 7.4.3).

Abdul Salique (2010), a TfL contract manager who represents TfL in its relationship with LOROL, was interviewed about the effects of introducing the "3 + 3" service. It should be noted that it is his job to hold LOROL to their contractual responsibilities while considering the experience of the Overground's passengers to whom TfL is ultimately accountable. He noted that "overall it is good for passengers... There has been a lot of good passenger benefit to a greater number of passengers than those that have been disadvantaged" (*i.e.* those traveling west of Willesden Junction). "It has also improved train performance and has made the timetable more robust and easier to recover from... Overall I would say we could move more people during the peaks when we were a bit constrained before."

9.4 Conclusions

The analysis in this chapter indicates the following preliminary conclusions regarding the recent tactical planning intervention and resulting "3 + 3" service on the NLL. Conclusions regarding EJT and the use of automatic data in tactical planning in general are presented in the following chapter.

- The timetable is, overall, more aligned with actual operating conditions (*i.e.* segment and line running times). That said, the "3 + 3" timetable may have been too ambitious in terms of decreasing running time between Stratford and Camden Road.

- The service plan and timetable are, on balance, more aligned with the experience and needs of passengers and operators. Headways have been evened out to correspond with the largely random incidence behavior of NLL passengers. The service patterns and timetable itself are simpler and more uniform than before the change, making them easier for passengers to understand and easier for operators to operate.

- Passenger journey times have improved. Decreases observed in the 9 months leading up to the "3 + 3" implementation are felt by London Overground management to have resulted from substantial infrastructure investment, maintenance, and management efforts. Additional decreases observed directly after the implementation are felt to have resulted from the change to the tactical plan. The latter improvements in passenger journey times are comparable to, if smaller than, the former improvements – 20% overall and 65% on the most heavily used segment from Stratford to Camden Road.

- The change to the "3 + 3" service was approximately cost-neutral. Whatever benefits were reaped by this tactical planning intervention on the NLL came at the cost of only two additional crew members and the analysis and planning effort required to see it through.

The causal relationships in the penultimate conclusion, particularly with respect to the effects of infrastructure investment, have not been rigorously analyzed. If (and only if) they are to be believed, it raises the question of whether an approximately cost-neutral tactical planning effort had comparable but smaller effects on passenger journey times as a substantial investment program and 9 months worth of maintenance and management efforts. Even if this question were answered affirmatively, it is not to say that tactical planning is a *substitute* for capital investment (which comes as the result of strategic planning). Rather, it would illustrate that the potential benefits and cost effectiveness of improved tactical planning using automatic data should not be overlooked in the overall management, investment, and planning of an urban railway.

Chapter 10

Final Remarks

This chapter presents final remarks regarding the research documented in this thesis. Section 10.1 summarizes and draws conclusions about the methodological and applied results of the previous chapters. Section 10.2 presents some recommendations for the use of automatic data by the London Overground and Section 10.3 suggests future research directions.

10.1 Summary and Conclusions

This section presents a summary and conclusions, first for the various analytical methods developed and used in this thesis, and next for the use of automatic data in tactical planning. It also outlines the specific methodological contributions of this thesis.

10.1.1 Analytical Methods

Loadweigh Calibration

Chapter 4 developed and applied a method for calibrating train loadweigh systems to measure passenger loads on trains. It used a simple linear regression model to compare loadweigh measurements with corresponding manual passenger counts so as to estimate the average passenger weight and tare weight of the loadweigh system. It found that the manual passenger counts provided, collected by a pair of surveyors at each station, were generally of insufficient quality to accurately calibrate loadweigh systems. The exception to this finding was at terminal stations where conditions allowed surveyors to count more accurately than at other stations in the network.

Analysis of the limited data set (from terminals only) suggested that the assumptions of an average passenger weight of 80kg, a 95% confidence interval of ±20 passengers (per train), and no tare weight, as recommended by an industry expert (Smale, 2010), are in fact reasonable. The number of paired observations of loadweigh measurements and manual counts taken at terminals was limited to only 49 out of the whole set of 1,253. The regression on this limited set produced results very close to prior findings including the following.

- It estimated the average weight of passengers at 81.4kg, with high statistical significance. This is only 1.8% different from the recommended value of 80kg.

- It estimated an error bound of 10.8kg. This is the upper bound of the standard deviation of random error in passenger loads inferred from loadweigh data. Assuming normally distributed random error, this amounts to a 95% confidence interval of ±21.2 passengers (per train), as compared to the recommended ±20 passengers.

- It estimated a relatively small overall tare weight of 328.6kg (4.0 passengers), but this estimate is not statistically significant.

- When data was segmented by individual train unit, it estimated nearly equal average passenger weights and similarly small and statistically insignificant tare weights *for each unit*. That is, there is no evidence to suggest that the tare weight is other than zero.

In terms of methodology, the linear regression model used in this chapter appears suitable for comparing loadweigh data to manual passenger counts. Because the residuals in these regressions were found to be of approximately constant variance, the method of ordinary least squares is adequate to estimate this model.

OD Matrix Estimation

Chapter 5 developed and applied a method, tailored to the circumstances of the London Overground, to estimate time period level origin-destination (OD) matrices from multiple types of automatic data. It developed an assignment model as required by OD estimation methods which integrate data on OD flows with counts of on-board passenger loads. It applied the assignment model to an OD estimation process using the Information Minimization method of Van-Zuylen and Willumsen (1980) to estimate an OD matrix for the Overground network for the AM Peak period. It validated the resulting estimate against boarding and alighting counts, and compared this validation to a validation of the existing Overground OD matrix estimated using from TfL's regional transportation models and surveys.

The following conclusions were drawn.

- Link flows from loadweigh measurements and/or manual on-board link counts can be combined, through a mathematical estimation process, with aggregate transactional data from the Oyster smartcard ticketing system to estimate time period level OD matrices for the Overground network.

- The overall accuracy of the OD estimate is improved by the addition of automatic entry and/or exit totals from gatelines at stations exclusive to the Overground.

- Of the wide range of available mathematical models and methods for assignment and OD estimation, a relatively simple approach is sufficient to use these data sources to improve the accuracy and currency of OD matrices for the Overground network compared with the existing OD matrix from the London Travel Demand Survey and RailPlan regional model.

- The key outputs of the network assignment model, which does not account for congestion or capacity constraints, are relatively insensitive to most embedded assumptions regarding passenger path choice. Specifically, the choice for most passengers of whether or not to use the Overground network does not change when most of the model's assumptions are violated.

- The Information Minimization method for OD estimation from link flows and a seed matrix is suitable to the problem faced by the Overground. It is simple to implement, is conceptually very similar to the matrix estimation method used by the London Underground, and has the very important feature that its results are not sensitive to overall scaling of the seed matrix when total number of passengers (*i.e.* the sum of the OD matrix) is not fixed.

- The OD estimation process developed here is insensitive to measurement error in the loadweigh data under two conditions. First, that the measurement error is unbiased and uncorrelated with the actual number of passengers, which has been found to be the case in this and other work. Second, that there is a sufficient quantity of loadweigh data (*i.e.* at least eight weeks) over which to estimate average loads.

- The OD estimation method developed here does not treat the seed matrix as a lower bound on the estimated matrix, and some estimated OD flows are, unrealistically, lower than their respective values in the Oyster seed matrix. This is not considered a serious problem for this method in practice because it affected only a very small portion of the OD flows and those it did affect were for the most part relatively small flows to begin with.

Broadly speaking, it is concluded that the methodology developed here would represent an improvement with respect to the Overground's current practices, but that there remains potential for further improvement (as discussed later in this chapter).

Passenger Incidence Behavior

Chapter 6 developed and applied a method to analyze passenger incidence (*i.e.* arrival) behavior at stations using AFC data. The variables by which incidence behavior was analayzed are, for each individual passenger journey, the scheduled headway relevant to that journey and the time between incidence and the next relevant scheduled departure (the "scheduled waiting time"). The method infers these variables given each journey's origin, destination, and time of entry. It uses schedule-based assignment to account for the heterogeneity of services that are likely to be used by different passengers, even at a single location (*e.g.* trunk-and-branch service). It was implemented using open standard timetable formats and free software tools.

The following conclusions were drawn about the methodology itself.

- It can be used to study passenger incidence behavior using large samples of disaggregate journey data from AFC systems such as the Oyster smartcard system. It is able to efficiently process thousands or millions of such data records.

- It can, for each passenger journey, estimate the waiting time and headway (with respect to the timetable), two of the most important quantities for studying passenger incidence, even under quite heterogeneous conditions.

With respect to the London Overground, the following have been concluded from the results of applying the method to relevant Oyster data.

- Passenger incidence behavior is heterogeneous across the Overground network and across times of day, and that the differences are broadly reflective of what has been found to date in the literature. Incidence appears to be substantially more random on the North London Line (NLL) than on the other Overground lines.

- On the lines with timetable-dependent incidence behavior (*i.e.* other than the NLL), passengers reduce their mean scheduled waiting time by over 3 minutes, or up to 30%, during daytime hours compared with random incidence behavior. On the NLL, such reductions are much smaller, in some cases nearly zero, in both relative and absolute terms.

This work has not attempted to rigorously identify the causes of these difference. Hypotheses drawn from existing literature on the subject and knowledge of the Overground network include (i) shorter headways (*i.e.* higher frequencies) and (ii) less reliable service on the NLL compared to other lines.

Service Quality Measurement from AFC Data

Chapter 7 explored the issue of service quality measurement using passenger journey times as measured by AFC systems (*e.g.* Oyster). It developed a method to measure excess journey time (EJT) – the difference between the observed journey time (OJT) and a journey time standard derived from the published timetable. EJT thus defined is a measure of relative service quality that strikes a useful balance between the passenger's and operator's perspectives. It has found lasting application at a number of large urban railways. Actual passenger journey times can now be measured (rather than modeled) directly from automatic data produced by AFC systems such as the Oyster smartcard.

The method developed here includes different models for how passengers set journey time standards depending on whether they are aware of the timetable (and specific departure times) or not, as implied by their incidence behavior. It estimates EJT through comparing the actual arrival time at the destination with the scheduled value as determined by a schedule-based assignment of each journey given its origin, destination, and incidence time. This method was found, despite its dependence on specific train departure and arrival times, to be, in aggregate, unbiased regardless of the randomness of passenger incidence. In this sense it is a unified method in that it can be applied regardless of specific passenger incidence behaviors, whereas most measures of service quality and relative service quality have made the assumption of random incidence. This result was rigorously proven for a single rail line without interchanges, but intuitively should hold for a rail network. This is a very useful result in practice. It allows for the estimation of aggregate EJT from only AFC (*e.g.* Oyster smartcard) data and published timetables in a simple unified manner, regardless of service frequencies or passenger behavior that vary across the network or over time.

EJT for individual passenger journeys on a given service was found to range from negative (*i.e.* early) by up to one headway to positive (*i.e.* late) by substantially more than one headway. A negative EJT for a single journey does not in and of itself represent a problem or inconsistency in the measurement scheme. However, it is difficult to interpret EJT for individual journeys, in part because of the ambiguity with respect to passengers' incidence behavior and implied journey time standards. Consequently, EJT is not a particularly useful measure for analyzing individual passenger journeys.

Aggregate EJT, on the other hand, is a measure of relative service quality with clear meaning. It expresses the average passenger's experience in terms of total journey time compared to what the timetable would imply, for a wide range of passenger incidence behaviors. Individual EJT measurements are easily aggregated both spatially and temporally. When aggregated, EJT is by its very nature passenger-weighted. Different aggregation functions, for example mean or sum, can be used to present different types of passenger orientation. Both such aggregations capture certain information about the passenger's experience that a service delivery measure such as on-time performance (OTP) does not.

Three particular issues were identified with the interpretation of EJT. Firstly, when a substantial proportion of passengers are incident shortly after a scheduled departure time based on an expectation that trains regularly depart late, and their expectations are correct, EJT can be net negative. This could occur despite the fact that passengers may arrive at their destinations on time or late compared to the schedule for the train *they* may have expected to take. Such a situation highlights the need to study passenger incidence behavior, as in Chapter 6, before analyzing EJT. No evidence of this problem was found for EJT on the London Overground network. Such could still occur in practice.

Secondly, changes to the timetable may result in changes in EJT even when no change in service quality has been experienced by passengers. As timetable revision is one of the most common tactical planning activities, this poses a particular problem for the use of EJT in *longitudinal* evaluation of changes in service quality. Lastly, while changes in passenger incidence behavior will not bias EJT results, neither will benefits that passengers capture from such changes (*i.e.* by reducing waiting time) be reflected. This too may confound longitudinal interpretation of EJT results, and underscores the need to study passenger incidence independently of EJT.

It may be the case that EJT and other measures of relative service quality are useful primarily as tools for cross-sectional analysis to inform, but not evaluate, tactical planning changes. Further conclusions regarding the application of EJT to railway tactical planning (and evaluations thereof) are discussed in the following section.

10.1.2 Applications of Automatic Data to Tactical Planning

The case study presented in Chapter 9, which used London Overground EJT results presented in Chapter 8, provided a rich example of the use of automatic data for tactical planning on the North London Line (NLL) of the Overground network. The models, methods, and results of the earlier chapters of this thesis contributed to both the development of the new tactical plan for the NLL and to the evaluation of the implementation of that plan. The effects of this implementation appear to have been positive on balance. This case study thus demonstrates the applicability of automatic data generally, and the data and methods developed in the thesis specifically, for tactical planning of an urban railway.

Various measures of service delivery and service quality, all generated from automatic data sources, contributed to the tactical planning exercise that led to the "3 + 3" service.

In terms of service delivery, dwell times, running times, and on-time performance (*i.e.* PPM) were analyzed using train signaling and control data. They indicated that running times in operation were substantially longer than in the timetable, and that excess dwell time, particularly on certain scheduled services and at certain locations, was one of the main causes of increased running times. Even when tremendous amounts of information are available on actual passenger journeys, traditional measures of service delivery were found still to be quite useful in starting the tactical planning process.

Service quality was analyzed primarily in a relative sense in terms of excess journey time (EJT) as measured using Oyster journey data and the published timetable. An aggregate analysis of total EJT helped direct attention to the experience of passengers traveling westbound on NLL in the AM Peak period. A more disaggregate analysis of EJT identified those scheduled services whose passenger markets suffered the longest journey times relative to the timetable. The analysis of service quality in terms of EJT largely confirmed the analysis of the various service delivery measures, particularly in terms of which scheduled services suffered the worst performance (*i.e.* those with the longest scheduled headways). EJT was found to add an element of the passenger's perspective to the tactical planning process. It can focus tactical planning attention to the segments of passengers who need it most, and can support and enhance analyses that have been initiated from the operator's perspective.

Two analyses of passenger demand also contributed to the tactical planning process. Passenger incidence behavior on the NLL was analyzed using Oyster journey data and the published timetable. It was found to be substantially more random than previously assumed, contributing to the decision to move the NLL to an even headway service. The use of automatic data can provide key insights into passenger behavior, helping to challenge standing assumptions and develop tactical plans better suited to passengers current behaviors.

Also analyzed was the origin-destination matrix of overall AM Peak passenger demand on the Overground network, estimated from aggregate Oyster passenger volumes, automatic gateline entry counts, and manual on-board passenger counts (which can and should be replaced in the future by automatic loadweigh data). This OD matrix indicated that the proposed reallocation of some service frequency away from the western end of the NLL towards the eastern end would benefit more passengers than it would harm. This reflects a common consideration in tactical planning which illustrates the need for a high quality OD matrix in almost any analysis.

The confluence of these analyses contributed to the development, proposal, and implementation of the even headway "3 + 3" service on the NLL and West London Line (WLL) in the AM and PM peak periods. The development of the timetable for this service was also influenced by the key concept inherent in the idea of measuring EJT – that standards can be set and lateness can be measured at the level of individual passenger journeys or OD flows. This led the developers of the timetable to adjust running times throughout the length of the NLL rather than only at the end of the line as was typical on the National Rail network. In this sense, EJT can be a useful tool to help shift tactical planning practices that may be less oriented towards the passenger's perspective than is desired.

Service delivery and quality on the NLL were analyzed longitudinally to evaluate the effects of introducing the "3 + 3" service on passengers and on the operation. Because the timetable changed so drastically in the "3 + 3" implementation, an additional measure of absolute service quality was included in the evaluation. Observed journey time (OJT) was estimated using only Oyster journey data. This and other measures were analyzed before and after the introduction of "3 + 3" service. PPM increased substantially and OJT decreased (*i.e.* they both improved). EJT decreased by substantially more than OJT for the line as a whole and in fact increased for the core portion of the line, which was the portion towards which the "3 + 3" service was targeted.

These discrepancies were found to be because the "3 + 3" timetable had lengthened running times over the whole line and shortened them over the core portion. This highlights the relative nature of EJT, illustrating its value as a relative rather than absolute measure. EJT provides good information about how the passenger experience compares to the timetable, but not necessarily a clear picture of how it has changed in an absolute sense. It is thus similar to on-time performance, but measured for and weighted by individual passenger journeys.

Consequently, it is concluded that measures of relative service quality are not sufficient for evaluating the effects of tactical planning changes, particularly those that modify the timetable. In general it may also be necessary to measure absolute service quality. That said, relative service quality measures such as EJT can be useful for the type of cross-sectional tactical planning analysis that was described in this case study. They can help identify and prioritize problems and suggest solutions, but are not always appropriate for judging the full effects of those solutions.

10.1.3 Methodological Contributions

The research in this thesis, while for the most part applied in nature, has produced two methodological innovations. These were discussed above, but are highlighted here for their contributions to the literature. These new methodologies, primarily useful for applying automatic data to railway tactical planning, include the following.

- Estimation of key variables for studying passenger incidence behavior from AFC data under heterogeneous service patterns. Previous studies of passenger incidence all selected places and times of observation (be it with manual or automatic data) to avoid ambiguity with respect to the services to be used by waiting passengers (*e.g.* excluding trunk stations on a line with multiple branches). They trivialized the measurement of key variables by selecting stations served by only a single service pattern and, in some cases, with a constant scheduled headway. The contribution of the method developed and tested in Chapter 6 (*i.e.* Algorithm 6.1) is that it allows the study of passenger incidence across a network using AFC data without these limitations.

- Unified estimation of aggregate EJT from AFC data under heterogeneous service patterns and passenger incidence behaviors. Existing methods for estimating EJT make the assumption that passengers are randomly incident and thus should expect to wait half the scheduled headway. These methods do not apply to passengers whose incidence is non-random and whose journey time standards depend on specific scheduled train departure and arrival times. They also pose an implementation challenge under heterogeneous service patterns where different passengers at the same location face different headways in practice (*e.g.* at trunk stations on a line with multiple branches).

The contribution of the method developed in Chapter 7 is that it makes no such assumption about passenger incidence behaviors at a given location or for a given service but is nevertheless unbiased. It makes assumptions about how passengers set their journey time standards *given* their incidence behavior (which are consistent with assumptions in other methods for randomly incident passengers). However, it was proven that the actual type of incidence behavior is irrelevant for measuring aggregate EJT with data from AFC systems such as Oyster.

Both of these methods use schedule-based assignment which was implemented for this thesis using open standard format public timetables and free and open source software tools. These methods should be of value in future research as described in the following section.

10.2 Recommendations for Data Collection on the London Overground Network

The general recommendation stemming from this thesis is that the London Overground should move as much as possible towards the use of automatic data to provide key tactical planning inputs and away from dependence on manual passenger counts and surveys. Specifically, it should adopt the strategy proposed in Section 3.2 and summarized in Table 3-1 to both reduce the cost and improve the timeliness of key inputs such as train level passenger loadings and time period level OD matrices. The key elements of this strategy are the following.

Train loads on individual scheduled services should be estimated directly from loadweigh data. Resources permitting, additional calibration should be conducted to estimate the average passenger weight and tare weight (per train unit or fleet wide). An average passenger weight of 80kg and a zero tare weight may be used in this estimation if no further calibration is conducted.

Average loads (on a given service at a given location) should be estimated from multiple loadweigh measurements in a given analysis period. The number of samples needed for a given level of accuracy should be determined from further analysis of larger samples of loadweigh data than were available for this thesis. It is likely that somewhere between one and two weeks worth of data will be sufficient to estimate average load to within ±10 passengers at a 95% confidence level.

Origin-destination (OD) matrices should, as described in Chapter 5, be estimated at the time period (*e.g.* AM Peak) level from the combination of different types of automatic data including the following.

- Seed matrices estimated from aggregate Oyster journey data. Each seed matrix should be estimated by averaging daily Oyster data from a representative sample of days (*e.g.* ten weekdays) in the period of interest (*e.g.* a fiscal quarter).
- Aggregate link flows, estimated from loadweigh data, indicating the daily average number of passengers riding Overground services in each direction between each

pair of consecutive stations in the given time period. These aggregate link flows should be estimated as the sum of link flows on individual scheduled services, as described above. When loadweigh data are used in estimating OD matrices, as many data as possible should be averaged to overcome the random measurement error introduced by loadweigh systems.

- Aggregate station entry and/or exit totals, estimated from station gatelines, at stations that are fully gated and served exclusively by the Overground. Like Oyster seed matrices, these may be estimated from a representative sample of days in the period of interest.

These data should be integrated using the assignment model developed in Chapter 5, which depends on RODS the network representation used in the London Underground's OD estimation process. They should be reconciled (to estimate a complete OD matrix) using the Information Minimization method.

Station boardings and alightings (at the time period level) should be estimated by assigning the estimated OD matrix to the Overground network (using the assignment model developed for OD estimation).

Total daily trips on the Overground network (at the time period level) should be estimated as the sum of the estimated OD matrix (for that time period).

This strategy should be extended as the Overground network is expanded. For example, with the opening of the East London Line (ELL), loadweigh data should be obtained from trains serving the ELL and Oyster and gateline data should be obtained for ELL stations. The RODS network representation should be extended to cover the ELL to support OD matrix estimation (for the Underground and the Overground).

10.3 Future Research

The analysis and results of this thesis suggest a number of directions for future research.

Loadweigh Calibration

Additional research should be conducted into the nature and magnitude of the various sources of error associated with passenger loads inferred from loadweigh data. The weakness of the analysis in this thesis stems primarily from the low quality of the manual counts against which loadweigh data were compared. To remedy this and other issues, the following are recommended.

- More calibration-quality data should be gathered (be it platform-based surveyors at terminals or from on-board surveyors en route) and paired with corresponding loadweigh data. The regression models from Chapter 4 should be used to re-estimate the calibration parameters and the error bound.

- In addition, additional calibration-quality data should be used to explore variability in the parameter estimates and error bound across different individual units of rolling stock.

- As identified by Nielsen et al. (2008*a*), additional analysis should be conducted at different times of year to assess the systematic variation in average passenger weights correlated with seasons and weather. It is possible that such variation could be ignored in practice, but this question should be explored.

Loadweigh calibration results will likely vary across different types of rolling stock. Different railways (including the London Overground) will incur different costs in the loadweigh calibration process, depending on the required or desired accuracy of the results. Best practices should be developed for railways to make the appropriate cost/accuracy tradeoff for their respective uses of loadweigh data.

OD Matrix Estimation

One specific aspect of the OD estimation methodology used in Chapter 5 of this thesis merits further research. The constraint of the seed matrix as a lower bound on the final OD estimate should be added to the Information Minimization formulation. It is trivial to add this constraint to the formulation, but it may make the model much more difficult if not impossible to solve efficiently. It is possible that a lagrangean analysis similar to that developed by Van-Zuylen and Willumsen (1980) would yield an efficient algorithm as it has for the existing formulation.

Other OD estimation methods, for example those based on generalized least squares, may also be applicable to this type of problem. A benefit of these methods is that they allow for the estimation to take into account the varying statistical quality of the inputs rather than treating all measured link flows as deterministic constraints. These methods were not used here because they appeared to be inapplicable when the seed matrix is estimated from only a fraction of journeys, as is the case with Oyster data. The possibility of augmenting these methods to overcome this issue should be explored.

The assignment model and OD estimation methodology used here both assume no congestion effects or capacity constraints in the network. This assumption was made primarily as an engineering simplification, but was justified in part by the structure of the Overground network. The London Underground and National Rail are known, by passengers and managers alike, to face serious congestion problems. Consequently, OD estimation over the entire network of railways in London using similar data would need to account for congestion and capacity. A host of models and methods, some discussed in Chapter 5, have been developed (by other authors) to treat these issues. They should be applied and tested on London's networks, both for the sake of exploring those models and methods and in support of the planning and management of London's railways.

Passenger Incidence Behavior

The method developed Chapter 6, which builds heavily on some of the basic concepts of schedule-based assignment, should be used to support further study of passenger incidence behavior. The work of Bowman and Turnquist (1981) has been influential in shaping the understanding of the relationships between headway, reliability, passenger behavior, and waiting time. Their work should be updated and extended using the method of this chapter to cheaply and easily gather large samples of passenger data across heterogeneous networks. Their work also depends on measurements of service reliability, which should be gathered from automatic vehicle tracking systems.

With such rich data available at such low cost, the study of passenger incidence should be disaggregated to understand the range and consistency of behavior of individual passengers, for example as by Csikos and Currie (2008). Moreover, longitudinal analysis should be conducted to understand the aggregate and disaggregate behavioral responses to changes in key service variables such as headway and reliability.

The London Overground network represents an ideal opportunity to conduct such a study – its passengers can clearly be studied via Oyster data, and its trains are tracked by a computerized signaling system. Once the East London Line opens, the network will have headways ranging from 5 to 30 minutes during most hours of the day. As the network grows and headways and reliability levels change, the evolution of passenger incidence behavior should be studied over time to understand passengers' reactions to such changes.

Nearly three decades have passed since the work of Bowman and Turnquist (1981). In that time, many strides have been made towards informing passengers in real time about the status of public transport services. Such information is now often distributed via in-station signs and announcements as well as over the internet to passengers' computers and, more importantly, mobile devices outside of stations. It is crucial to advance the understanding of passenger incidence to include the effects of real-time information. This requires careful thinking and research designs, but should be able to take advantage of the methodology developed here.

Service Quality Measurement from AFC Data

The unified estimator developed in Chapter 7 for mean EJT with respect to the published timetable should be further analyzed. While it was found to be unbiased for journeys on a single rail under a range of circumstance, the following additional areas of analysis were identified.

- *Interchange journeys* – the estimator should be rigorously analyzed for journeys on a railway network with interchanges.

- *Reliability* – estimators for other higher-order statistics of the EJT distribution, which indicate reliability rather than average performance levels, should be proposed and analyzed. One difficulty in this analysis is that, even with all journey time standards set based on the published timetable, the distribution of EJT depends on how passengers are incident and set their standards whereas mean EJT does not.

The research by this and other authors on using data from AFC systems (*e.g.* Oyster) to measure absolute and relative service quality has only scratched the surface of understanding the complete passenger experience. For example, the research by Chan (2007), Wilson et al. (2008), and Uniman (2009) have developed measures of reliability based on assumptions about how *randomly incident* passengers set standards given their previous travel experiences. This work analyzed EJT (*i.e.* mean performance, not reliability) using standards set based on the timetable for a range of passenger incidence behaviors; it was found however to be confounded by certain types of experience-based incidence behavior. Hopefully, these different threads can be synthesized to develop measures of relative and absolute service quality, in terms of mean performance as well as reliability, for all types of passengers with any type of incidence behavior or journey time standards.

Further research should be conducted on the application of the results of this thesis and related work towards the management and planning of urban railways. For example, while the focus of this thesis was on tactical planning, there may be opportunities to apply service quality measurements towards performance and contract management. These tasks have different requirements from tactical planning in that they are more highly structured activities, often with financial stakes. While EJT may not be appropriate for such uses, primarily because of its sensitivity to changes in the timetable and to certain exceptional passenger incidence behaviors, other AFC-based measures of service quality, such as total journey time or reliability buffer time, may be applicable.

Appendix A

London Overground Station Information and Abbreviations

Table A-1 shows information for each station on the London Overground network, including presence of ticket gatelines and availability of interchanges to the London Underground (LU) and National Rail (NR) networks. NLC is the National Location Code. Lines and segments are defined in Appendix B.

Table A-1: London Overground stations

						Interchange	
Station Name	Code	NLC	Line	Segment	Gated	LU	NR
Acton Central	ACC	1404	NLL	NLLW	Yes	No	No
Blackhorse Road	BHO	522	GOB	GOB	Yes	Yes	No
Barking	BKG	514	GOB	GOB	Yes	Yes	Yes
Bushey	BSH	1395	WAT	WATN	No	No	Yes
Brondesbury Park	BSP	1438	NLL	NLLC	No	No	No
Brondesbury	BSY	1437	NLL	NLLC	Yes	No	No
Caledonian Road & Barnsbury	CIR	1439	NLL	NLLE	No	No	No
Clapham Junction	CLJ	5595	WLL	WLL	Yes	No	Yes
Camden Road	CMD	1440	NLL	NLLE	Yes	No	No
Canonbury	CNN	1441	NLL	NLLE	Yes	No	No
Carpenders Park	CPK	1442	WAT	WATN	No	No	No
Crouch Hill	CRH	7406	GOB	GOB	No	No	No
Dalston Kingsland	DLK	1429	NLL	NLLE	Yes	No	No
London Euston	EUS	1444	WAT	WATS	Yes	No	Yes
Finchley Road & Frognal	FNY	1445	NLL	NLLC	No	No	No
Gospel Oak	GPO	1409	INT	INT	Yes	No	No
Gunnersbury	GUN	591	NLL	NLLW	Yes	Yes	No
Hampstead Heath	HDH	1413	NLL	NLLC	Yes	No	No
Headstone Lane	HDL	1434	WAT	WATN	No	No	No
Harlesden	HDN	596	WAT	WATC	Yes	Yes	No
Highbury & Islington	HHY	603	NLL	NLLE	Yes	Yes	Yes
Hackney Central	HKC	6977	NLL	NLLE	Yes	No	No

Hackney Wick	HKW	6978	NLL	NLLE	No	No	No
Homerton	HMN	6979	NLL	NLLE	Yes	No	No
Harrow & Wealdstone	HRW	597	WAT	WATC	Yes	Yes	Yes
Harringay Grn Lns	HRY	7401	GOB	GOB	No	No	No
Hatch End	HTE	1398	WAT	WATN	No	No	No
Kilburn High Road	KBN	1415	WAT	WATS	No	No	No
Kensal Green	KNL	617	WAT	WATC	Yes	Yes	No
Kensal Rise	KNR	1448	NLL	NLLC	No	No	No
Kenton	KNT	620	WAT	WATC	Yes	Yes	No
Kensington Olympia	KPA	3092	WLL	WLL	No	Yes	Yes
Kentish Town West	KTW	1449	NLL	NLLE	No	No	No
Kew Gardens	KWG	621	NLL	NLLW	Yes	Yes	No
Leyton Midland Road	LEM	7402	GOB	GOB	No	No	No
Leytonstone H Rd	LER	7403	GOB	GOB	No	No	No
North Wembley	NWB	659	WAT	WATC	Yes	Yes	No
Queens Park London	QPW	680	WAT	WATC	Yes	Yes	No
Richmond	RMD	686	NLL	NLLW	Yes	Yes	Yes
South Acton	SAT	1452	NLL	NLLW	No	No	No
Stonebridge Park	SBP	717	WAT	WATC	Yes	Yes	No
South Hampstead	SOH	1451	WAT	WATS	No	No	No
South Kenton	SOK	709	WAT	WATC	No	Yes	No
Shepherds Bush	SPB	9587	WLL	WLL	Yes	No	Yes
Stratford	SRA	719	NLL	NLLE	Yes	Yes	Yes
South Tottenham	STO	7404	GOB	GOB	No	No	No
Upper Holloway	UHL	1524	GOB	GOB	No	No	No
West Brompton	WBP	755	WLL	WLL	Yes	Yes	Yes
Watford High Street	WFH	1455	WAT	WATN	Yes	No	No
Watford Junction	WFJ	1402	WAT	WATN	Yes	No	Yes
Woodgrange Park	WGR	7467	GOB	GOB	No	No	No
West Hampstead	WHD	1421	NLL	NLLC	Yes	No	No
Willesden Junction	WIJ	766	INT	INT	Yes	Yes	No
Wembley Central	WMB	751	WAT	WATC	Yes	Yes	Yes
Walthamstow Queens Rd	WMW	7407	GOB	GOB	No	No	No
Wanstead Park	WNP	7408	GOB	GOB	No	No	No

Appendix B

London Overground Line and Segment Abbreviations

The existing London Overground network can be divided up into lines and line segments as described in Table B-1. Note that the two interchange stations that lie on multiple lines, Gospel Oak and Willesden Junction, are considered separately. There are other stations on the network with interchanges to other networks (*i.e.* National Rail and London Underground) but these are the only two with interchanges between different London Overground services.

Line	Segment	Description
NLL		North London Line, Stratford ⇔ Richmond
	NLLE	NLL East, Stratford ⇔ Kentish Town West
	NLLC	NLL Central, Hampstead Heath ⇔ Kensal Rise
	NLLW	NLL West, Acton Central ⇔ Richmond
WLL		West London Line, Clapham Junction ⇔ Willesden Junction
WAT		Watford DC Line, Watford Junction ⇔ Euston
	WATN	WAT North, Watford Junction ⇔ Headstone Lane
	WATC	WAT Central, Harrow & Wealdstone ⇔ Queens Park
	WATS	WAT South, Kilburn High Road ⇔ Euston
GOB		Gospel Oak to Barking Line, Gospel Oak ⇔ Barking
INT		London Overground interchange stations (Gospel Oak, Willesden Jn)

Table B-1: London Overground lines and line segments

Appendix C

Schematic Map of TfL Rail Services

The image in Figure C-1 on the following page shows a schematic map of TfL rail services, including the London Underground, London Overground (with the new East London Line), and the Docklands Light Railway. The map also illustrates the 9 fare zones from which rail fares in London are determined.

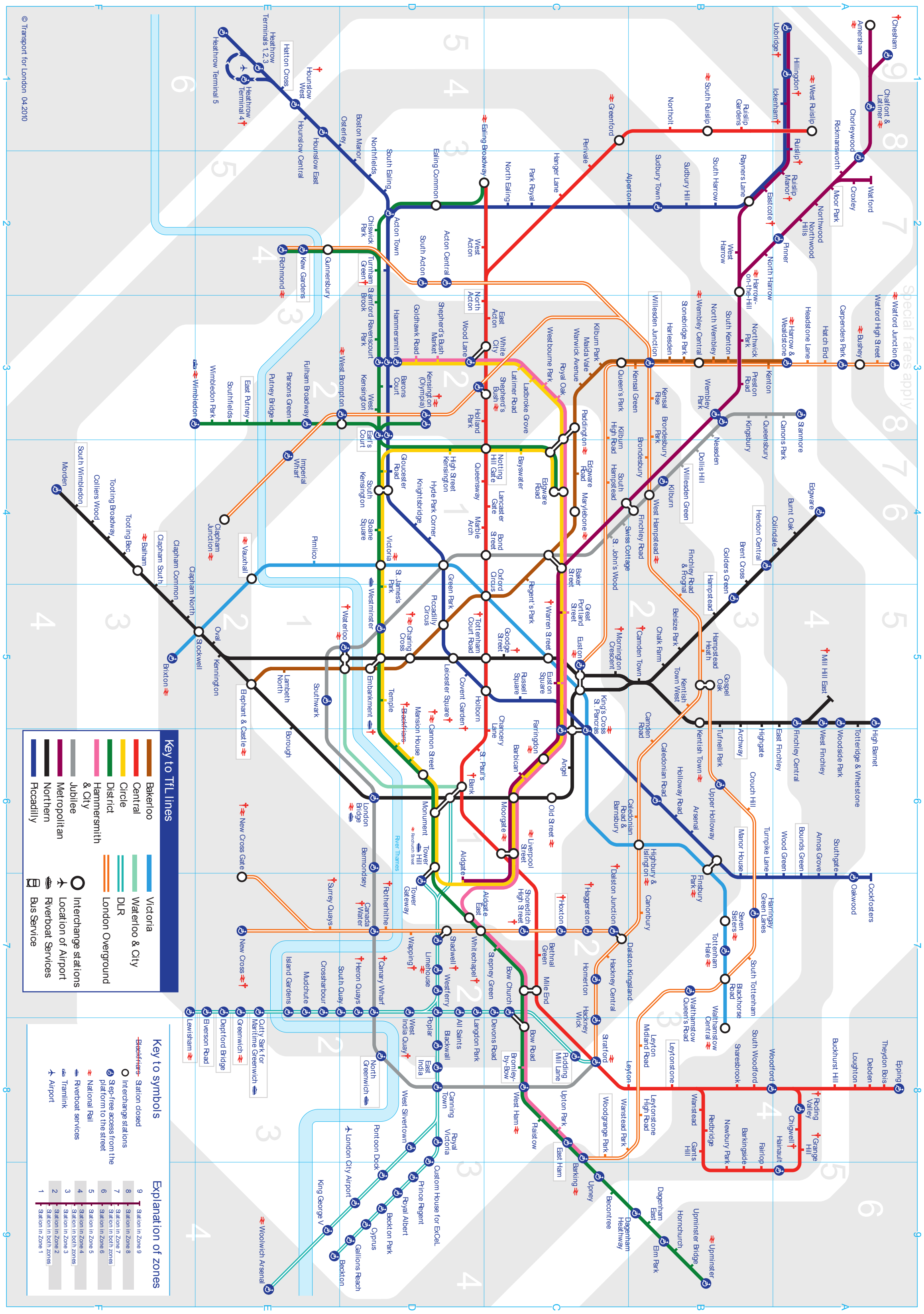


Figure C-1: Schematic map of TfL rail services

Source: Transport for London (2010*d*)

Appendix D

Assignment Model Algorithm for Operator Aggregation

This appendix provides additional detail on implementing the operator clamping aggregation described in Section 5.3.1. The most subtle part of this computation is that of finding the inner (clamped) OD flow for a given path for a given outer (unclamped) OD flow. The complex part of this computation is illustrated the below excerpt from the *ODNet* program developed to implement the assignment model of Chapter 5.

The excerpt shows the declaration of a `Clamp` class (in the Java programming language); an object of this class would be created for each possible path for each outer OD flow. After creating this object, the links in the given path would be examined in order. For each boarding link, the `board` method would be called; for each alighting link, the `alight` method. At the end of such a process, the `ods` member of the `Clamp` object would contain a list of all of the inner clamp probabilities – the fraction of the outer OD flow assigned to each inner OD flow.

```
public class Clamp {

  String opCode;
  HashMap<String,Double> ods = new HashMap<String,Double>();
  AlightLink prevAlight = null;
  List<ClampEntry> clamps = new ArrayList<ClampEntry>();
  double currShare = 0;

  // Start with the "operator code" of the
  // operator for which the clamp is being computed
  public Clamp(String opCode) {
    this.opCode = opCode;
  }

  public void board(BoardLink bl) {

    // the service link which the boarding link connects to
    ServiceLink sl = bl.getServiceLink();
```

```
  // with interavailability, the given operator may only provide some
  // share of the total frequency in this combined service link
  double freq = sl.getOperatorFrequency(opCode);

  double share = 0;
  if(freq != 0) {
    share = freq/sl.getFrequency();
  }

  if(share < currShare) {
    clampTo(prevAlight, share);
  }
  else if(share > currShare) {
    ods.put(bl.getStationID(), share);
    currShare = share;
  }
}

public void alight(AlightLink al) {
  prevAlight = al;
}

public void finish() {
  if(prevAlight != null) {
    clampTo(prevAlight, 0);
  }
}

private void clampTo(AlightLink al, double share) {
  if(share < currShare) {
    double throughShare = share/currShare;

    if(throughShare < epsilon) {
      throughShare = 0;
    }

    double alightShare = (1 - throughShare);

    Set<String> os = new HashSet<String>(ods.keySet());
    for(String o : os) {
      String d = al.getStationID();
      double oShare = ods.get(o);

      ClampEntry ce = new ClampEntry(o,d,oShare * alightShare);
      clamps.add(ce);
```

```
        double newShare = throughShare * oShare;
        if(newShare < epsilon) {
          ods.remove(o);
        } else {
          ods.put(o, newShare);
        }

        currShare = share;
      }
    }
  }

  private class ClampEntry {
    String o;
    String d;
    double share;

    public ClampEntry(String o, String d, double share) {
      super();
      this.o = o;
      this.d = d;
      this.share = share;
    }
  }
}
```

Appendix E

Terms and Abbreviations

This appendix lists and defines a number of the terms and abbreviations used in this thesis. It does not include abbreviations for specific London Overground stations, lines, or line segments; these are described in Appendices A and B.

AFC: Automatic Fare Collection.

DLR: Docklands Light Railway.

EJT: Excess Journey Time; the difference between a passenger's OJT and SJT.

ELLX: East London Line eXtension; the project to rebuild and extend the old East London Line as part of the London Overground network.

GLA: Greater London Authority; the governing body for the greater London area.

Headway: The time between two successive public transport services.

Interavailable: Refers to services with identical stopping patterns on a given corridor.

JTM: Journey Time Metric; the name of the scheme by which the London Underground measures EJT.

LO: London Overground.

LOROL: London Overground Rail Operations, Ltd; the private concessionaire with operating responsibility for London Overground services.

LU: London Underground.

MIT: Massachusetts Institute of Technology.

NLRIP: North London Railway Infrastructure Project; an investment project to improve capacity and reliability on the London Overground network.

NR: National Rail.

OD: Origin-destination.

OJT: Observed Journey Time; the observed duration of a passenger journey, for example as measured by AFC data.

OLS: Ordinary Least Squares; a method by which to estimate mathematical models such as linear regressions.

OSI: Out-of-Station Interchange; a no-cost interchange on the London railway network that involves exiting and re-entering the system.

OTP: On-Time Performance; the fraction of trains arriving some timepoint within some threshold from the scheduled time.

OXNR: Oyster eXtension to National Rail; the project to extend the Oyster system to National Rail network in London.

Oyster: The smartcard ticketing system for London's public transport network.

Pax: Passengers.

PAYG: Pay as You Go; the fare class of passengers who pay for individual journeys using a stored value on their Oyster card.

PPM: Public Performance Measure; the name of the scheme by which OTP is measured on the UK National Rail network.

RODS: Rolling Origin-Destination Survey; the process by which OD matrices are estimated for the London Underground.

SJT: Scheduled Journey Time; the scheduled duration of a passenger journey as derived from published timetables.

TfL: Transport for London; the municipal transport authority for the London.

Timepoint: A location at which public transport service arrivals, passings, or departures are timed.

TOC: Train Operating Company; a private concessionaire, such as LOROL, with operating responsibility for a National Rail franchise.

Bibliography

Abkowitz, M., Slavin, H., Waksman, R., Englisher, L. and Wilson, N. H. M. (1978), Transit service reliability, Technical Report UMTA-MA-06-0049-78-1, US DOT Transportation Systems Center, Cambridge, MA.

Abrahamsson, T. (1998), Estimation of origin-destination matrices using traffic counts – a literature survey, Technical report, International Institute for Applied Systems Analaysis, Laxenburg, Austria.

ACT (2008), North London Line performance review for London Overground Rail Operations Limited (LOROL). Consultant's unpublished report.

ACT (2009), North London Line performance review 2009 for London Overground Rail Operations Limited (LOROL). Consultant's unpublished report.

AECOM (2006), 'East London Line business case update'.

Ahuja, R. K., Magnanti, T. L. and Orlin, J. B. (1993), *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall.

BART (2010), 'BART Quarterly Performance Report 2010 Q1'. `http://www.bart.gov/docs/QPR_Report_FY2010q1_FINAL.pdf` (Accessed March 25, 2010).

Bates, J., Polak, J., Jones, P. and Cook, A. (2001), 'The valuation of reliability for personal travel', *Transportation Research E* **37**, 191–229.

Bell, M. (1991), 'The estimation of origin-destination matrices by constrained generalised least squares', *Transportation Research Part B: Methodological* **25**(1), 13–22.

Bell, M. G. H. and Schmocker, J. D. (2004), *A Solution to the Transit Assignment Problem*, Operations Research/Computer Science Interface Series, Kluwer Academic Publishers, chapter 14, pp. 263–279.

Ben-Akiva, M. (1987), Methods to combine different data sources and estimate origin-destination matrices, *in* M. Gartner and N. H. M. Wilson, eds, 'Transportation and Traffic Theory: Proceedings of the 10th International Syposium on Transportation and Traffic Theory', Elsevier, New York, pp. 459–481.

Ben-Akiva, M., Bergman, M. J., Daly, A. J. and Ramaswamy, R. (1984), Modelling inter urban route choice behaviour, *in* 'Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, Delft, the Netherlands', VSP Intl Science, pp. 299–330.

Ben-Akiva, M. and Lerman, S. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press.

Ben-Akiva, M., Macke, P. P. and Hsu, P. S. (1985), 'Alternative methods to estimate route-level trip tables and expand on-board surveys', *Transportation Research Record* **1037**, 1–11.

Ben-Akiva, M. and Morikawa, T. (1989), Data fusion methods and their applications to origin-destination trip tables, *in* 'Selected Proceedings of The Fifth World Conference on Transport Research: Transport Policy, Management & Technology, Towards 2001', Vol. IV, Yokohama, Japan, pp. 279–293.

Bertsimas, D. and Tsitsiklis, J. N. (1997), *Introduction to Linear Optimization*, Dynamic Ideas.

Bierlaire, M. (2002), 'The total demand scale: a new measure of quality for static and dynamic origin-destination trip tables', *Transportation Research Part B: Methodological* **36**(9), 837–850.

Bombardier Transportation (2004), Electrostar Loadweigh Data Gathering (Final Results Analysis), Technical report.

Bowman, L. A. and Turnquist, M. A. (1981), 'Service frequency, schedule reliability, and passenger wait times at transit stops', *Transportation Research A* **15**(6), 465–471.

Bratton, O. (2008), Unpublished interviews and electronic communications with Oliver Bratton, Head of Performance and Planning, London Overground Rail Operations LTD.

Bratton, O. (2009), Unpublished interviews and electronic communications with Oliver Bratton, Head of Performance and Planning, London Overground Rail Operations LTD.

Bratton, O. (2010), Unpublished interviews and electronic communications with Oliver Bratton, Deputy Operations Director, London Overground Rail Operations LTD.

Brimbacombe, A. (2010), Unpublished interviews and electronic communications with Adele Brimbacombe, Performance Contracts Manager, London Overground Rail Operations LTD.

Buneman, K. (1984), 'Automated and passenger-based transit performance measures', *Transportation Research Record* **992**, 23–28.

Caliper Corporation (2007), *Travel Demand Modeling with TransCAD 5.0: User's Guide.*

Camilletti, P. (1998), *High Frequency Calculations Explained*, London Buses.

Camilletti, P. (2003), *Low Frequency Calculations Explained*, London Buses.

Carrel, A. (2009), Diagnosis and assessment of operations control interventions: Framework and applications to a high frequency metro line, Master's thesis, Massachusetts Institute of Technology.

Cascetta, E. (1984), 'Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator', *Transportation Research Part B: Methodological* **18**(4-5), 289–299.

Cascetta, E. and Nguyen, S. (1988), 'A unified framework for estimating or updating origin/destination matrices from traffic counts', *Transportation Research Part B: Methodological* **22**(6), 437–455.

Cascetta, E. and Postorino, M. N. (2001), 'Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks', *Transportation Science* **35**(2), 134–147.

Casello, J. M. and Hellinga, B. (2008), 'Impacts of express bus service on passenger demand', *Journal of Public Transportation* **11**(4), 1–23.

Ceder, A. (2007), *Public Transit Planning and Operation: Theory, Modelling, and Practice*, Butterworth-Heinemann, Oxford, England.

Cepeda, M., Comineetti, R. and Florian, M. (2006), 'A frequency-based assignment model for congested transit networks with strict capacity constraints: Characterization and computation of equilibria', *Transportation Research B* **40**, 437–459.

Cham, L. C. (2006), Understanding bus service reliability: A practical framework using avl/apc data, Master's thesis, Massachusetts Institute of Technology.

Chan, J. (2007), Rail OD matrix estimation and journey time reliability metrics using automated fare data, Master's thesis, Massachusetts Institute of Technology.

Cominetti, R. and Correa, J. (2001), 'Common-lines and passenger assignment in congested networks', *Transportation Science* **35**(3), 250–267.

Csikos, D. and Currie, G. (2008), 'Investigating consistency in transit passenger arrivals', *Transportation Research Record* **2042**, 12–19.

Csikos, D. R. and Currie, G. (2007), The impacts of transit reliability on wait time: Insights from automated fare collection system data, *in* '86th Annual Meeting of the Transportation Research Board'.

Darroch, J. N. and Ratcliff, D. (1972), 'Generalized iterative scaling for log-linear models', *The Annals of Mathematical Statistics* **43**(5), 1470–1480.

Florian, M. (2004), *Finding shortest time-dependent paths in Schedule-Based transit networks: a Label Setting algorithm*, Kluwer Academic Publishers, chapter 3, pp. 43–52.

Friedman, R. R. (1976), Statistical models of the mean and standard deviation of passenger wait time in urban bus transit, Master's thesis, Northwestern University.

Froloff, E., Rizzi, M. and Saporito, A. (1989), Theory and practice of bus service management, Technical report, RATP.

Frumin, M., Yuan, Y. and Weston, G. (2010), The analytical relationship of waiting time variance and headway distribution for randomly arriving passengers. Unpublished manuscript.

Furth, P., Hemilly, B., Muller, T. and Strathman, J. (2006), *TCRP Manual 113: Using Archived AVL-APC Data to Improve Transit Performance and Management*, Transportation Research Board, Washington, DC.

Furth, P. and Muller, T. (2006), 'Service reliability and hidden waiting time', *Transportation Research Record* **1955**, 79–87.

Furth, P. and Muller, T. (2007), 'Service reliability and optimal running time schedules', *Transportation Research Record* **2034**, 55–61.

Gaitskell, A. (2008), Unpublished interviews and electronic communications with Andrew Gaitskell, Information Systems Analyst, TfL Fares & Ticketing.

Google (2009), 'General Transit Feed Specification'. http://code.google.com/transit/spec/transit_feed_specification.html (Accessed November 13, 2009).

Government Offices (2010), 'Government Offices — London'. http://www.gos.gov.uk/gol/factgol/London/?a=42496 (Accessed February 4, 2010).

Graphserver (2009), 'Graphserver: The Open-Source Multi-Modal Trip Planner'. http://bmander.github.com/graphserver/.

Greater London Authority (2010*a*), 'About'. http://www.london.gov.uk/about.jsp (Accessed February 4, 2010).

Greater London Authority (2010*b*), 'Greater London Authority - Transport'. http://www.london.gov.uk/gla/transport.jsp (Accessed February 4, 2010).

Greene, W. H. (2007), *Econometric Analysis*, sixth edn, Prentice Hall.

Guo, Z. (2008), Transfers and Path Choice in Urban Public Transport Systems, PhD thesis, Massachusetts Institute of Technology.

Hammerle, M., Haynes, M. and McNeil, S. (2005), 'Use of automatic vehicle location and passenger count data to evaluate bus operations: Experience of the chicago transit authority, illinois', *Transportation Research Record* **1903**, 27–34.

Hao, J. and Kocur, G. (1992), *An implementation of a shortest augmenting path algorithm for the assignment problem*, American Mathematical Society, pp. 453–468.

Henderson, G., Adkins, H. and Kwong, P. (1990), 'Toward a passenger-oriented model of subway performance', *Transportation Research Record* **1266**, 221–228.

Henderson, G., Adkins, H. and Kwong, P. (1991), 'Subway reliability and the odds of getting there on time', *Transportation Research Record* **1297**, 10–13.

Henderson, G., Kwong, P. and Heba, A. (1991), 'Regularity indices for evaluating transit performance', *Transportation Research Record* **1297**, 3–9.

Hong, Y. (2002), Transition to smart card technology: How transit operators can encourage the take-up of smart card technology, Master's thesis, Massachusetts Institute of Technology.

Hsu, P. S. (1985), Estimation of Parameters for Multiple Spatially and Temporally Distributed Populations, PhD thesis, Massachusetts Institute of Technology.

Interfleet Technology (2004), Review of passenger counting techniques, Technical report.

Jolliffe, J. K. and Hutchinson, T. P. (1975), 'A behavioral explanation of the association between bus and passenger arrivals at a bus stop', *Transportation Science* **9**, 248–282.

Kittelson & Associates, Inc et al. (2003*a*), *TCRP Report 100: Transit Capacity and Quality of Service Manual*, 2nd edn, Transportation Research Board, Washington, DC.

Kittelson & Associates, Inc et al. (2003*b*), *TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System*, "Transportation Research Board", Washington, DC.

Kurauchi, F., Bell, M. G. H. and Schmocker, J. D. (2003), 'Capacity constrained transit assignment with common lines', *Journal of Mathematical Modelling and Algorithms* **2**, 309–327.

Lam, W. H. K. and Lo, H. P. (1991), 'Estimation of origin-destination matrix from traffic counts: a comparison of entropy maximizing and information minimizing models', *Transportation Planning and Technology* **16**(2), 85–104.

Lamond, B. and Stuart, N. F. (1981), 'Bregman's balancing method', *Transportation Research B* **15**(4), 239–248.

Larson, R. C. and Odoni, A. R. (2007), *Urban Operations Research*, 2nd edn, Dynamic Ideas.

Lee, D. S.-M. (2002), Understanding capacity and performance of urban rail transit terminals, Master's thesis, Massachusetts Institute of Technology.

Lo, H. P., Zhang, N. and Lam, W. H. K. (1996), 'Estimation of an origin-destination matrix with random link choice proportions: A statistical approach', *Transportation Research Part B* **30**(4), 309–324.

Lo, H. P., Zhang, N. and Lam, W. H. K. (1999), 'Decomposition algorithm for statistical estimation of OD matrix with random link choice proportions from traffic counts', *Transportation Research Part B* **33**(5), 369–385.

London Buses (2009), *Report Functional Description – Glossary. Spearmint, Embedding Project Management in Surface Transport.*

London Transport (1999), Journey Time Metric (JTM): An Overview.

LOROL (2010), 'LOROL - Home'. `http://www.lorol.co.uk/` (Accessed February 8, 2010).

Lu, A., Aievoli, S., Ackroyd, J., Carlin, H. C. and Reddy, A. (2009), 'Passenger environment survey: Representing the customer perspective in quality control', *Transportation Research Record* **2112**, 93–103.

Luethi, M., Weidmann, U. A. and Nash, A. (2007), Passenger arrival rates at public transport stations, *in* '86th Annual Meeting of the Transportation Research Board'.

Maher, M. J. (1987), 'Bias in the estimation of o-d flows from link counts', *Traffic engineering and control* **28**(12), 624–627.

Maunder, G. (2003), *Rolling Origin and Destination Survey: The Complete Guide*, London Underground Limited.

Meyer, M. D. and Miller, E. J. (2001), *Urban Transportation Planning: A Decision-Oriented Approach*, second edn, McGraw-Hill.

Multisystems, I. (2003), *TCRP Report 92: Strategies for Improved Traveler Information*, Transportation Research Board, Washington, DC.

Multisystems, Inc et al. (2003), *TCRP Report 94: Fare Policies, Structures and Technologies: Update*, Transportation Research Board, Washington, DC.

National Rail (2010), 'National Rail Enquiries - Train Operating Companies'. http://www.nationalrail.co.uk/tocs_maps/tocs/ (Accessed February 4, 2010).

Network Rail (2009), *Delay Attribution Guide*.

Nguyen, S. and Pallottino, S. (1988), 'Equilibrium traffic assignment for large scale transit networks', *European Journal of Operations Research* **37**, 176–186.

Nielsen, B. F. (2009*a*), Unpublished electronic communications with Bo Friis Nielsen, PhD, Associate Professor of Stochastic Modelling, Department of Informatics and Mathematical Modeling, Technical University of Denmark.

Nielsen, O. A. (1998), *Two New Methods for Estimating Trip Matrices from Traffic Counts*, Elsevier, chapter 13, pp. 221–250.

Nielsen, O. A. (2009*b*), Unpublished electronic communications with Otto Nielsen, PhD, Professor in Traffic Modeling, Centre for Traffic and Transport, Technical University of Denmark.

Nielsen, O. A., Nielsen, B. F., Brun, B. and Frederikson, R. D. (2008*a*), 'Automatic OD-matrix estimation based on counting and weighing trains', Presentation slides from European Transport Conference 2008, Association for European Transport, The Netherlands.

Nielsen, O. A., Nielsen, B. F., Brun, B. and Frederikson, R. D. (2008*b*), 'Automatic OD-matrix estimation based on counting and weighing trains', Online abstract from European Transport Conference 2008, Association for European Transport, The Netherlands. http://www.etcproceedings.org/paper/automatic-od-matrix-estimation-based-on-counting-and-weighing-trains (Accessed November 10, 2009).

Nuzzolo, A. (2003), *Transit path choice and assignment model approaches*, Pergamon, chapter 4, pp. 93–124.

Nuzzolo, A. and Crisalli, U. (2004), *The Schedule-Based Approach in Dynamic Transit Modeling: A General Overview*, Operations Research/Computer Science Interface Series, Kluwer Academic Publishers, chapter 1, pp. 1–24.

Nuzzolo, A., Russo, F. and Crisalli, U. (2001), 'A doubly dynamic schedule-based assignment model for transit networks', *Transportation Science* **35**(3), 268–285.

Office of Rail Regulation (2008), National rail trends yearbook, 2007 - 2008, Technical report.

Office of Rail Regulation (2009), 'Station usage Excel sheet'. http://www.rail-reg.gov.uk/upload/xls/station_usage_0708.xls (Accessed February 8, 2010).

Office of Rail Regulation (2010), 'Glossary'. http://www.rail-reg.gov.uk/server/show/nav.001002/chooseLetter0/Show%20all (Accessed February 11, 2010).

Osuna, E. E. and Newell, G. F. (1972), 'Control strategies for an idealized public transportation system', *Transportation Science* **6**(1), 52–71.

Paramahamsan, H. (1999), Fundamental properties of synthetic O-D generation formulations and solutions, Master's thesis, Virginia Polytechnic Institute and State University.

Perl.org (2010), 'The Perl Programming Language'. `http://www.perl.org/`.

Prashker, J. and Bekhor, S. (2004), 'Route choice models used in the stochastic user equilibrium problem: A review', *Transport Reviews* pp. 437–463.

Puong, A. (2000), Dwell time model and analysis for the MBTA Red Line. Unpublished report.

Rahbee, A. (2006), Rail operations analysis and strategies for service quality improvement, Technical report, Chicago Transit Authority.

Rahbee, A. B. (1999), Rail transit operations analysis: Framework and applications, Master's thesis, Massachusetts Institute of Technology.

Rail for London Limited and MTR Laing Metro Limited (2007), 'London rail concession: Concession agreement'.

Reddy, A., Cramer, A., Cucarese, J., Lu, A. and Tran, M. (2009), 'Performance measurements on mass transit - new york city transit authority case study', *Transportation Research Record* **2111**, 125–138.

Salique, A. (2010), Unpublished interviews and electronic communications with Abdul Salique, Plan Delivery Manager, TfL London Rail.

Schmocker, J. D. and Bell, M. G. H. (2009), *The Build-Up of Capacity Problems During the Peak Hour: A Dynamic Frequency-based model*, Springer, pp. 217–239.

Schmocker, J. D., Bell, M. G. H. and Kurauchi, F. (2008), 'A quasi-dynamic capacity constrained frequency-based transit assignment model', *Transportation Research B* **42**, 925–945.

Smale, T. (2010), Unpublished interviews and electronic communications with Tony Smale, Director, Demon Information Systems LTD.

Smales, C. (2009), Unpublished interviews and electronic communications with Carol Smales, Chief Transport Economist, TfL London Rail.

Smales, C. (2010), Unpublished interviews and electronic communications with Carol Smales, Chief Transport Economist, TfL London Rail.

Small, K. and Verhoef, E. (2007), *The Economics of Urban Transportation*, Routledge, New York.

Southern Railway LTD (2009), Example of PLD benefit for journalist. Internal corporate document.

Spiess, H. (1990), A gradient approach for the od matrix adjustment problem, Technical report, EMME/2 Support Center, Aegerten, Switzerland.

Spiess, H. and Florian, M. (1989), 'Optimal strategies: A new assignment model for transit networks', *Transportation Research B* **23**(2), 83–102.

Sun Microsystems (2009), 'Developer resources for java technology'. `http://java.sun.com`.

The R Project (2009), 'The R project for statistical computing'. http://www.r-project.org/.

Tong, C. O. and Wong, S. C. (1999), 'A stochastic transit assignment model using a dynamic schedule-based network', *Transportation Research B* **33**, 107–121.

Transport for London (2007*a*), 'Creating London Overground'. http://www.tfl.gov.uk/assets/downloads/creating-london-overground.pdf (Accessed February 7, 2010).

Transport for London (2007*b*), 'Investment programme 2008: 2005/06 to 2009/10'.

Transport for London (2009*a*), *Business Case Development Manual*.

Transport for London (2009*b*), 'Docklands Light Railway Factsheet'. http://www.tfl.gov.uk/assets/downloads/corporate/dlr-factsheet-july-2009.pdf (Accessed February 8, 2010).

Transport for London (2009*c*), 'Investment programme 2009: 2009/10 to 2017/18'.

Transport for London (2009*d*), 'London Buses Factsheet'. http://www.tfl.gov.uk/assets/downloads/corporate/london-buses-factsheet-july-2009.pdf (Accessed February 8, 2010).

Transport for London (2009*e*), 'Oyster Factsheet'. http://www.tfl.gov.uk/assets/downloads/corporate/oyster-factsheet-july-2009.pdf (Accessed February 8, 2010).

Transport for London (2009*f*), 'Travel in london: Key trends and developments'.

Transport for London (2010*a*), 'Adult travelcard'. http://www.tfl.gov.uk/tickets/faresandtickets/seasontickets/4805.aspx (Accessed February 8, 2010).

Transport for London (2010*b*), 'Getting around with Oyster'. http://www.tfl.gov.uk/assets/downloads/tickets/getting-around-with-oyster-january-2010.pdf (Accessed February 8, 2010).

Transport for London (2010*c*), 'Key Facts'. http://www.tfl.gov.uk/corporate/modesoftransport/londonunderground/1608.aspx (Accessed April 20, 2010).

Transport for London (2010*d*), 'Large print Tube map'. http://www.tfl.gov.uk/assets/downloads/large-print-tube-map.pdf (Accessed May 18, 2010).

Transport for London (2010*e*), 'Oyster rail services map'. http://www.tfl.gov.uk/assets/downloads/oyster-rail-services-map.pdf(AccessedFebruary5,2010).

Uniman, D. L. (2009), Service reliability measurement framework using smart card data: Application to the london underground, Master's thesis, Massachusetts Institute of Technology.

Van-Zuylen, H. and Willumsen, L. (1980), 'The most likely trip matrix estimated from traffic counts', *Transportation Research Part B* **14**(3), 281–293.

Vescovacci, M. A. (2001), Junction capacity and performance in rail transit, Master's thesis, Massachusetts Institute of Technology.

Vuchic, V. R. (2005), *Urban Transit: Operations, Planning and Economics*, 1 edn, Wiley.

Vuchic, V. R. (2007), *Urban Transit Systems and Technology*, Wiley.

Wardman, M. (2001), 'A review of british evidence on time and service quality valuations', *Transportation Research E* **37**, 107–128.

Warner, D. (2010), Unpublished interviews and electronic communications with David Warner, Senior Business Analyst, TfL London Rail.

Wilson, N. H. M., Attanucci, J. P., Chan, J. and Cui, A. (2008), The use of automated data collection systems to improve public transport performance, *in* 'Proceedings of the 10th International Conference on Application of Advanced Technologies in Transportation'.

Wilson, N. H. M. and Lin, T. (1993), 'Dwell-time relationships for light rail systems', *Transportation Research Record* **1361**, 296–304.

Wilson, N. H. M., Nelson, D., Palmere, A., Grayson, T. and Cederquist, C. (1992), 'Service-quality monitoring for high-frequency transit lines', *Transportation Research Record* **1349**, 3–11.

Wilson, N. H. M. W. (2008), Current short range planning practice. Lecture slides from MIT class 1.258: Public Transportation Service and Operations Planning. Spring, 2008.

Wong, K. I., Wong, S. C., Tong, C. O., Lam, W. H. K., Lo, H. K., Yang, H. and Lo, H. P. (2005), 'Estimation of origin-destination matrices for a multimodal public transit network', *Journal of advanced transportation* **39**(2), 139–168.

Wong, S. C. and Tong, C. O. (1998), 'Estimation of time-dependent origin-destination matrices for transit networks', *Transportation Research Part B: Methodological* **32**(1), 35–48.

Zhao, J. (2004), The planning and analysis implications of automated data collection systems: Rail transit OD matrix inference and path choice modeling examples, Master's thesis, Massachusetts Institute of Technology.

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