

# Open Research Online

---

The Open University's repository of research publications  
and other research outputs

## Optimising the Loading Diversity of Rail Passenger Crowding using On-Board Occupancy Data

### Thesis

#### How to cite:

Ball, Simon David (2016). Optimising the Loading Diversity of Rail Passenger Crowding using On-Board Occupancy Data. MPhil thesis The Open University and Transport Research Laboratory.

For guidance on citations see [FAQs](#).

© 2016 Transport Research Laboratory

Version: Version of Record

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

Full name and degrees	<b>Simon Ball, BA (Hons)</b>
Title	<b>MPhil thesis – Optimising the loading diversity of rail passenger crowding using on-board occupancy data</b>
Degree for which it is offered	<b>MPhil</b>
Discipline to which it pertains	<b>Maths, Computing and Technology</b>
Date of submission	<b>14/05/2016</b>
Date of award	<b>22/11/2016</b>
Affiliated Research Centre	<b>Transport Research Laboratory (TRL)</b>

## Abstract

Crowded conditions on trains can lead to lower passenger satisfaction, discourage rail travel, result in negative economic impacts and are a factor in a number of health and safety hazards. In the UK there is an annual survey of rail passenger crowding, although the measures used do not reflect coach-by-coach variations, nor do they reflect variations across the peak period.

In this MPhil thesis I investigated the application of weight-based automatic passenger counting data to deliver more even loadings on trains through the provision of new real-time and static solutions. In addition I investigated the potential benefits of such solutions in terms of reduced dwell times and reduced crowding. The overall concept proposed was to make the most of the existing available capacity; for example, so that no-one is standing when seats are available. Through analysing a large sample of air suspension data, I identified station-specific trends where some coaches were over capacity while others had spare capacity. I also conducted a critical review of academic research into on-train crowding and solutions that seek to optimise 'loading diversity'.

This study contributes to this emerging subject area in several ways: I propose two new metrics to describe inter-coach loading diversity that, unlike existing metrics, contain information relative to the capacity; I have revealed a link between the inter-coach loading diversity metrics and estimated boarding times, with trains classified as 'very uneven' on departure typically having dwell times of approximately five to ten seconds greater than services that were classified as being 'even' with a similar total number of passengers on board; and finally I have applied classification supervised learning techniques to predict the load factor for a given service and these predictors were an improvement over taking the historical average.

## Contents amendment record

Version	Date	Description	Reviewed
-	28/01/2015	Probation report	JW,GK
V1	14/05/2016	Submitted for examination before the viva panel assessment	JW,GK
V2	16/08/2016	Updated to address comments from examiners	Examiners
V3	18/08/2016	Several sub-sections removed from Section 5 to allow publication	Train operator
V4	12/12/2016	Formatting changes to meet OU regulations	N/A

## Disclaimer

This report has been produced by the Transport Research Laboratory. The information contained herein is the property of TRL Limited and does not necessarily reflect the views or policies of the customer for whom this report was prepared. Whilst every effort has been made to ensure that the matter presented in this report is relevant, accurate and up-to-date, TRL Limited cannot accept any liability for any error or omission, or reliance on part or all of the content in another context.

© Transport Research Laboratory 2016

## Acknowledgements

The author would like to thank Jill Weekley, Gerd Kortuem and Alan Stevens for their supervision and advice, the TRL Academy who provided funding for this work and the train operator who permitted access to the various datasets. The author is also grateful to: Jean Hopkin who helped facilitate the focus group; Divij Jhamb who provided information on some solutions; Alistair Weare who advised on human factors considerations; Marcus Jones who reviewed sections of the thesis; and Lucy Coogan for her unconditional support.

## Contents

<b>Abstract</b>	<b>ii</b>
<b>Glossary of terms and abbreviations</b>	<b>x</b>
<b>1 Introduction</b>	<b>11</b>
1.1 Motivation: on-train crowding	11
1.2 Standpoint	14
1.3 Structure of thesis	15
<b>2 Literature review: existing research</b>	<b>18</b>
2.1 ICT and data analytics in public transport	18
2.1.1 Intelligent Transport Systems (ITS): data from source to consumption	19
2.1.2 'Found data' and the digital footprint from ubiquitous computing	21
2.1.3 Crowdsourcing	24
2.1.4 Influencing public behaviour in transport	26
2.2 On-train crowding	28
2.2.1 Definitions of capacity and crowding	28
2.2.2 Automatic passenger counting (APC) data and rail passenger crowding	31
2.2.3 Automatic fare collection (AFC) data and rail passenger crowding	34
2.2.4 Localised crowding and 'loading diversity'	40
2.2.5 Relationship between crowding and boarding times	46
2.3 Review of existing research into solutions to optimise loading diversity	48
2.3.1 Inter-coach, real-time information	49
2.3.2 Inter-coach, near-real-time seat reservations	55
2.3.3 Inter-coach, static information	56
2.3.4 Inter-coach, static passive behavioural modifiers	57
2.3.5 Inter-train, real-time information	60
2.3.6 Inter-train, static information	61
2.4 Techniques	66

2.4.1	Definitions of data mining and machine learning	66
2.4.2	Overview of particular techniques	71
2.4.3	Applications of data mining techniques in transportation	73
<b>3</b>	<b>Research focus, aims and questions</b>	<b>76</b>
3.1	Research focus	76
3.2	Research aims	80
3.3	Main research questions	82
<b>4</b>	<b>Methodology</b>	<b>84</b>
4.1	Task 1 – Inter-coach loading diversity: analysis of existing data	84
4.2	Task 2 – Inter-train loading diversity: analysis of existing data	86
4.3	Task 3 – Perceived causes and effects of loading diversity, both inter-coach and inter-train	87
<b>5</b>	<b>Findings: occupancy data</b>	<b>89</b>
5.1	Task 1 – Inter-coach loading diversity: analysis of existing data	89
5.1.1	RQ 1 – Quantification of inter-coach loading diversity	89
5.1.2	RQ 2 – Link between inter-coach loading diversity and dwell times	118
5.2	Task 2 – Inter-train loading diversity: analysis of existing data	122
5.2.1	RQ 3 – Quantification of inter-train loading diversity	122
5.2.2	RQ 4 – Prediction of inter-train loading diversity	133
<b>6</b>	<b>Findings: staff workshop</b>	<b>142</b>
6.1	Task 3 – Perceived causes and effects of loading diversity, both inter-coach and inter-train	142
6.1.1	RQ 5 – Perceived causes and effects of loading diversity	142
<b>7</b>	<b>Discussion</b>	<b>147</b>
7.1	Implications of findings	147
7.1.1	RQ 1 – Quantification of inter-coach loading diversity	147
7.1.2	RQ 2 – Link between inter-coach loading diversity and dwell times	150

7.1.3	RQ 3 – Quantification of inter-train loading diversity	151
7.1.4	RQ 4 – Prediction of inter-train loading diversity	151
7.1.5	RQ 5 – Perceived causes and effects of loading diversity	153
7.2	Limitations	154
7.2.1	Occupancy data	154
7.2.2	Dwell time data	156
7.3	Effectiveness of solutions to influence loading diversity	156
7.3.1	(a) How can the solutions be categorised, both inter-coach and inter-train?	157
7.3.2	(b) How would passengers interact with the solutions, both inter-coach and inter-train?	160
7.3.3	(c) What would be the impacts of the solutions, both inter-coach and inter-train?	162
7.3.4	(d) What are the most appropriate solutions, both inter-coach and inter-train?	166
<b>8</b>	<b>Conclusions, recommendations, contribution and further work</b>	<b>168</b>
8.1	Conclusions	168
8.2	Recommendations	171
8.3	Contribution to scholarship	172
8.4	Further work	173
<b>9</b>	<b>References</b>	<b>175</b>

## List of tables

Table 1 – Platform gaps and options: extract from Table 5.8 of the Network Rail Stations RUS (reproduced from Network Rail 2011)

Table 2 – Categories of machine learning algorithms (reproduced from Brownlee 2013)

Table 3 – Comparison of crowding classifications

Table 4 – Proportion of observations where some coaches had standing passengers while other coaches had spare seats (Sample 2)

Table 5 – Proportion of observations for ‘busy’ services where some coaches had standing passengers while other coaches had spare seats (Sample 2)

Table 6 – Proposed new metric of ‘Two busiest and two quietest coaches’: definition of six classifications to describe inter-coach loading diversity

Table 7 – Partition by ‘Two busiest and two quietest coaches’ metric (Sample 2)

Table 8 – Partition by ‘Two busiest and two quietest coaches’ metric for ‘busy’ services (Sample 2)

Table 9 – Proposed new metric of ‘Rear-Middle-Front’ inter-coach loading diversity classification: definition of 27 ordered classes

Table 10 – Description of Naïve Bayes models

Table 11 – Model predictions compared to actual observations (Sample 2, test dataset)

Table 12 – Decision tree predictions compared to actual observations (Sample 2, test dataset)

Table 13 – Proposed categorisation of ‘solutions toolkit’ to optimise loading diversity

Table 14 – Human factors considerations for each category in the toolkit

Table 15 – Considerations for pilots and impact evaluations for each category in the toolkit



## List of figures

Figure 1 – Weekday entrances (red) and exits (blue) at Canary Wharf Underground station (reproduced from Ceapa et al. 2012; © 2012 ACM, Inc. Included here by permission)

Figure 2 – Passenger flows on 11 March 2011, Tokyo, Heat Map View (left) and Route Map View (right) (reproduced from Itoh et al. n.d.)

Figure 3 – ‘Ratio of car occupancy to train average’: Toronto, Yonge Subway, Wellesley Station, southbound direction, Jan 11th 1995, 99 trains, 66,263 passengers (reproduced from TRB 2003)

Figure 4 – Train load across the extended AM peak period: Toronto, Yonge Subway, Wellesley Station, southbound direction, Jan 11th 1995 (reproduced from TRB 2003)

Figure 5 – Pilot of real-time smartphone app (reproduced from NS 2013)

Figure 6 – Pilot of real-time platform information system (reproduced from De Vos 2013)

Figure 7 – Fixed signs pointing to less busy ends of the platform on the London Underground

Figure 8 – Expected occupancy levels when buying tickets online (reproduced from SBB n.d.)

Figure 9 – Posters used in the South West Trains pilot project (reproduced from ORR 2012)

Figure 10 – Overview of tasks

Figure 11 – Distribution of estimated number of passengers on train upon departure (Sample 1)

Figure 12 – Distribution of estimated number of passengers in each coach upon departure (Sample 1)

Figure 13 – Distribution of occupancy in each coach upon departure (Sample 1)

Figure 14 – Distribution of occupancy in each coach upon departure, with four classes (Sample 1)

Figure 15 – Distribution of occupancy in each coach upon departure, with four classes (Sample 2)

Figure 16 – Comparison of estimated passenger counts against ticket data (subset of Sample 2, where it was possible to match to ticket counts; n=9,964)

Figure 17 – Percentage of variance explained for each number of clusters tested (average across all ten repetitions)

Figure 18 – Proportion of observations where some coaches had standing passengers while other coaches had spare seats, by time of day (Sample 2)

Figure 19 – Proportion of passengers who were standing, by time of day (Sample 2)

Figure 20 – Proportion of passengers who were standing but could have been sitting down, by time of day (Sample 2)

Figure 21 – Proposed new metric of ‘Two busiest and two quietest coaches’: illustration

Figure 22 – Partition by ‘Two busiest and two quietest coaches’ metric, by time (Sample 2)

Figure 23 – ‘Rear-Middle-Front’ inter-coach loading diversity classification applied to departures from one station (Sample 2)

Figure 24 – Distribution of dwell time for services that arrived late (Sample 3)

Figure 25 – Average dwell time for the six classes of inter-coach loading diversity, by total number of passengers on-board, short trains (Sample 3, trains that arrived late)

Figure 26 – Average dwell time for the six classes of inter-coach loading diversity, by total number of passengers on-board, long trains (Sample 3, trains that arrived late)

Figure 27 – Distribution of load factor (total passengers / capacity) upon departure (Sample 2)

Figure 28 – Load factor average (top) and banded (bottom), by scheduled departure time period (Sample 2)

Figure 29 – Load factor average (top) and banded (bottom), by scheduled departure half-hour time period (Sample 2)

Figure 30 – Load factor average (top) and banded (bottom), by day of week (Sample 2)

Figure 31 – Load factor average (top) and banded (bottom), by month (Sample 2)

Figure 32 – Load factor average (top) and banded (bottom), by school holiday (Sample 2)

Figure 33 – Load factor average (top) and banded (bottom), by late-running (Sample 2)

Figure 34 – Average load factor +/- standard deviation for one station (Sample 2)

Figure 35 – Average load factor +/- standard deviation for another station (Sample 2)

Figure 36 – Relationship between inter-coach and inter-train occupancy (Sample 2)

## Glossary of terms and abbreviations

AFC	Automatic fare collection
APC	Automatic passenger counting
Average	Used to describe the mean average unless stated otherwise
CIS	Customer information systems
Critical load point	“the location where the passenger load on a service is highest on arrival at (AM peak) or on departure from (PM peak) a city”
DfT	Department for Transport
FOI	Freedom of information request
GJT	Generalised journey time
ICT	Information and communication technology
MOIRA2	The revenue and demand forecasting model used in the rail industry
NRPS	National Rail Passenger Survey
ORR	Office of Rail Regulation
PiXC	Passengers in excess of capacity
RSSB	The organisation that manages the rail industry research programme
RUS	Network Rail Route Utilisation Strategy reports
Standard class load factor	The number of standard class passengers expressed as a percentage of the maximum allowable standard class passenger capacity for that service
Transport Focus	The independent watchdog for Britain’s rail and bus passengers
TSUG	Transport Statistics User Group

## 1 Introduction

### 1.1 Motivation: on-train crowding

On-train crowding is an increasingly pressing issue for many UK rail passengers and train operators alike. In October 2010, the Public Accounts Committee produced a report called 'Increasing Passenger Rail Capacity' (PAC 2010), which raised concerns on "substantial increases in already unacceptable overcrowding levels [on trains]" and called for the DfT to consider obligations to limit rail passenger crowding in all franchise agreements.

A qualitative research study commissioned by Transport Focus, the independent watchdog for Britain's rail and bus passengers, identified rail crowding as the second most important measure of performance from the perspective of passengers (Transport Focus 2011). The study found that the principal measures of performance in descending order of importance to passengers were: punctuality and reliability of services; comfort and space on board and being able to get a seat; fares; customer service; and journey times. Transport Focus also conducts a large survey twice per year called the National Rail Passenger Survey (NRPS). The results from the Spring 2014 survey showed that out of a list of 19 questions for on-train facilities and services, the question on "sufficient room for all passengers to sit/stand" had the sixth highest proportion (21%) who responded "dissatisfied or poor" (Transport Focus 2014).

The Office of Rail Regulation (ORR) is the health and safety authority for Britain's railways and is responsible for regulating health and safety aspects of passenger crowding. They say that, "crowding can be inconvenient, uncomfortable... make passengers feel unsafe... and lead to unpleasant travelling conditions, especially when passengers must stand very close together for prolonged periods" (ORR 2014).

Crowding affects many parts of the UK rail network and is particularly acute for some routes into London. In January 2015 the London Assembly wrote to Network Rail (London Assembly 2015) stating that two of their main concerns were the high projected increase in rail demand and a focus on passengers: "High levels of growth will put severe pressure on the national rail network serving London... significant enhancements will therefore be necessary on many routes and corridors... The Committee urges Network Rail to consider its responsibilities to passengers, and to prioritise those interventions which will: deliver sufficient capacity, particularly sufficient

seats, to prevent passengers from having to stand on journeys of over 20 minutes in length; minimise station overcrowding; improve punctuality and reliability of trains...".

Piczenik (2013) presented to the Transport Statistics User Group (TSUG)<sup>1</sup> on how rail passenger crowding in the UK is modelled and forecasted, which included a discussion on the effects of crowding. It was suggested the implications of rail crowding are as follows:

- "Peak spreading to less crowded trains... in the passenger's preferred travel time
- Suppression of journeys as passengers decide not to travel... lost revenue for the operators
- Shift to other modes of travel resulting in loss of revenue and highway congestion
- Reduction in work activity undertaken on the train impacting on the economy
- Passenger safety risks
- Delays due to additional boarding and alighting times at stations: both primary delay due to exceeding timetabled stop time and also reactionary delay to other trains
- Additional costs for the train operating companies in exercising 'best endeavours' when PiXC benchmarks are exceeded - although not contractual<sup>2</sup>
- Lower National Rail Passenger Survey (NRPS) scores which can result in requirement for the train operator to invest in remedial measures."

Specifically with respect to the impact of crowding on rail demand, Piczenik explained how a crowding penalty is included in MOIRA2, the revenue and demand forecasting model used in the rail industry. Essentially MOIRA2 uses the concept of Generalised Journey Time (GJT) to take into account various penalties relating to waiting times, interchanges, performance and crowding. The crowding penalty is a scale factor that is

---

<sup>1</sup> The TSUG is a monthly forum that discusses use of statistics and data in transport.

<sup>2</sup> As discussed in Section 2.2.1, the PiXC is no longer used as a target that train operators must meet.

multiplied by the in-vehicle time. There are different values depending on whether the passenger is seated or standing and how crowded the train is, i.e. the standard class load factor. The precise values of the crowding penalty also vary for different services. The purpose of the MOIRA2 model is to enable train operators to assess revenue and demand forecasts in future or alternative scenarios, for example timetabling changes or fare increases. These crowding penalty factors provide a useful insight into how the industry models the effects of crowding.

Various other studies have found slightly different values for the crowding penalty, although the principle of applying a penalty scale factor to the value of time is the same. A thorough review of crowding penalties can be found in Wardman and Whelan (2011).

RSSB, formerly known as the Rail Safety and Standards Board, manages the rail industry research programme on behalf of industry sponsors, with funding mainly from DfT and from collaborative partnerships. RSSB (2005) conducted a piece of cross-industry research to establish an understanding of the health and safety issues that may arise as a result of crowding on main line and underground railways. There were a range of suggestions on the actual definition of crowding from different perspectives; for example, industry defined it as “not being able to shut the doors”, whereas passengers defined it as “no spare seats” for inter-city journeys and “no handholds or standing for more than 10-20 minutes” for commuter journeys. Also, the literature review conducted by RSSB as part of the study revealed that some definitions related to objective elements such as density and the available space, whereas other definitions focused on more subjective elements on the perception of crowding.

Stakeholder interviews for RSSB defined four qualitative levels of perceived crowding on the platform: ‘No crowding – All of body visible’; ‘Low to moderate crowding – Only body and head visible’; ‘Severe crowding – Only shoulder and head visible’; and ‘Unacceptable crowding – Only head visible’ (RSSB 2003). Building on this, RSSB also proposed four similar qualitative levels of perceived crowding on the train (RSSB 2005):

- *“No crowding – Available seats, available luggage space, unrestricted movement throughout train.*

- *Low to moderate crowding – Standing for short periods of time (less than 20 minutes), available handholds, luggage space taken but not obstructing movement, some restricted movement in moving down train.*
- *Severe crowding – Standing for moderate periods (20 minutes to 1 hour), small amounts of luggage restricting movement, cannot get catering trolley through train, train delayed at station due to alighting and boarding times.*
- *Unacceptable crowding – Standing for long periods (>1 hour), standing for short periods without access to handhold, luggage obstructing movement throughout train, passengers or staff cannot move through train, cannot shut train doors.”*

Industry workshops and stakeholder groups were used to create a list of crowding-related hazardous events. This list was ranked in terms of a qualitative estimation of risk as follows: “1. Slips, trips or falls in station area; 2. Ill health e.g. fainting; 3. Assaults; 4. Platform train interface; 5. Evacuation; 6. On-train incidents; 7. Other station area accidents; 8. Train accidents. It was concluded that “In context with the risk from other non-crowding related hazards on the railway, the risk associated with crowding was found to be small, [although] the perception of risk from crowding appears higher, among the public and rail users [than compared to the findings of the study]”.

In this thesis, the terminology ‘crowding’ is taken to mean ‘passenger density’ rather than relating to the more qualitative definition of ‘perceptions of crowding’.

---

## **1.2 Standpoint**

---

Through this thesis I sought to address the issue of rail passenger crowding in the UK from a data analytics standpoint. The primary focus was on extracting more value from existing ‘found data’ by reusing air pressure data from suspension systems as a measure of carriage loadings along a train, to help assess the case for investing in initiatives to optimise passenger loadings. The overall concept proposed was not to increase total capacity, but rather to help make the most of the existing available capacity, for example, so that no-one is standing when seats are available.

Over recent years there has been a proliferation of new and larger datasets, which has been heralded by many as an opportunity for innovation, productivity and economic growth (Manyika et al. 2011). There is much ongoing work to develop new tools and

techniques to exploit these opportunities, which has seen the emergence of several related disciplines, some of which are briefly introduced here. Microsoft has described this phenomenon as the “Fourth Paradigm and the age of data-intensive scientific discovery” (Hey et al. 2009), whereas the term ‘data science’ is used by others and has been defined as “extracting predictive and prescriptive insight from large-scale complex data, data exploration, data discovery and delivering surprise” (Data Science London 2014). Others have defined data science as “the field of study which is concerned with the collection, preparation, analysis, visualisation and management of data,... which builds upon and incorporates the expertise of many different disciplines to successfully extract meaning from data” (Knowledge Transfer Network & Smith Institute 2014). There exists a large number of data mining and machine learning techniques, drawing on different aspects of maths, statistics and computing. Zaki and Meira (2014) have published a free textbook with introductions to some of these techniques, which they say “form the basis for the emerging field of data science, which includes automated methods to analyse patterns and models for all kinds of data, with applications ranging from scientific discovery to business intelligence and analytics”.

A related phrase is ‘big data’, which has taken on different meanings for different people. Some definitions focus on the large amounts of data involved, i.e. ‘volume’ (Kelly and Hamm 2013); other sources consider a wider definition, which as well as high volumes includes data that is frequently refreshed i.e. ‘velocity’ and/or data that does not have one consistent structure i.e. ‘variety’ (Kincaid 2014). Big data has quite a specific meaning for some to mean “massively parallel software running on up to thousands of servers”, in particular using software tools such as Hadoop and Spark (Adhikari 2011). However, use of the phrase also has become widespread with other meanings to the point that some commentators in both the print and online media are now sceptical: “it has become a synonym for data analysis, which is confusing and counter-productive” (Mims 2013); “big data is a vague term, often thrown around by people with something to sell” (Harford 2014).

Regardless of the terminology it is clear that there are new emerging opportunities for deriving insights as well as developing new products and services across a wide range of domains. One such domain is transportation, which has seen the integration of ICT across different modes in recent years and hence new datasets.

### **1.3 Structure of thesis**

The thesis is structured as follows:



- §2 is a critical review of relevant academic literature;
- §3 describes the research focus, aims and questions;
- §4 outlines the methodology and the datasets used, along with rationale for the proposed approaches;
- §5 presents the findings of the main data analysis task and is supplemented by the findings of a workshop with frontline staff in §6;
- §7 discusses the implications of the findings; and
- §8 completes the thesis with conclusions, recommendations, a summary of the contribution to scholarship and suggested areas for further work.

## Summary

Various organisations have raised concerns over the level of rail passenger crowding on some trains in the UK. There is research to suggest that conditions that are near or over capacity discourage passengers from using the train; this effect is modelled using ‘crowding penalties’, which are included in revenue and demand forecasting models. Crowding can also lead to negative economic impacts, such as a reduction in work activity on the train and a shift to other modes, which may lead to an increase in road congestion.

Qualitative research suggests that crowding was found to be a factor in a number of health and safety hazards, although in the context of the risk from other non-crowding related hazards on the railway, the risk associated with crowding was found to be “small”. The research also suggested a need to distinguish between objective elements such as passenger density and subjective elements such as perceptions of crowding. In this thesis, the terminology ‘crowding’ is taken to mean ‘passenger density’ rather than relating to the more qualitative definition of ‘perceptions of crowding’.

Over recent years, in many different domains, there has been a proliferation of new and larger datasets and this has been accompanied by the emergence of data science as a discipline with new techniques and software. Through this thesis I sought to address the issue of rail passenger crowding in the UK from a data analytics standpoint. The primary focus was on extracting more value from existing ‘found data’ by reusing air pressure data from suspension systems as a measure of carriage

loadings along a train, to help assess the case for investing in initiatives to optimise passenger loadings.

## **2 Literature review: existing research**

The literature review undertaken covers several related areas and the objectives were as follows:

- Understand existing research within the topic areas
- Identify gaps and the current boundaries of the research
- Identify methods used by others that may be applicable
- Help focus the research questions of this study.

The literature review begins in §2.1 by situating the thesis in the context of ICT and data analytics in public transport. In §2.2 the review focuses on research into on-train crowding. This is followed in §2.3 with a more specific review of existing solutions to optimise ‘loading diversity’. This review is concluded with a brief discussion of relevant data mining techniques, both in general and in a transportation context in §2.4; I applied some of these techniques in the research.

I systematically searched a variety of transport research and computer science journals and conference proceedings, including but not limited to: Transport Research Parts A (Policy) and C (Emerging Technologies); Journal of Public Transportation; Transportation Research Record, Journal of the Transportation Research Board; RSSB research catalogue; Pervasive and Mobile Computing; ACM Computing Surveys; Intelligent Transport Systems UK Public Transport Interest Group, Transport Statistics User Group; International Conference on Data Mining (ICDM); Conference on Knowledge Discovery and Data Mining (KDD); International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp). This was supported by a more general internet search and discussions with colleagues for sources outside academic literature. I used a variety of search terms including but not limited to: rail passenger crowding; loading diversity; passenger occupancy; overcrowding; automatic passenger counting; automatic fare collection; smart card.

---

### **2.1 ICT and data analytics in public transport**

---

This section explores different aspects of ICT in public transport and aims to situate the thesis in the wider context of related research. There are several categories for uses of

data to be considered in a transportation context (Ball 2014) and these are explored in the following sections:

- Data flows from ‘source to consumption’ in Intelligent Transport Systems (§2.1.1);
- ‘Found data’ whereby the ‘data exhaust’ from existing technologies are put to a secondary use; related to this is the ‘digital footprint’ of transport users using ubiquitous computing systems (§2.1.2);
- Crowdsourcing, i.e. actively collecting transport data from people (§2.1.3);
- Using data to influence public behaviour (§2.1.4).

### **2.1.1 Intelligent Transport Systems (ITS): data from source to consumption**

Intelligent Transport Systems (ITS) has been defined as “an umbrella term for a range of information and control systems aimed at... improving travel efficiency and reducing the risk of accidents... the core technologies employed are sensors, communication links, processing elements, displays and actuators” (Gillan 2009). By this definition a necessary component in all ITS is the flow of data from source to consumption. The Internet of Things (IoT) is a related term used across multiple domains that expresses the concept of everything being connected in some way, some people have called it ‘things that think’. As such, ITS can be considered as a subset of the more general IoT concept. A key point for all ITS is that various data quality aspects need to be considered at the system design phase to meet the required level so that the data is fit for purpose for the correct operation of the system. Similarly, data security needs to be considered for vulnerabilities to attacks and also privacy where personal details are involved (Ball et al. 2009).

In 2005 the DfT set out its ‘ITS Policy Framework for the Roads Sector’ and identified seven themes for ITS: improving road network management; improving road safety; better travel and traveller information; better public transport on the roads; supporting the efficiency of the road freight industry; reducing negative environmental impacts; supporting security, crime reduction and emergency (DfT 2005).

Within public transport, an example of ITS is passenger information systems. In addition to traditional information channels, such as posters and timetables, over recent

years the provision of passenger information by electronic means has become widespread. However, these are often for single modes of transport and as such the EC has identified multimodal traveller information and planning services as a priority through the ITS Action Plan (EC 2010); work is ongoing across Europe to deploy such multimodal systems (EC 2015). A relatively recent innovation is that transport operators are increasingly using social media websites, such as Twitter, as an information channel for both broadcasting travel alerts and engaging in two-way communication with passengers (Gault et al. 2014, Transport Focus 2012).

A related concept to ITS is that of 'Smart Cities', where data is combined for several city functions such as transportation, energy, buildings and public safety (IBM n.d.). Research initiatives include the Smart Streets project, which aimed to "develop an open data hub for road maintenance [to] facilitate new innovation across the industry leading to cost reductions and new revenue opportunities" (Smart Streets Hub n.d.). City councils, such as Birmingham, have developed their own smart city road maps detailing their plans in this area (Birmingham City Council 2012).

One subset of ITS is the deployment of very many sensors across a region for monitoring purposes. The Message project demonstrated the potential of diverse, low cost sensors to provide data for the planning, management and control of the environmental impacts of transport activity at an urban, regional and national level. This included implementation of 'pervasive sensors' at a trial site in Gateshead on vehicles and people to act as mobile real-time environmental probes, sensing transport and non-transport related pollutants and hazards (Bell et al. 2009).

Another subset of ITS is cooperative vehicle systems, which involves communication and sharing of information dynamically between vehicles or between vehicles and infrastructure. The scope ranges from applications which warn the driver, to those which potentially take control from the driver in safety-critical situations. Driverless cars and other automated vehicles are arguably an advanced extension of ITS and are expected to generate substantial benefits: "They will make driving easier, allow people to be more productive and offer greater mobility to a wider range of people than ever before, they will also help improve road safety, reduce emissions and ease congestion" (DfT 2015). A range of 'semi-autonomous' technologies are becoming increasingly widespread, which are likely to be a stepping stone towards fully autonomous vehicles (Ball 2011).

## **2.1.2 'Found data' and the digital footprint from ubiquitous computing**

The terminology 'found data' is used to describe situations where the 'data exhaust' from existing technologies are put to a secondary use, either through developing new real-time applications or deriving insights from historical analysis. In the Private Sector, organisations are working to understand and exploit the value of their data (Knowledge Transfer Network & Smith Institute 2014), while in the Public Sector many datasets are being made openly available (HM Government 2012).

Ubiquitous computing is the concept of people being able to interact with computing anywhere; it is defined by the Pervasive and Mobile Computing journal (n.d.) as "ambient intelligence where network devices embedded in the environment provide unobtrusive connectivity and services all the time". Sometimes when people use ubiquitous computing systems they will generate what some might call a 'digital footprint', for example smartphone GPS tracks or Oyster card usage histories. These can be a rich 'passive' source for data mining, where applications may range from personalised recommendation systems to country-wide traffic state estimation. The 'digital footprint' is a type of 'data exhaust', although the distinction is that the digital footprint relates to a person rather than a machine. As mentioned above, wherever personal details are involved it is vital to ensure that privacy aspects are considered and addressed.

There exists an emerging field of transport research around 'found data' and the 'digital footprint' from ubiquitous computing; some example topics of research in this area are summarised below.

### **2.1.2.1 Mobility data and mobility profiling**

The areas of mobility data and mobility profiling are concerned with techniques and applications of monitoring people in motion. A good overview of the area is provided by Renso et al. (2013) and there are many examples of research in this area, for example as can be found on the Italian KDD Lab website (KDD n.d.).

GPS vehicle tracks have been used to predict the end-to-end route of a vehicle based on past trips. For example, Microsoft built personalised predictive models based on a large historic sample of GPS tracks around Seattle (Froehlich and Krumm 2008). The software company, Pivotal, conducted a similar demonstration although with a real-time element (Dobson 2014).

A review paper by Calabrese et al. (2014) provides a useful overview of research and techniques in using mobile phone cellular network data in urban environments, referencing research in the areas of population distribution, land use, mobility patterns, local events and social networks. Regarding research into mobility patterns, they cite examples of monitoring the movements of both groups and individuals and the degree to which human behaviour is predictable. However, none of these examples relate specifically to rail travel.

Inrix have demonstrated a technique using cellular data from mobile phones to understand peoples' trip origins for a particular train route (Graham and Petrie 2012). As an example, they filtered out users travelling into London Paddington station by rail on one weekday morning and then plotted where the passengers started their journeys. They estimated that through this technique (in partnership with one mobile operator) they could monitor 20-30% of all movements on every mode every day. Although this is quite impressive and offers great potential for analysing origins and destinations, it is perhaps still not sufficiently accurate for passenger counting on specific trains.

In addition to long-range technologies, short-range technologies such as Wi-Fi and Bluetooth are also commonly used to track the movement of people. For example, organisations such as Path Intelligence have developed techniques to monitor footfall movements for the retail industry (Path Intelligence 2015). Such techniques have also been applied to station environments to understand in detail how passengers interact with the station environment. However, similar to cellular data, the penetration rate of equipped devices is not accurately known and so these technologies are not as accurate as other technologies for passenger counting.

#### **2.1.2.2 Insights into how users interact with cycle hire schemes**

Cycle hire schemes with multiple stands across city regions typically use electronic systems so that users can hire pedal cycles without the need for human interaction (Bike Sharing Map n.d.). As a by-product such systems typically generate data on stand occupancy and origin and destination pairs for each trip. There is an active research community that is using such data to generate insights on how users interact with cycle hire schemes (IFSTTAR 2012).

Kaltenbrunner et al. (2010) used stand occupancy data for the Barcelona cycle hire scheme to detect temporal and geographic mobility patterns within the city. They then built models that were able to predict the number of available bikes for any stand a few

hours into the future. When the London Barclays Cycle Hire (since renamed to Santander Cycle Hire) was launched users were required to apply for a key to access to the system; after six months this was changed so that anyone in possession of a debit or credit card could gain access. Lathia et al. (2012) investigated how the change affected the system's usage and found that there was greater weekend usage and a number of stations underwent a complete change. Researchers at City University have used the same origin-destination dataset to generate interesting visualisations of the London scheme (Wood et al. 2011).

A pilot trial of electrically-assisted bicycles was conducted in Brighton, in which data on the usage of the electric assistance coupled with a GPS track was sent via a smartphone back to a central database. The data was analysed to investigate how participants used the three levels of assistance when travelling up hills (Kiefer and Behrendt 2015).

#### **2.1.2.3 Transport mode detection from smartphone data**

Another active area of related research is using data from smartphones to detect the mode of transport of travellers. A variety of methods exist that use a combination of accelerometer data, GPS and static reference data. Libby (2008) developed a method based solely on accelerometer data in order to distinguish between sitting/standing, walking, cycling and vehicle passenger travel; likewise Hemminki et al. (2013) used just accelerometer data. Some studies have used purely GPS data to predict transport mode (Bolbol et al. 2012, Zheng et al. 2010), while others such as Reddy et al. (2010) have developed a technique that used a combination of GPS and accelerometer data. Several other methods have been developed that use various other data sources such as bus and rail data in addition to GPS and accelerometer data (Li et al. 2011, Stenneth et al. 2011, Manzoni et al. 2011, Thiagarajan et al. 2010). Applications of such techniques are varied, one example being the CarbonDiem smartphone app, which provides information to the user on their CO<sub>2</sub> consumption (CarbonDiem n.d.).

#### **2.1.2.4 Social media mining in transport**

An emerging area of research is mining of social media as a 'passive' or 'implicit' data source. This is distinct from the use of social media as an 'active' or 'explicit' data source i.e. crowdsourcing (see §2.1.3) or the use of social media by transport operators as an information channel for engagement with transport users (see §2.1.1).



A recent collaborative research project collected and analysed tweets relating to travel to and from football matches. They investigated the methods, challenges and opportunities for enhancing transport data collection through social media (Grant-Muller et al. 2015), as well as considering the policy implications for such techniques (Grant-Muller et al. 2014). A similar project analysed passenger sentiments using Twitter data relating to the Chicago metro system (Collins 2013). IBM conducted sentiment analysis for commuters in different cities across Europe by mining Twitter data (IBM 2013). Some 'observatories' have recently emerged for transport-related tweets, such as TransportBuzz (Watt 2013) and CommuteLondon (CommuteLondon n.d.).

There is substantial research into social media mining in other domains. In the UK the Analysing Social Media Collaboration (ASMC) are a "multidisciplinary group of researchers interested in analysing data from social media such as Twitter with the aim to understand the role they play in social phenomena". The Collaborative Online Social Media Observatory (COSMOS) is a related software platform that aims to reduce the technical and methodological barriers to accessing and analysing social media and other forms of open digital data (COSMOS Project n.d.).

Nielsen (2006) discusses that a limitation of social media mining is the 'participation inequality', i.e. the situation where a minority of users usually account for a disproportionately large amount of content. The rule of thumb is that "In most online communities, 90% of users are 'lurkers' who never contribute, 9% of users contribute a little, and 1% of users account for almost all the action". Such a sample can give a biased understanding of the community, because the overall system may not be representative of average users, because "many differences almost certainly exist between people who post a lot and those who post a little".

With regards to limitations of social media mining in transport, some research has found that the seriousness of transport as a 'problem' in people's lives can become exaggerated and that it can be a means of "celebrating rather than solving" transport problems (Lyons 2012).

### **2.1.3 Crowdsourcing**

Ross (2012) defined crowdsourcing as "engaging the masses to produce large data/information sets... the most common type being voluntary, personal interest or group interest, active online data contributions". Furthermore it was claimed that in a transportation context crowdsourcing is "nearly always voluntary but 'rewards' can be

offered, can be connected to a niche community but not exclusively, [there is] often a geographical element but not exclusively, [there are] mostly active contributions e.g. mapping for OpenStreetMap but can be passive e.g. floating vehicle data, mostly online and increasingly via mobile". It may be argued that transport crowdsourcing is a type of ITS, although with the data source being from human-created content rather than generated from sensors.

There have been several recent research projects that have developed and piloted transport crowdsourcing smartphone apps: the 'Tiramisu' app was a real-time arrival and bus crowding information app, which was powered by crowdsourcing (Steinfeld et al. 2011); Lancaster University developed the 'OurTravel' app which provided multi-modal information for particular communities through crowdsourcing (Harding et al. 2013); the 'GetThere' app was developed by the University of Aberdeen to gather traveller information for rural buses from bus users (Corsar et al. 2013).

Other organisations have also built proof-of-concept transport crowdsourcing apps: researchers at the University of Cambridge and UCL developed the 'TubeStar' app, which actively collected feedback from passengers on their experiences on London Underground journeys via Twitter (Ryan 2012); TRL built the 'How's My Trip' app that aimed to collect feedback on journey quality and the perceived value of time from users (Hopkin et al. 2014); a project by the Transport Systems Catapult called 'CommonRoute' collected journey quality feedback via a mobile website (TSC 2014).

A key issue with transport crowdsourcing is identifying the motives for encouraging participation; all of the crowdsourcing apps listed above were pilot projects and as such had relatively low uptake. Some studies provided a financial incentive for participation in the research study, such as 'Tiramisu', whereas in other studies, such as 'Tube Star', there was only the 'altruistic incentive' in that users were sharing information directly with other travellers. Some studies concluded that transport crowdsourcing apps would be useful for service quality monitoring or issue detection (which does not rely receiving data from a representative sample of users), but may not be appropriate to measure passenger satisfaction due to issues of self-selection bias and the 'participation inequality' (Hopkin et al. 2014).

Several transport crowdsourcing websites exist, whereby members of a community with a particular group interest share information. Examples include: CycleStreets, a "UK-wide cycle journey planner system designed by cyclists for cyclists" (CycleStreets n.d.); JustPark, which is a Private Sector enterprise that connect drivers in search of

parking with anyone who has a spare space (JustPark n.d.); FixMyTransport, which although no longer supported was a website where members of the public could identify issues with public transport, raise queries with the relevant authorities and ‘up-vote’ important issues (FixMyTransport n.d.).

#### **2.1.4 Influencing public behaviour in transport**

Transport operators have attempted to influence public behaviour in transport for many years, ranging from general policies intended to create safer and more efficient transport systems, through to strategies to influence the time that people travel and mode of transport used, such as ticket pricing and road pricing. The Canadian VTPI maintain the Online Transport Demand Management (TDM) Encyclopaedia and they define TDM as a “general term for strategies that result in more efficient use of transportation resources” (VTPI 2014). Their encyclopaedia provides information on a wide range of traditional demand management strategies.

There have been recent research projects that have used ICT to encourage use of more sustainable transport. One such example was the EC-funded project, Sunset, in which participants used a smartphone app to build up their own profile and were encouraged to use sustainable travel with incentives of ‘rewards’ with a social gaming element (Bijlsma 2014). Users were set challenges, such as travelling outside of the peak period if possible. A variety of methods for user feedback were tested, which included cartoons of a pig whose mood reflected the environmental costs of the user’s travel choices and also a man rowing fast or slow as a measure of personal fitness. There were personalised messages and recommendations, such as ‘Why don’t you try walking to the shops today?’. The app included two-way communication between operators and users and was piloted in both Leeds and Sweden. The app used a combination of GPS, cellular data and Wi-Fi to record position accurately and there was an online dashboard for querying the data. The findings suggested that some users were surprised with the information and didn’t realise the extent of their current transport habits. Some people modified their behaviour and shifted outside the peak times, whereas some wanted to but couldn’t. It was suggested that this is a common finding and that better information is a good first step towards behaviour change. (Grant-Muller 2015).

A similar initiative was a research project called ‘Persuasive Advisor for CO2-reducing cross-modal trip planning (PEACOX)’. This project developed a smartphone application that aimed to “provide travellers with personalised multi-modal navigation tools that

allow, help and persuade them to travel and drive ecological friendlier... that focus on minimising energy consumption and pollutant emissions". Deliverables included guidelines and recommendations for the design and implementation of persuasive systems in the context of personal mobility and trip planner systems (Schrammel et al. 2015).

ICT initiatives exist to encourage car drivers to adapt their driving style to be more fuel efficient. One such example is the 'ecoDrive' app by Fiat, which gives the driver a score out of 100 based on speed and acceleration data. Initially this was in the form of a smartphone app, but more recent versions have involved an interface with on-board data, for example gear selection and pedal usage (Fiat 2010).

### Summary

Intelligent Transport Systems (ITS) is an umbrella term for a range of information and control systems aimed at improving travel efficiency and reducing the risk of accidents, with core technologies being sensors, communication links, processing elements, displays and actuators. Because of the flow of data from source to consumption ITS can be considered as a subset of the more general Internet of Things (IoT) concept. Examples of ITS in public transport include passenger information systems and smart ticketing.

The terminology 'found data' is used to describe situations where the 'data exhaust' from existing technologies are put to a secondary use, either through developing new real-time applications or deriving insights from historical analysis. One example research area is that of cycle hire schemes, whereby researchers are using stand occupancy and origin and destination trip pairs to generate insights on how users interact with such systems. The more general areas of mobility data and mobility profiling are concerned with techniques and applications of monitoring people in motion and can be based on long-range technologies, such as GPS and cellular or short-range technologies, such as Bluetooth. Other active areas of related research include using data from smartphones to detect the mode of transport of travellers and also the mining of transport-related tweets as a passive data source.

There have been several recent research projects that have actively collected data from travellers through developing and piloting transport crowdsourcing apps. Some studies concluded that transport crowdsourcing apps would be useful for service quality monitoring or issue detection, but may not be appropriate to measure passenger

satisfaction due to issues of self-selection bias and the 'participation inequality'. Encouraging uptake and understanding people's motivation is also a key challenge for transport crowdsourcing.

Transport operators have attempted to influence public behaviour in transport for many years, ranging from general policies intended to create safer and more efficient transport systems, through to strategies to influence the time that people travel and mode of transport used. Examples of research projects were identified where ICT has been used to encourage use of more sustainable transport; some studies found that providing better information on available alternatives was effective in encouraging behavioural change.

## 2.2 On-train crowding

A review of relevant research literature has been conducted to investigate the state of the art with respect to on-train crowding. The review begins with definitions of train capacity and the methods that are used for measuring crowding on trains in the UK (§2.2.1). This is followed by consideration of data-intensive methods that have been used to monitor and predict rail passenger crowding, using both automatic passenger counting (APC) data e.g. door sensors and weight sensors (§2.2.2) and also automatic fare collection (AFC) data e.g. smart cards (§2.2.3). There is a review of research into local crowding and 'loading diversity' (§2.2.4) and also the relationship between crowding and boarding times (§2.2.5).

### 2.2.1 Definitions of capacity and crowding

The Department for Transport (DfT) publishes an annual report called "rail passenger numbers and crowding on weekdays in major cities in England and Wales" (DfT 2014). The methodology defines on-train capacity as follows (DfT 2013a):

*"The standard class capacity includes the number of standard class seats on the service and may include an allowance for standing room. No allowance for standing is made on a service when the time between stations before (AM) or after (PM) the*

*critical load point<sup>3</sup> is more than 20 minutes, but it is allowed when it is 20 minutes or less. The allowance for standing varies with the type of rolling stock but, for modern sliding door stock, it is typically approximately 35 per cent of the number of standard class seats. For most train operators the standing allowance is based on an allowance of 0.45m<sup>2</sup> of floor space per passenger. However, for South West Trains' commuter rolling stock a figure of 0.25m<sup>2</sup> is used, and for Southeastern's class 376 'metro' style stock and for London Overground a figure of 0.35m<sup>2</sup> is used. In some cases train operators do not have standing capacities calculated for their rolling stock based on the available floor area. In these cases the standing capacities have been estimated as 20 per cent of the number of standard class seats for long distance rolling stock, and 35 per cent of the number of standard class seats for commuter rolling stock."*

The DfT collect passenger counts from the train operators once per year in the autumn and in their annual report present summary statistics on weekday peak crowding and also passenger numbers. Historically, the counts were one-off manual counts, although increasingly many of the counts use automatic passenger counting (APC) techniques and so are collected on multiple days and averaged. The APC data is collected either from door sensors or weight-based sensors and this is discussed in further detail in §2.2.2.

The crowding statistics are for the peak periods for trains arriving in major cities in the AM peak (07:00 to 09:59) and departing from major cities in the PM peak (16:00 to 18:59). The crowding statistics are calculated for standard class passengers and are on the busiest point of each train service, i.e. the 'critical load point', as defined above. The main measure for crowding is called 'Passengers in excess of capacity' (PiXC) and is defined as follows:

*"PiXC is the number of standard class passengers on a service that are in excess of the standard class capacity at the critical load point. It is the difference between the standard class passenger load at the critical load point and the standard class capacity, or zero if the passenger load is within the capacity. Capacities include the number of standard class seats, and also include a standing allowance if the time between stations at the critical load point is 20 minutes or less. For each train*

---

<sup>3</sup> Critical load point := "the location where the passenger load on a service is highest on arrival at (AM peak) or on departure from (PM peak) a city"

*operator the numbers of passengers in excess of capacity on each service are aggregated together and expressed as a percentage of the total standard class critical load.”*

In previous years, PiXC was used as the crowding performance target for train operators, with a maximum allowable PiXC of 4.5% in one peak (either morning or afternoon) and 3% across both peaks. However, the PiXC is no longer used as a target, but it is still annually published to give a consistent measure that allows crowding to be compared between different routes and over time.

One limitation of the PiXC measure is that it is based on average passenger loads over the peak period and aggregated across different services for particular train operators so does not capture variations between services. The disaggregated data is commercially sensitive and so is not published, although the DfT does publish a biannual briefing note called the “top ten overcrowded services” (DfT 2013b). This defines another measure for crowding, called the “standard class load factor”, which is defined as:

*“The number of standard class passengers expressed as a percentage of the maximum allowable standard class passenger capacity for that service. For example, a train which has the same passenger load as the passenger capacity would have a load factor of 100%.”*

This publication lists the ten services with the highest standard class load factor, which for Spring 2012 ranged from 149% to 184%.

Another limitation is that the crowding statistics are based on the capacity and loading for the whole train and so coach-by-coach variations are not taken into account. This is suggested in the methodology report (DfT 2013a), “For example, at major terminals passenger numbers are often higher at the end of a train that is closest to the entrance/exit on the platform, meaning that passengers travelling at one end of a train can perceive a higher level of crowding than those at the other end, and that passengers can be standing in one carriage when there are empty seats in another”. Research on the extent to which the coach-by-coach occupancy is uneven is explored further in §2.2.4.

In an international literature review (Charles and Hale 2009), researchers from the University of Queensland introduced other performance indicators, which included:



“standing time duration and ratios of passenger-km to seat-km”; and “pass-ups, i.e. the regularity at which passengers are unable to board due to overcrowding”.

There is limited information on station platform capacity, although in a FOI request (Smith 2011), it was explained that for London stations there is no official maximum platform capacity, rather the “station staff make their judgements based on the density of people observed in specific areas of the station”.

### Summary

The DfT conducts an annual survey of rail passenger crowding, which is increasingly based on passenger counts from automated systems. On-train standard class capacity is defined as the number of standard class seats plus a ‘standing allowance’, which is different for different types of service and rolling stock. The DfT uses two measures of rail crowding: ‘passengers in excess of capacity’ (PiXC) and ‘standard class load factor’.

The DfT acknowledge limitations of these measures, namely that they do not reflect coach-by-coach variations, nor do they reflect variations across the peak period.

### 2.2.2 Automatic passenger counting (APC) data and rail passenger crowding

The supporting notes to the DfT rail crowding statistics report explain that there are two main types of automatic passenger counting (APC) systems in the UK (DfT 2013a):

- ‘Load weighing’ – this is equipment fitted to trains that ‘weighs’ the train at certain points, estimating the number of passengers on board by assuming an average weight per passenger.
- ‘Infra-red’ – this uses infra-red sensors fitted around each door on the train to count the numbers of passengers boarding and alighting at each station.

Some door sensors also use video image processing or lasers instead of or in addition to infra-red technology. Train operators may have a variety of uses for this data, such as planning timetables and how rolling stock is deployed, informing ticket pricing and marketing campaigns. Due to the commercial nature of this data it is not published at disaggregated level, although some train operators publish typical information about



crowding on individual train services to passengers, either online or through posters at stations (ORR 2012).

The DfT is in the process of procuring a new database of all UK APC data for use by themselves and the train operators (Capita Symonds 2012), which will “provide an important insight into the usage and volumes of passenger traffic across the rail network [and] identify ‘hotspots’ and trends”.

The existing deployment of APC on UK trains varies by train operator. A report by the Public Accounts Committee states that the DfT requires all new trains to have it, and that some train operators are retro-fitting APC to old trains (PAC 2010). As of Nov 2010, it was estimated that 39% of carriages were fitted with some type of APC equipment.

The accuracy of APC door sensors is often quoted by manufacturers as being at least 95% (TRB 2008). The accuracy of weight based sensors may well be less than this, because assumptions are required for the average weight of passengers. Another advantage of door sensors is that they are able to provide bi-directional boarding and alighting flows, whereas weight-based sensors cannot. However, door sensors may be susceptible to ‘drift’ in the running total on-board, e.g. this may occur in the event that the alighting count is accurate, but the boarding count is over-counting. Also, to provide accurate passenger counts for each coach, door sensors are required between carriages.

The literature review did not find any academic research on the use of APC data in rail in the UK. This may be because the data is owned by the train operators and is commercially sensitive. A small amount of academic research was found on the use of APC data for other modes of transport and in other countries.

Rajbhandari et al. (2003) conducted research into inter-city bus dwell times using APC data. They suggested that factors that may affect the boarding time included: the number of boarding passengers; the number of alighting passengers; and the number of passengers standing. They developed linear and non-linear regression models to describe the relationship between dwell times and these factors. They found that the non-linear model, which included the number of standing passengers, gave the best fit suggesting that boarding times were greater when there was crowding on-board. They also found that the dwell time was not affected by time of day or service type. It is interesting that Rajbhandari et al. were able to use APC data to establish a link

between crowding and boarding times; however, there is only limited applicability of the results to rail, because there are additional factors, such as payment to the driver for boarding passengers. Nevertheless, it suggests that it may be possible to use APC data to investigate the relationship between crowding and dwell times in rail. Other examples of research into the link between crowding and dwell times using other sources of data will be discussed in §2.2.5.

The TRB (2008) produced a synthesis of information and advice for US public transport agencies on the use of APC. However, it primarily is based on case studies of APC used in buses and in some cases light rail.

Other technologies, in particular mobile phones, have also been used for monitoring pedestrian flows more generally (see §2.1.2). However, these technologies are not as widely used as door sensors or weight-based sensors by train operators for passenger counting because of lower accuracy.

### **Summary**

The two main types of APC sensors used in the UK rail industry are door sensors and weight sensors, with around 40% of the fleet equipped as of 2010. An advantage of door sensors is that they are typically more accurate than weight-based sensors, because assumptions are not required around the average weight of passengers. Also door sensors are able to provide bi-directional boarding and alighting flows, whereas weight sensors cannot. However, door sensors may be susceptible to 'drift' in the running total on-board and also to provide accurate passenger counts for each coach, additional door sensors are required between carriages.

The literature review did not find any academic research on the use of APC data in rail in the UK, which may be because the data is owned by the train operators and is commercially sensitive. This perhaps suggests that there is a gap for novel research if it were possible to gain access to this data source. Some academic research was found into the use of APC data for estimating dwell times of buses, although this is not directly applicable, because there are other factors not present in rail, such as payment to the driver for boarding passengers.

### **2.2.3 Automatic fare collection (AFC) data and rail passenger crowding**

Automatic fare collection (AFC) systems in public transport have been deployed in many countries around the world using smart card technology. Rather than using traditional paper tickets, public transport passengers can purchase tickets without queuing and can complete their journey using a single smart card, phone or contactless bank card. Smart card schemes are well established in many cities and regions internationally, with examples such as the Oyster card in London and the Octopus card in Hong Kong. In some cases the area covered by the smart cards is regional or country-wide, although more commonly the area covered by smart cards is city-wide. In the UK there are plans for greater deployment of smart card schemes across the country (Borg 2014) and a pilot project in Norfolk is ongoing, which will evaluate the effectiveness of a 'managed service' approach in deploying smart and integrated ticketing across a rural county (DfT 2013c).

There has been much research using the data generated from the user 'touch-ins' and 'touch-outs' across a wide range of situations. Pelletier et al. (2011) gives a good overview of strategic, tactical and operational uses of AFC data, as well as considering the research challenges. Some more specific examples of research using AFC data include the following.

- Personalised recommendation systems – Researchers at UCL have proposed approaches for developing personalised systems using AFC data, in particular for providing personalised trip time estimates (Lathia et al. 2013, Lathia et al. 2010), as well as giving recommendations for the type of ticket users should purchase (Lathia and Capra 2011).
- Estimating occupancy on buses – Oyster card data is increasingly being used to measure occupancy and crowding on buses in London instead of traditional manual surveys; this is not a trivial task because data is only collected when entering a bus (Wang et al. 2011).
- Measuring public transport accessibility – Ferrari et al. (2013) built a model from AFC data and a journey planner to evaluate the level of accessibility for various trips across London.

These studies illustrate that AFC data is an exciting area of research with potential to be applied to a variety of situations. The rest of this section focuses on a smaller selection of research studies, in which AFC data is used to investigate crowding.

### 2.2.3.1 Predicting crowding using AFC data

Ceapa et al. (2012) undertook analysis of peak crowding at stations on the London Underground using one month of Oyster card entrance and exit data.

After cleaning the data, they conducted a spatio-temporal analysis, which revealed that the station crowding patterns were very regular during the week due to commuting, but typically had higher variance at the weekends. For the weekday data, they identified at least three types of station: ‘Residential’ with high entrances in the AM Peak and high exits in the PM Peak e.g. Finchley Central; ‘Business’ with high exits in the AM Peak and high entrances in the PM Peak e.g. Canary Wharf; and ‘Transport Hub’ with both high entrances and high exits in both the AM Peak and PM Peak e.g. Waterloo. Figure 1 shows the average weekday entrances (blue) and exits (red) at Canary Wharf Underground station, which showed very low standard deviation (shaded area) in the evening peak, along with three ‘sub-peaks’ around 17:10, 17:40 and 18:10.



**Figure 1 – Weekday entrances (red) and exits (blue) at Canary Wharf Underground station (reproduced from Ceapa et al. 2012; © 2012 ACM, Inc. Included here by permission)**

They then undertook a more systematic classification of the stations using an agglomerative hierarchical clustering technique, i.e. this is a ‘bottom up’ approach in which each observation starts in its own cluster and then pairs of clusters are merged together. They used a technique related to ‘Dynamic Time Warping’ (DTW), which is a popular algorithm used in data mining and time series clustering analysis, as well as in other fields. Senin (2008) describes the DTW algorithm as “being extremely efficient as the time-series similarity measure which minimises the effects of shifting and distortion in time by allowing elastic transformation of time series in order to detect similar shapes with different phases”. Specifically Ceapa et al. chose to use an approximation to DTW called ‘FastDTW’; this had the advantage of being more suitable for large time series datasets. From visual inspection, they chose to terminate the clustering algorithm at six clusters. The results of the clustering gave small intra-cluster distances suggesting that

there was good compactness within each cluster; there was also large inter-cluster distances suggesting good separation between the clusters.

As there was no official definition for station crowding, they defined a proxy measure for crowding, which was the proportion of touch-ins and touch-outs at the station relative to the maximum number observed in the data. This measure meant that a value of 1 represented the station at its peak level of crowding and a value of 0 indicated no entrances or exits within the period.

They then built and evaluated three prediction models and investigated the effect of several parameters on the accuracy of the results, using half of the data for training the model and half for testing. The three predictors were as follows:

- 'Historic value' – This was the most basic predictor, taking the one corresponding value in the training dataset for the corresponding day of week and time of day.
- 'Historic mean' – This was similar to the 'Historic value' predictor, but used the mean of all values in the training dataset for the corresponding day of week and time of day.
- 'Historic trend' – This attempted to improve on the other two predictors by taking into consideration crowding level for the current time.

The project concluded that the high regularity of commuting travel meant that patterns are indeed predictable, even with as little as two weeks of training data. The three predictors based on historic data achieved very good results in predicting crowding, which suggested that providing information on crowding levels to passengers would be “highly feasible”.

They proposed one use of this analysis as follows: “Identifying times when public transport is overcrowded could help travellers change their travel patterns, by either travelling slightly earlier or later, or by travelling from/to a different but geographically close station”. However, the study did not assess how passengers would make use of such information, but recommend this for future studies.

The finding that weekday travel on the London Underground has low variability is not surprising; for the most part this is likely explained by high levels of commuting trips, many of which will involve interchanges with timetabled train services. Nevertheless, it is novel that even relatively simple predictors yielded accurate results due to this

regularity. A limitation of the study should be noted in that it did not consider any seasonal effects; the study was based on data for the 31 days in March 2010, which was a ‘typical’ month in that it contained no school holidays or bank holidays.

The fundamental approach of station-by-station analysis, clustering and then prediction may be applicable to heavy rail, although it is not clear whether journeys would also have low levels of variability and in turn high levels of predictability. This would likely be affected by the proportion of passengers on a particular route who were travelling for commuting, business, leisure or other journey purposes.

As discussed above, the study used the proportion of touch-ins and touch-outs relative to the maximum number observed in the data as a proxy measure for crowding. This measure was fit for purpose in the context of the study in assessing the predictability of demand, although a limitation of the study is that there were no attempts to calibrate this with available capacity, i.e. to understand actual levels of crowding. This could be attempted in a more detailed study, which might involve analysis of service timetables, walking times, predicted interchanges, proportion of tickets that are still paper-based, validation with manual counts etc. As such, it seems that it would be easier to use APC data if available rather than AFC data for measuring levels of crowding.

#### **2.2.3.2 Agent-based modelling using AFC data**

Legara et al. (2014) built an Agent-Based Model (ABM) of the light rail network in Singapore using AFC smart card data. Agents have been defined as “autonomous decision-making units with diverse characteristics” (Macal and North 2010). For example, this might be a model of a flock of birds, where each individual bird is considered an agent and their behaviours are modelled individually.

The AFC dataset covered one week and contained 14 million journeys with information on time and location of touch-ins and touch-outs, as well as with a journey ID and anonymised card ID. The model consisted of a network of 29 stations on one line, with train dispatch intervals being normally distributed with a three-minute headway during the peak and a six-minute headway at other times. They used the AFC data to validate their model through both visual inspection and statistical goodness of fit tests. They plotted output from the model on the travel time distributions between origin-destination station pairs and then also plotted real data against these curves.

The model contained information on the loadings and available capacity of all trains at a particular time. The main measure used in the model was the number of ‘pass-ups’,



i.e. where passengers are not able to board a train, because there is no available capacity. They ran several model scenarios where the number of travellers was increased to represent future scenarios with population growth. They also ran scenarios where they decreased the available capacity. They found that when decreasing the available capacity, there was a tipping point where the cumulative travel time increased exponentially. This was because the model reached a point where trains consistently departed at full capacity and the number of passengers who could not board increased throughout the peak period.

The focus of this study was in modelling future growth scenarios for a subway system with very high demand, where the main measure of crowding used was the number of 'pass-ups'. Although on occasion some heavy rail trains do depart with passengers not able to board, this is typically a rare occurrence in the UK and so is not an applicable metric. A modelling approach can be a useful method to test alternative scenarios, although such models for the UK railways are already well established as discussed in §1.1.

#### **2.2.3.3 Visualisation of passenger flows using AFC data**

Itoh et al. (n.d.) proposed two types of visualisation for occupancy on metros using two years of AFC data for Tokyo's subway, a 'Heat Map View' and a 'Route Map View'. The Heat Map View uses the horizontal axis for the time and the vertical axis for the different lines and stations. There is one cell for each 10 minutes and for each station in a particular direction. The colour of the cell represents the level of crowding, normalised by standard deviation; red represents occupancy on that day where it is higher than the historic average for that cell and blue represents occupancy that is lower than the historic average. The Route Map View illustrates the number of passengers by the height of stacked 3D bands along the route, but also uses colour to describe how the values at a particular time compare to the historic average in the same way as the Heat Map View. Figure 2 shows the two views in Tokyo on a date of an earthquake; this shows that at 14:00 the subway was running as usual, but soon after the earthquake at 14:46 flows were much lower than usual and then when the lines were reopened in the evening the flows were much higher than usual.

These visualisations are an interesting way to communicate operational information for multiple routes and services, specifically for how a particular day compares with 'normal conditions', i.e. in communicating the level of service disruption.

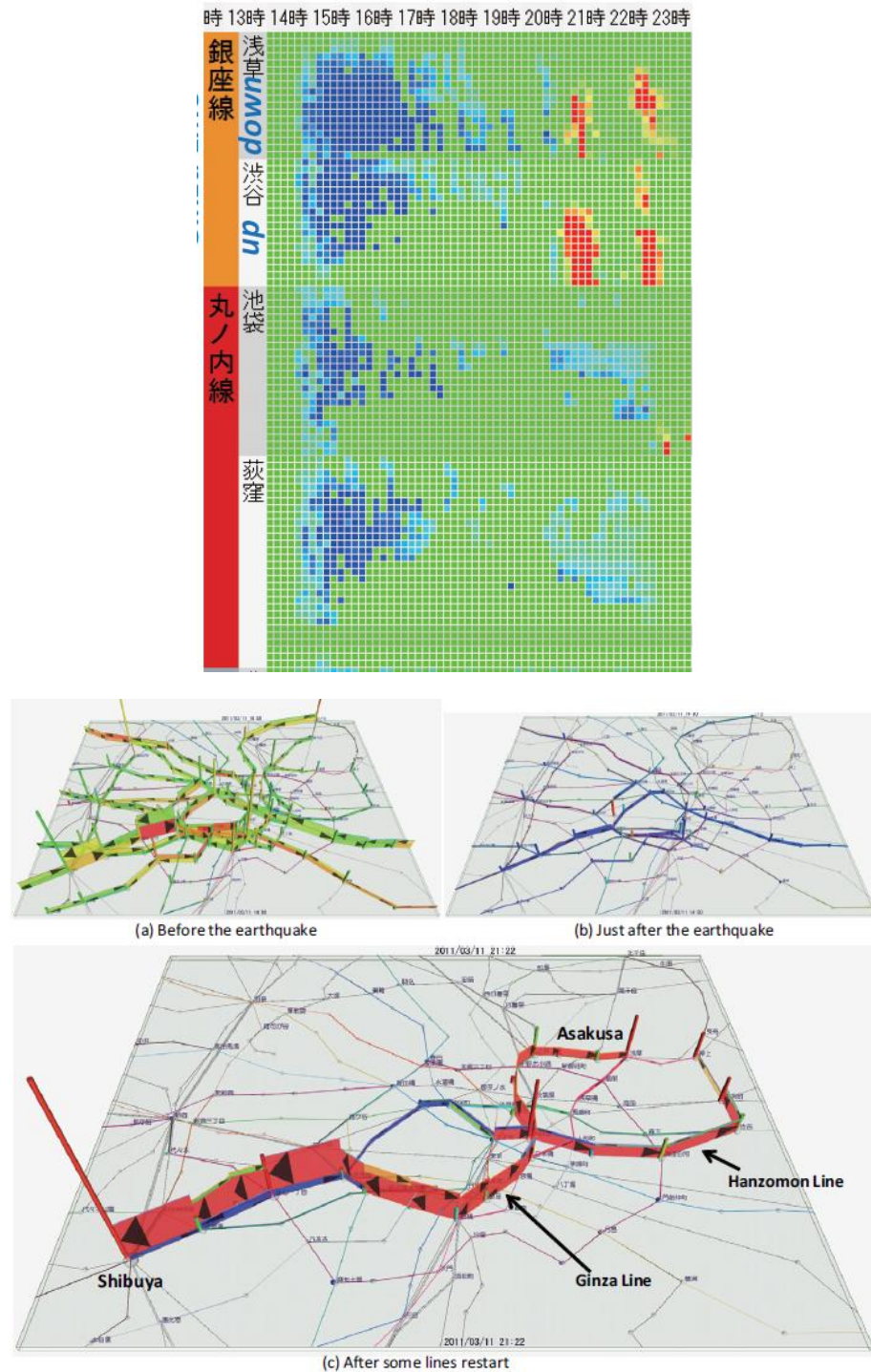


Figure 2 – Passenger flows on 11 March 2011, Tokyo, Heat Map View (left) and Route Map View (right) (reproduced from Itoh et al. n.d.)

## Summary

There is a substantial amount of academic research into AFC data across a range of topic areas. The literature review has focused on applications of AFC data to estimate levels of crowding, both on stations and on trains. The structure of AFC data is typically that of 'touch-ins' and 'touch-outs' and so does not provide a direct link to crowding and



occupancy data, unlike APC data. Nevertheless, there are a range of approaches that may be applicable.

Ceapa et al. (2012) used the approach of a preliminary spatio-temporal visual analysis, followed by a more rigorous clustering approach to classify the stations, followed by use of predictive techniques. They used an approximation to the DTW clustering algorithm to classify the time series data. They showed that if the variability is low, then simple predictors can give quite accurate results, even with a relatively small amount of data.

The modelling approach used by Legara et al. (2014) focused mainly on the number of 'pass-ups' as the main measure of crowding. This may be applicable to a very busy light rail scenario, but is less applicable to heavy rail.

The review identified some novel visualisation techniques that may be applicable, although such visualisations should be in addition to more rigorous techniques.

## **2.2.4 Localised crowding and 'loading diversity'**

As introduced in §2.2.1, uneven distribution of passengers between coaches can lead to localised crowding on a train when seats are still available, meaning that passengers travelling at one end of a train can perceive a higher level of crowding than those at the other end of the train. This section explores research into this localised crowding or 'loading diversity'.

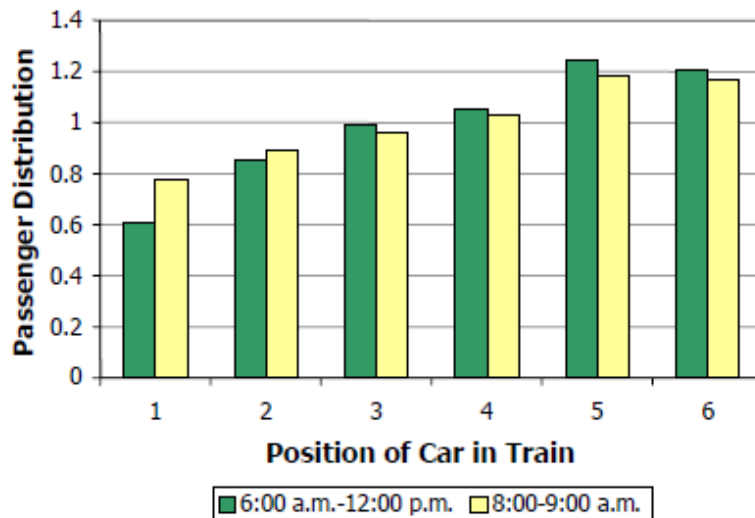
### **2.2.4.1 Definition of 'loading diversity'**

The TRB (2003) issued guidance that introduced the concept of 'loading diversity', of which they defined three types: "1. Loading diversity within a car<sup>4</sup>"; "2. Loading diversity among cars of a train"; and "3. Unevenness of passenger demand during the peak hour". They described the first type of loading diversity as being the situation in which "the highest standing densities occur around doorways" and suggested this can be reduced by increasing the number of doors. For the second type of loading diversity, where there are more passengers in some coaches than other coaches, they defined a

---

<sup>4</sup> The terminology 'cars' is used in the TRB report, which is equivalent to the word 'coaches' used in the thesis.

measure that they call ‘average peak hour passenger distribution between cars of trains’, or perhaps more usefully ‘ratio of car occupancy to train average’. This is a measure where a value equal to 1 represents an individual coach load equal to the average load of all cars in the train. An example for one platform on Toronto’s subway is reproduced in Figure 3; this was aggregated across 99 trains in a six-hour period on one day.



**Figure 3 – ‘Ratio of car occupancy to train average’: Toronto, Yonge Subway, Wellesley Station, southbound direction, Jan 11th 1995, 99 trains, 66,263 passengers (reproduced from TRB 2003)**

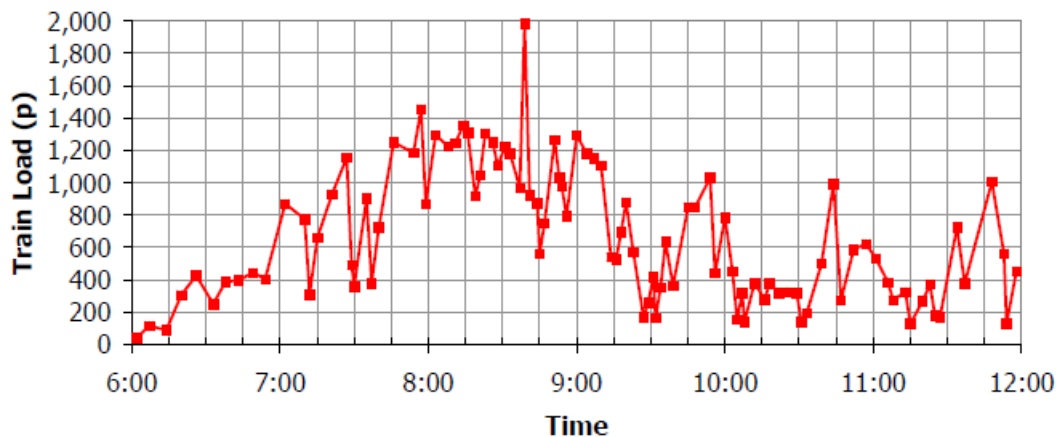
In this example, during the morning peak period, the rear of the train (Coaches 5 and 6) was typically more heavily loaded than the front of the train; this ranged from 1.26 (+26%) in the busiest coach to 0.61 (-39%) in the emptiest coach. They suggested that this could be explained by a major transfer station at the next station along (Bloor Street), which has the interchange at the rear end of the platform. In this example, there was less variation in the ratio of car occupancy to train average during the one-hour peak period (08:00 to 09:00) suggesting passengers were more likely to spread evenly along the full length of the train at the busiest times. For individual trains in the survey period, the ratio of car occupancy to train average ranged from 2.56 (+156%) to 0.11 (-89%). They suggested that the main cause of this second type of loading diversity is that “cars that are closer to station exits and entrances will be more heavily loaded than more remote cars”. Furthermore, they suggested that, “this inefficiency can be minimised by staggering platform entrances and exits between ends, centres, and third points of the platforms”.

They described the third type of loading diversity as being where passenger demand is spread unevenly over the peak hour period. The distribution of train loads for the

extended peak period is shown in Figure 4 for the same example as above. There was a spike at 08:35, illustrating that there can be high variability within the peak hour period. A measure was proposed to describe the ‘peak-within-the-peak’ that was called the ‘peak hour factor’, defined as:

$$\text{“Peak hour factor} = (\text{passenger volume in peak hour}) / (4 * \text{passenger volume in peak 15 mins})\text{”}$$

The peak hour factor gives a value between 0.25 and 1, where a higher value means that the passenger demand is more evenly spread over the peak hour.



**Figure 4 – Train load across the extended AM peak period: Toronto, Yonge Subway, Wellesley Station, southbound direction, Jan 11th 1995 (reproduced from TRB 2003)**

The three-part definition of ‘loading diversity’ provides a useful starting point to consider the issue of uneven occupancy on trains. The first type was considered not of interest here in that it is largely affected by carriage design, which was not the primary focus of this research. The second type of ‘inter-coach’ loading diversity was considered to be of interest, although the current TRB metric (‘ratio of car occupancy to train average’) does have a limitation in that it only explains the loading relative to other coaches and does not include information on the capacity; for example some quiet services would be classified as having very high loading diversity, even though there may be no coaches at or near capacity. The third type of ‘inter-train’ loading diversity was also of interest, although the ‘peak hour factor’ measure may not be appropriate for use in heavy rail because of larger headways between trains. As these loading diversity measures are not readily applicable, there is a gap to explore alternative measures, using these as a starting point. It may be possible to combine the measures proposed by the TRB with existing DfT measures, such as ‘standard class load factor’ or PiXC (see §2.2.1) on a coach-by-coach basis.

The TRB report suggested that loading diversity could be addressed through measures such as staggering entrances and exits along the platform and providing more doors on trains. These approaches would be relatively high-cost or would need to be considered initially when designing the station. A review of possible measures to address inter-coach and inter-train loading diversity will be given in §2.3.

#### **2.2.4.2 Research studies into different aspects of loading diversity**

Kim et al. (2014) conducted a study to investigate the reasons why passengers choose a particular coach on the Seoul subway. Thus they attempted to explain the underlying causes for unevenness of passenger loads across individual coaches of a train, i.e. the second type of loading diversity as defined in §2.2.4.1. They conducted 340 face-to-face interviews at one station on the Seoul metro during the weekday morning peak over a period of four weeks. The surveyors randomly picked passengers who were waiting for their train to arrive. The survey involved two main questions: whether or not they chose a specific coach intentionally; and if so what was the motivation for their choice. Before conducting the survey, they determined four groups of variables to correlate against their responses:

- Individual-specific characteristics – age, gender, marital status, income and other socio-economic factors
- Trip-related variables – variables related to a respondent's current trip, such as trip purpose, trip frequency, prior travel experience, awareness of station layout etc
- Physical environment around the platform contains – entrances/exits, transfer gates, elevators, and escalators.
- Attitudinal or behavioural propensities – good memory, health condition, punctuality etc.

The headline results from the survey were that about three-quarters of the respondents reported choosing a specific coach intentionally. Of these when asked to explain their motivation: 70% said “to minimise the walking distance to exit when they disembarked at a destination station”; 17% said “to minimise the distance from the entrance when they boarded at an origin station”; and the remaining 13% said “to pursue comfort while travelling”. They constructed a statistical model to determine which variables explained the answers.

A relationship of note was that young females were more likely to choose a coach to avoid high levels of crowding. Another was that commuters were more likely to minimise their walking distance at the destination station, as were passengers who were classified to have better mnemonic ability, i.e. those with better memory. The survey was conducted for a subway during the weekday morning peak and so it would be expected that there would be a high proportion of commuters. Given that it was found that commuters were more likely to minimise walking distance on alighting the train, this combined with the make-up of the sample would go some way to explaining the 70% who said this was their main motivation for choosing their coach. These results are likely applicable to situations with high proportions of commuters, but perhaps less so where there are low proportions of commuters. The passenger survey approach used in this study could also be applied to heavy rail in determining the motivations of passengers in their choice of coach.

Sohn (2013) built a model to determine the optimal stopping position for trains for a hypothetical metro line, in order to make the passenger load more evenly dispersed. He did this by applying a genetic algorithm to solve the proposed model, which “considerably improved the distribution of passenger loading”. Such research is useful if designing and building new stations, but is of limited applicability to existing stations.

Lee et al. (2012) conducted a simulation study for the Seoul Subway into the benefits of providing real-time coach-by-coach congestion information to passengers waiting for the approaching train on the platform. The simulation was for the most crowded section of Line 4 between Danggogae and Chungmuro stations. There is only limited published information on the simulation, although it was concluded that dwell time could be reduced by 15% if this were implemented. Modelling the impacts of coach-by-coach congestion information could be a successful approach for estimating the benefits of such systems. The 15% reduction in dwell time suggested by the modelling would be a substantial time saving; however, there is only limited information on the scenarios and assumptions made and so these results should be treated with caution.

Wiggenraad (2001) conducted observational studies at seven Dutch stations, which included observations of where passengers waited and also where they boarded. The study involved around 20 surveyors covering the full length of the platforms with clipboards, for a period of four hours at each station (07:00-09:00 and 10:00-12:00). This yielded a total of 130 observations of train services. The study focused predominantly on boarding and dwell times, but also involved summary analysis of the distribution of passengers along the platform. A key finding was that at some stations

although there was a concentration of passengers waiting close to the stairs, upon the arrival of the train the passengers distributed themselves more evenly over the platform for boarding.

The traditional approach of using surveyors to monitor the boarding location of passengers would still be applicable today, although is a resource-intensive method. If door sensors were already fitted on the trains (see §2.2.2), it would be possible to analyse the proportion of passengers boarding at each door, which would be a substantially more efficient approach. Another alternative would be to conduct a similar analysis through using video surveys to record behaviour along the platform; this would have the advantage of being able to 'fast forward' to the arrival times of the trains and also to assess the distribution of where passengers waited.

Researchers from UCL (Fujiyama 2014) are also currently working in the area of loading diversity. An employee of London Underground, David Dobson, has recently started a part-time PhD at UCL conducting investigations into more even loadings between coaches on the Underground, although the research is still at an early stage. Taku Fujiyama from UCL also has an interest in this area from the perspective of boarding times and dwell times. Researchers in South Korea and the USA are also working in this area.

### Summary

The TRB have proposed a three-part definition of 'loading diversity' to describe the distribution of passengers: i) near doors within a coach; ii) from coach to coach; iii) between trains across the peak period. The metrics associated with these definitions are not readily applicable to heavy rail, although they do provide a useful starting point to explore alternative measures of uneven occupancy on trains. The first type was considered not of interest here in that it is largely affected by carriage design, which was not the primary focus of this research. The second type of loading diversity was considered to be of interest, although the 'ratio of car occupancy to train average' is only a relative measure and does not include information on the capacity. The third type of loading diversity was also of interest, although the 'peak hour factor' measure may not be appropriate for use in heavy rail because of larger headways between trains.

The review identified only a few research papers in the area of loading diversity. Kim et al. (2014) found that about three-quarters of subway users in South Korea reported choosing a specific coach intentionally and of these when asked to explain their motivation there were a variety of reasons, although 70% said “to minimise the walking distance to exit when they disembarked at a destination station”. This paper goes some way to investigate the causes of uneven loadings, but it should be noted that the survey was conducted for a subway during the weekday morning peak and so may not be representative of other scenarios.

### **2.2.5 Relationship between crowding and boarding times**

It may be expected that the extent to which passengers are spread unevenly between coaches would be related to boarding times. This section briefly explores existing research into train boarding times; this includes both real-world observations and laboratory experiments.

Wiggenraad (2001) conducted observations at Dutch railway stations and found that during the first phase of boarding where passengers are crowded at the door the mean alighting and boarding time per passenger was around one second. He found that for wider doors there were 10% shorter time values and for narrower doors there were 10% longer time values.

Lee et al. (2007) conducted a review of existing research into boarding times. They suggested that there are four types of factors that affect train boarding times: passenger mobility; platform design; vehicle design; and crowding effects. They also proposed a model based on empirical data to describe the non-linear relationship between boarding times and the level of crowding on the train.

UCL constructed a mock-up of a train carriage in their pedestrian laboratory environment and conducted controlled experiments, whereby participants board and alight the carriage under different conditions. They undertook a study for the design of the Thameslink carriages (Fujiyama et al. 2008), which tested different passenger flows for a range of design variables, such as door width, platform height and platform markings. Regarding the relationship between crowding in the vestibule and passenger flow rate, they found that “when density of the destination (e.g. the vestibule for the



boarding flow) becomes two people per square metre or greater, the flow rate starts decreasing”.

There are several existing research studies that have investigated the factors affecting boarding times through both real-world and laboratory experiments; a range of results have been collected for the average boarding and alighting times per passenger under various conditions. The only study identified that used APC data to investigate boarding times was Rajbhandari et al. (2003), as discussed in §2.2.2. There is potential for more widespread use of APC (door sensor) data in this way, which would have advantages of larger sample sizes and also cheaper data collection if sensors were already installed. The studies identified focused on average results for one set of doors, rather than taking into account the maximum boarding time across all doors.

The relationship between loading diversity and boarding times is likely two-fold. Firstly, the boarding time for a whole train is equal to the maximum boarding time of each individual door, thus if a disproportionate number of passengers board and alight from one particular door this will result in higher overall boarding time. Secondly, there is research from multiple studies to suggest that once the vestibule reaches a certain threshold of crowding, the flow rate at which passengers can board reduces. Thus it is intuitive that optimising boarding patterns along the platform will result in lower boarding times, although this relationship is likely non-linear if there is an interactive effect of crowding in vestibules.

Dwell time is the time that train is at the station; one definition is the time spent at the station between the wheels stopping and starting, while an alternative definition is the time between the doors unlocking and locking. Dwell time is closely related to boarding time, in that boarding time makes up a large proportion of the dwell time in many cases. Other aspects of dwell time include waiting for the scheduled departure time if trains are running early or if there is contingency in the timetable.

Boarding times and in turn dwell times are one source of delay for smooth operation of trains. There are several punctuality targets for train operators and Network Rail, although the main target is the ‘public performance measure’ (PPM), which is the percentage of trains that arrive at their terminating station on time, i.e. within five minutes of the planned arrival time for London or regional services, or within ten minutes for long distance services (Network Rail n.d.).



## Summary

There are several existing research studies that have investigated the factors affecting boarding times through both real-world and laboratory experiments. The mean alighting and boarding time per passenger is typically around one second, although this increases or decreases depending on various conditions, such as passenger mobility, platform design, vehicle design and crowding effects.

The relationship between loading diversity and boarding times is likely two-fold. Firstly, the boarding time for a whole train is equal to the maximum boarding time of each individual door, thus if a disproportionate number of passengers board and alight from one particular door this will result in higher overall boarding time. Secondly, there is research from multiple studies to suggest that once the vestibule reaches a certain threshold of crowding, the flow rate at which passengers can board reduces. Thus it is intuitive that optimising boarding patterns along the platform will result in lower boarding times, although this relationship is likely non-linear if there is an interactive effect of crowding in vestibules.

## 2.3 Review of existing research into solutions to optimise loading diversity

This section gives an overview of research into solutions that are designed to optimise loading diversity. The review draws upon both academic and other sources and covers initiatives from concepts to pilots to implemented solutions. This involves approaches that affect passengers both consciously and subconsciously, with both real-time and static solutions. All solutions considered in some way use the on-board occupancy data that is captured automatically. For the purposes of making comparisons, initiatives have been grouped into two categories depending upon their intended effect:

- **Inter-coach:** these are intended to influence how passengers distribute themselves between coaches along a train, i.e. to optimise the second type of loading diversity as defined in §2.2.4.1.
- **Inter-train:** these are intended to influence the choice of train passengers make so as to encourage a move from crowded trains to earlier or later trains where more space is available, i.e. to optimise the third type of loading diversity as defined in §2.2.4.1.

The purpose of this review of solutions is to understand what concepts have been proposed or piloted and to explore the extent to which there exists evaluation evidence on the impact of these interventions. In turn this will highlight the gaps in the knowledge.

### **2.3.1 Inter-coach, real-time information**

Solutions in this category provide real-time information to inform passengers waiting on the platform of congested zones on the approaching train. The intention is to allow people to stand at appropriate locations along the platform so that they can board coaches with available space and avoid coaches that are full. Several examples were identified in the review:

- There has been a large-scale pilot in the Netherlands of a system that provides real-time information on coach-by-coach occupancy via an app and on-platform displays (§2.3.1.1).
- A simulation by the Korean Transport Institute concluded that real-time coach-by-coach occupancy information could reduce dwell times by 15% (§2.2.4.2).
- London Underground have completed the first phase of a research study into 'crowding indicators' and plan to conduct three demonstrators (§2.3.1.2).
- The new Siemens Desiro City trains to be used on Thameslink is planned to provide platform displays on passenger loadings (§2.3.1.2).
- A number of other conceptual proposals have been identified which demonstrate different approaches that could be used for displaying coach-by-coach occupancy information to passengers (§2.3.1.2).

#### **2.3.1.1 Pilots of automated real-time passenger information systems using on-board automatic passenger counts**

This project was undertaken by the consortium of NS, ProRail, STBY and Edenspiekermann in the Netherlands. The project revolved around the development of a real-time passenger information application system providing users with information on on-board real-time coach-by-coach occupancy. The information was made available to the user in the form of a mobile app as well as a digital display on the station platform.

Phase 1 was a trial of an automatic system that used on-board passenger counters to generate real-time occupancy data that were used to update a smartphone application with information to passengers on where to board (NS 2013). The trial was undertaken by NS, the long-distance train operator in the Netherlands. NS considered different technologies for the passenger counts, such as measuring the weight of the train, Wi-Fi usage, electronic checking data or monitoring seats, although they eventually chose to use door sensors to generate counts at the compartment level. Before the complete roll-out of the test, cameras were installed in the prototype train to check the accuracy of the sensors.

The APC test took place on the line from Roosendaal to Zwolle; 11 trains were equipped with over 270 counting sensors in total. The test involved two user groups, a group of NS employees and a group of 800 passengers, who both had access to the specially designed “iNStApp”, see Figure 5. Boarding and alighting passengers were counted, after which the occupancy of the train was calculated and classified as Red, Amber or Green. Count data were transmitted wirelessly to a central server, along with other data about the train, where they were then used to provide real-time information via a smartphone app about the passenger loading along the train to passengers waiting at stations for the train to arrive. The real-time information was updated around two minutes after the train left a station. Other information supplied by the app included the location of the 1st and 2nd class coaches, quiet coaches, entrances for bicycles and wheelchairs and train Wi-Fi information.



Figure 5 – Pilot of real-time smartphone app (reproduced from NS 2013)

The benefits of the system were stated as being that “passengers with a seat and a less crowded carriage are more likely to have an enjoyable journey and feel better about the service”. In an online interview, a spokesperson from NS spoke about feedback from passengers (Upton 2014): “We are very satisfied about the automatic passenger counting test. We notice that our customers greatly appreciate being able to see where there are seats available on the train and have a more comfortable journey as a result. Additionally, we can make great use of the occupancy figures for other applications within our company [such as the planning of the timetable and disruption management], making it a contributor to innovation within NS.”

This initial smartphone pilot by NS was novel in that it demonstrated that the technologies do exist to be able to build such a system and the scale of the pilot seems to have been sufficient to successfully demonstrate the technologies. Infra-red sensors were installed on all internal and external doors, partitioning each coach into two or three compartments; this level of granularity is higher than some other studies, which have proposed only coach-by-coach granularity. The approach taken by NS would likely be a more expensive implementation because more sensors would be required. Only limited information has been published on the results of the pilot and so it is difficult to draw conclusions on its success. It might be expected that using a smartphone app to display the information may only have limited take-up; in the subsequent phases of the pilot other ways to display the information were investigated.

In Phase 2 of the research programme there were observational studies at platforms, qualitative research with passengers and preparatory work for further pilots of real-time information systems. This work was commissioned by ProRail, the company responsible for the rail network and the management of train stations in the Netherlands, and was delivered by a research organisation called STBY and a design company called Edenspiekermann. There were three sub-projects (Enninga et al. 2013):

- ‘What motivates passengers on the platform’
- ‘Finding your way around the platforms’
- ‘Improving boarding and alighting’

The first project focused on how and why passengers move around platforms. STBY mapped the pedestrian flows on platforms through observations and shadowing; this was then supplemented by interviews with travellers to help understand the motives

behind the flows. The project concluded that there are three distinct categories of travel that determine a passenger's informational needs according to whether they are involved in routine travel, in incidental travel, or their travel experience has been disrupted. The second project was a more in-depth study involving travel diaries and interviews with train travellers. For about three weeks, travellers were asked to keep track of their train journeys using a specially designed diary. They answered questions about their trip, made sketches to indicate their movements and indicated their level of satisfaction at various stages. Each participant was then interviewed about their answers. There were "co-creation" workshops to define the main problems and identify possible improvements (Hummels 2013). Based on the customer insights, Edenspiekermann developed several possible design solutions to improve the boarding process. The design ideas were evaluated and improved on through further workshops with all relevant stakeholders. One concept was about providing train travellers with platform displays with information about the level of occupancy of the various train compartments. This concept overlapped with the work already done by NS on their real-time smartphone app and was taken forwards as a pilot in the next phase.

Phase 3 involved a pilot at Den Bosch station of a platform edge LED display system to provide a real-time display, using a Red-Amber-Green colour code, of where current seating availability was greatest, encouraging pedestrians to move along the platform, see Figure 6. The pilot was conducted jointly by the four organisations, drawing on experiences from the two earlier projects (Holthuis 2013).



Figure 6 – Pilot of real-time platform information system (reproduced from De Vos 2013)

Similar to the smartphone app that NS developed, the platform display provided train travellers with live information about the availability of seats in the train they were waiting for. The system architecture was essentially the same as the trial of the smartphone app: infra-red door sensors collected real-time occupancy data on the 11 trains that were part of the previous pilot and this was then transmitted wirelessly to the central server. The platform displays also gave other information on the actual composition of the train: the exact position of the train and train doors; position of the first and second class sections, as well as the special train functions like disabled ramps.

The pilot, which claimed to be the world's first real-time boarding information system of its kind, was operational for three months as a pilot at Den Bosch railway station from February to April in 2013. STBY analysed the usage of the app and the platform screen and also collected data on passenger satisfaction (De Vos 2013). During the test period many people used the new services, with about 700 travellers providing their feedback to STBY through online questionnaires and in-depth interviews.

Results from the pilot showed that, "The customer gave NS a significantly higher valuation at the end of the trial than before". Other benefits were that the passenger counts "provide the operator with an accurate, real-time display of how their trains are being used [and]... can help in time-table planning, improving how rolling stock is distributed, disruption management and how stations are planned". NS have said that they are looking at the costs of such a system in consideration for a wider roll-out. As of August 2015, the consortium was planning further trials.

The pilot of the real-time occupancy screens was successful in that it demonstrated such systems are technically feasible and there is some evidence to suggest that passengers valued the system. Emphasis was placed on user needs when designing the screens and the resulting 180-metre LED screens were quite ambitious. However, no published material was found with evaluation evidence of how passengers interacted with the system. The pilot could have been improved through quantitative monitoring of the behaviour of passengers, possibly using a 'Control' and 'Treatment' approach; i.e. monitoring the distribution of passengers along the platform (similar to Wiggendaad (2001) in §2.2.4.2) before it was installed, then again once the system was implemented. In particular, it would have been of interest to observe the level of compliance with the system, i.e. how many passengers stood in zones relative to the Red-Amber-Green colours displayed. Indeed, it is a possibility that too many people would stand where it was lit Green and this may in turn result in longer boarding times



and higher levels of crowding; for this reason more detailed evaluations should be conducted before such systems are implemented widely.

#### **2.3.1.2 Various – Vision concepts**

London Underground has recently announced that their Innovation Team has completed the initial phase of a project called, 'Crowding indicators research' (TfL 2015). Instrata, a digital SME from Cambridge, undertook the research, which included a literature review, passenger observations, a staff consultation at Victoria station and focus groups. A July 2015 London Underground newsletter said that, "A number of concepts were developed that aimed to make better use of data from trains to ease loading and reduce dwell times. The concepts focus on providing better information about frequency and crowding earlier in passengers' journeys (on a variety of displays) and then using platform digital projection and signage to display dynamic information about train crowding before trains enter the platform". The planned next phase of research will involve three demonstrators.

According to reports in the media the train operator in Stockholm, Stockholmståg, are planning to release an app that will "indicate which coaches are more or less crowded... [using] the wheel pressure in the trains" (Holland 2015). No further information was available on the initiative.

Siemens published a concept in which there was an indication on the platform of the location of the least crowded coaches (Siemens 2009). It is understood that a system like this may be installed for its Desiro City trains, which are being built for the Thameslink franchise. The promotional video does not describe how the information would be presented nor how the data would be collected, but it is understood that automatic passenger counters are to be fitted to all new rolling stock (see §2.2.2). A related application is also proposed for which the air conditioning system uses CO<sub>2</sub> sensors to match cooling to the passenger loadings.

A number of other conceptual proposals have been identified which demonstrate different approaches that could be used for displaying coach-by-coach occupancy information to passengers:

- A display for coach-by-coach occupancy for multiple approaching trains has been proposed for the London Underground (Taylor 2011). This would be displayed on the opposite wall to the platform on a large display and would have a traffic light colour code for each coach; it is proposed that this would

show multiple trains and be combined with information on arrival times and the current location of the trains on the network. It was suggested that loading information would be obtained from passenger body heat or carriage weight.

- The 'Comfort Zone Display' is a vision concept designed by 4-iD Design Network of Barcelona (Kumar 2010). The display boards at a station would show train passengers the coach-by-coach occupancy of the incoming train with images of people and a traffic light colour code. The displays would also show information on other services available in each coach.
- A start-up technology company in London has proposed a coach-by-coach occupancy display (Opencapacity n.d.). Their website describes a "content delivery product for public transport using the latest open source technologies combined with sensor innovations to create OpenCapacity – a novel, modular platform to measure, analyse and communicate the available space on public transport". It includes a concept for displaying the loading information in monochrome on existing LED CIS boards; passenger loading data would be obtained from any type of automatic sensors.

### **2.3.2 Inter-coach, near-real-time seat reservations**

Solutions in this category aim to use seat reservation information to help optimise coach-by-coach occupancy.

#### **2.3.2.1 'Seat Reservation Display System'**

DILAX, a provider of automatic passenger counting systems, has launched a new product for real-time seating information on trains. The DILAX 'Seat Reservation Display System' combines data from seat reservations and on board infra-red passenger counting systems (Dilax n.d.). They say that the "information can be shown on displays by the seat, as well as on the platform before boarding... another possible option is to offer a mobile app which enables passengers to quickly identify and locate their seats before and during boarding".

There is only limited information available and it is not clear exactly what the proposed product would do or how it would work. Nevertheless, there is potential for combining seat reservation information with APC data in some way, perhaps in enhancing real-time information systems as described in the previous section. Alternatively the



allocation of seat reservations to particular coaches could be optimised to assign certain Origin-Destination trips for advanced tickets to particular coaches. These could use historic APC data to train the forecasting algorithms on the most appropriate coach for each reservation. For such systems there would be no human factors considerations, because no decision would need to be made by the passenger.

### 2.3.3 Inter-coach, static information

Solutions in this category provide static information on inter-coach loading based on historical trends. Information would be static and so would not have the capability of adapting to specific circumstances in the way that dynamic measures may.

#### 2.3.3.1 Fixed signs pointing to less busy ends of the platform

Some London Underground stations have signs instructing passengers to walk to less busy ends of the platform. One such example is Platform 3 at Victoria London Underground station, which has a fixed sign with an arrow saying “Pass along the platform”, as shown in Figure 7.



**Figure 7 – Fixed signs pointing to less busy ends of the platform on the London Underground**

This approach may be applicable at some stations where there are repeated patterns in the distribution of boarding passengers along the platform. APC data could be used to investigate the variability of such patterns and to identify stations for which such

fixed measures may be appropriate. Ideally such analysis would be conducted on boarding flows, i.e. with door sensor data rather than weight data.

#### **2.3.3.2 Historic inter-coach crowding information**

Upon the award of the Essex Thameside franchise, a press release announced plans to, “Inform passengers on platforms which carriages on each service are least likely to be busy, initially using historical data on commuting patterns but with the aim of providing real-time information within five years” (Collins 2014). The train operator has already delivered an initiative for inter-train crowding information (see §2.3.6.3), although no further information on the inter-coach initiative has been identified.

### **2.3.4 Inter-coach, static passive behavioural modifiers**

Solutions in this category attempt to sub-consciously modify behaviours of travellers to induce more desirable positioning along the platform. Such measures would seek to perform a similar function to the above category, but rather than providing information, would seek to modify passenger behaviour through the use of ‘nudge’ tactics. For example, rather than inform people that it is better to stand at one end of the platform, a measure may simply encourage them to stand there by making it the easiest course of action. Several suggested examples were found from two sources:

- The Network Rail Stations RUS describes different methods that can be used. Example measures listed include moving information screens or removing obstacles to encourage people to congregate at a different point (§2.3.4.1).
- RSSB have produced guidance on management of rail crowding, which includes suggestions for influencing passenger positions along the platform. RSSB have also used pedestrian modelling to investigate the impacts of moving CIS displays (although on the main concourse rather than along platforms) (§2.3.4.2).

#### **2.3.4.1 Network Rail toolkit for platform overcrowding**

The Network Rail Route Utilisation Strategy (RUS) for Stations contains a toolkit to address station overcrowding (Network Rail 2011). It considers different ‘gaps’ and ‘options’ for three different station zones: ‘Access Zone’, ‘Facilities Zone’ and ‘Platform Zone’. Some options to address crowding in the Platform Zone are reproduced below.

**Table 1 – Platform gaps and options: extract from Table 5.8 of the Network Rail Stations RUS (reproduced from Network Rail 2011)**

Gap	Toolkit gap	Option	Toolkit option	Comments
PZ1	In inclement weather, passengers congregate under the canopy, causing congestion and increased station dwell time if all the train doors are not used.	PZ1.1	Extend the canopy further along the length of the platform, or provide shelters.	Current works at Clapham Junction are extending the platform canopies in order to encourage passengers to use the full length of the platform during inclement weather. Passengers currently congregate under the canopy and do not make full use of the platform length. This leads to bunching of passengers and slower boarding times potentially increasing train dwell times.
PZ2	Passengers may assemble near departure screens, especially during times of disruption.	PZ2.5	Consider installing further displays along the length of the platform.	The positioning of a CIS screen may cause passengers to congregate in one small area of a platform, as illustrated by the Littlehaven case study. The impact of this can be crowding in one area slowing boarding time and affecting train performance. Spreading passengers out along the platform has the potential to give more efficient boarding and reduce the impact on the train service by reducing dwell time.
		PZ2.6	Consider de-cluttering the platform or concourse to improve passenger circulation and allow the screens to be seen from further away.	In some instances other structures, equipment or signage may obscure the sight lines to the CIS. By decluttering the station CIS can be seen from further away reducing congestion.
PZ3	If unsure how many carriages are on the train, passengers may not use the appropriate length of the platform, causing bunching and potentially extending dwell times.	PZ3.2	Consider placing door markings with carriage numbers on the platform.	This is only likely to be feasible if identical rolling stock types are used on all services, and if the stop boards are positioned in such a way that different length trains will always have the doors in the same position. It also requires high levels of braking accuracy on the part of train drivers.
PZ4	Passengers requiring a specific part of the train (eg first class, reserved	PZ4.2	Signage on the platform to indicate where each carriage will stop.	While this information can be helpful, the signage can be misleading if the train is not in its usual formation.

Gap	Toolkit gap	Option	Toolkit option	Comments
	seat etc) may not know where to wait.			
PZ5	Arriving trains may already be crowded making it difficult for passengers to board.	PZ5.2	Use CIS to inform passengers where the less crowded part of the train is likely to be.	As information systems become more sophisticated it might be possible to provide real-time information from the on-train Automatic Passenger Counting system about where the busiest part of the train is.
		PZ5.6	Consider whether the train internal layout can be reconfigured to remove crowding from around the doors, by, for example, making it easier for passengers to stand in the body of the vehicle, or by widening the doors, or by increasing the vestibule area.	An example of modifications to rolling stock to increase its capacity without increasing its length is South West Trains refurbishment of Class 455s to augment the standing capacity, remove seating from around the doors and providing grab rails. These adaptations have the potential to reduce the time taken for passengers to board the train.
PZ7	Congestion at platform exit/entrances.	PZ7.2	If the width of the platform exacerbates crowding seek to widen the platform or by removing obstacles.	At Seven Sisters a structure on Platform 2 may be removed from to increase the width of platform space in the area of the entrance to the Underground which is a pinch point.
		PZ7.3	Provide new exit/entrance or enlarge existing entrance.	The Preston case study is an example where extra capacity is proposed at a platform entrance/exit in order to address congestion at this location in the station.

#### 2.3.4.2 RSSB research and guidance on rail crowding

RSSB conducted research into managing crowding on trains and in stations and produced guidance (RSSB 2009a), (RSSB 2009b). Section 3 of the guidance relates to crowd control and states:

*“There are a range of practical crowd measures you can use to manage your passengers effectively and deliver real benefits for on-train situations. These include better use of your available capacity, encouraging passengers to distribute to other services and helping them locate a suitable position on the platform.”*

The guidance then has a section on each of the following seven initiatives:

- Encouraging passengers to use other services
- Using passenger numbers and distributions
- Passenger positioning on platforms
- Declassifying first class
- Adjusting train stopping patterns and timings
- Implementing changes to stations and platforms
- Maintaining and increasing capacity

RSSB conducted a research study into station design and crowd management, which involved two main stages of work (RSSB 2012). Firstly, focus groups with passengers were undertaken to identify the user requirements that would lead to behaviour change. Secondly, six business solutions for releasing capacity were developed, which comprised three elements: technology feasibility; economic modelling of each of the solutions; and pedestrian modelling to demonstrate the impact of the solutions. The focus of the business solutions was primarily on ticketing, although one solution was to remove or relocate the CIS displays. This related to the main CIS displays on the concourse, rather than the displays along the platform and so is only marginally relevant here. Nevertheless, the pedestrian modelling showed that removal or relocation of the main CIS displays at three large stations would reduce the level of crowding on the concourse. This suggests a similar pedestrian modelling technique may be applicable to assess the impacts of modifications to CIS displays on the platform.

### **2.3.5 Inter-train, real-time information**

Measures in this category seek to encourage travellers to choose an earlier or later train that is less crowded, based on real-time data.

#### **2.3.5.1 Real-time occupancy information on buses**

The University of South Florida bus service provides real-time information on a website on bus movements and 'Percent Full' occupancy from automatic sensors, enabling

users to decide whether to take the next bus or wait for a less crowded one. The project was implemented by a company called Syncromatics.

### **2.3.6 Inter-train, static information**

Inter-train static measures similarly attempt to encourage travellers to use alternative services, but base this on long-term trends rather than real-time data. This would perhaps focus on times of peak and off-peak demand and would attempt to spread out the peaks and troughs. Several examples were found, both provision at the station through posters, as well as elsewhere through information provided at the point of sale for online tickets:

- The Swiss train operator, SBB, provides information on the expected level of crowding on specific trains at the point where passengers buy their tickets on their website (§2.3.6.1).
- South West Trains have carried out a pilot with ORR, using colour-coded posters showing which trains are typically the most crowded. Other train operators, such as London Midland have also done this, which has received praise from the DfT (§2.3.6.2).
- The C2C smartphone app provides information on the expected level of crowding on different trains in the peak period for some stations (§2.3.6.3).
- UCL have undertaken analysis of overcrowding on the Underground using Oyster card data. The project concluded that the high regularity of commuting travel means that patterns are predictable and that providing information on crowding levels to passengers would be "highly feasible" (§2.2.3.1).
- A number of projects have been identified that share user-generated (i.e. 'crowdsourced') data on the typical levels of overcrowding on particular services (§2.3.6.4).

#### **2.3.6.1 Expected occupancy levels when buying tickets online**

The Swiss train operator, SBB, provides information on the expected level of crowding on specific trains at the point where passengers buy their tickets on the website, see Figure 8 (SBB n.d.). The information is based on historic data from automatic door sensors, which have been in use on their trains for several years. There are three levels: "Low to Medium"; "High"; and "Very High".

Connections for Th, 10.07.14									
1		Basel SBB	dep 17:07	0:53	0	ICE	1.  2.	from CHF 16.00	
		Zürich HB	arr 18:00					<a href="#">Fare/Buy</a>	
2		Basel SBB	dep 17:13	1:11	0	IR	1.  2.	from CHF 16.00	
		Zürich HB	arr 18:24				Very high occupancy expected	<a href="#">Fare/Buy</a>	
3		Basel SBB	dep 17:33	0:53	0	IC	1.  2.	from CHF 16.00	
		Zürich HB	arr 18:26					<a href="#">Fare/Buy</a>	
4		Basel SBB	dep 17:36	0:59	0	TGV	1.  2.	from CHF 16.00	
		Zürich HB	arr 18:35					<a href="#">Fare/Buy</a>	

[Show all details](#) [Print](#) [Timetable Booklet](#)

[Earlier](#) [First connection](#) [Last connection](#) [Later](#)

[e-mail](#) [Word](#) mySBB: [Departure](#) [Arrival](#) [Relation](#)

**Legend**  
 Low to medium occupancy expected  
 High occupancy expected  
 Very high occupancy expected  
 Restaurant  
 BICYCLES: No self-service loading by sender

**Figure 8 – Expected occupancy levels when buying tickets online (reproduced from SBB n.d.)**

The critical factor as to the success of encouraging passengers to take alternative trains is likely to be in ensuring that the information is made available at a point in time where the user is reasonably able to alter their travel plans. For example, if on arrival at a station a traveller is informed that the next train is full and that the following train is less full, the individual's decision on whether to act on that information is likely to be influenced by the arrival time of the following train. How long a traveller is prepared to wait is itself likely to be influenced by a host of other factors, such as:

- How long their overall journey is expected to take
- The purpose of their travel (e.g. potential scheduling commitments)
- Trust in the information provided (both as to the predicted waiting time and the certainty that they will get a seat)
- Facilities at the station (leisure, food and drink, toilets, seating, waiting rooms)
- Weather conditions (linked to waiting facilities)
- Whether or not they are booked onto a particular service

Measures are therefore likely to be more effective if the information can be provided in advance, for example when buying a ticket online. However, if this information is



provided in advance, it would likely need to be based on historic trends as with the SBB website, rather than real-time information.

As the information is static, passengers may accept that the guidance will not be right all the time. However, if the information is consistently inaccurate, many passengers will ignore the guidance; the predictions should therefore be made to be as accurate as possible. APC data, either from door sensors or weight sensors could be used to generate such predictions taking into account variability due to day of week, seasonal effects and events such as football matches and concerts.

### 2.3.6.2 Pilots of rail passenger crowding posters

South West Trains and the ORR carried out a trial of posters to show passengers which trains are typically the most crowded, helping those who have the option to choose a different train that avoids crowding (ORR 2012). Posters with seat availability information were put up at five SWT stations (Basingstoke, Farnborough, Motspur Park, New Malden, Surbiton) between November 2011 and February 2012, see Figure 9. Questionnaires were used to collect passengers' views on whether they had seen the information and if it had any impact on their choice of train.



Figure 9 – Posters used in the South West Trains pilot project (reproduced from ORR 2012)



Survey results showed that 34% of respondents had seen the information and of those, over two thirds found it at least fairly useful and almost 10% found it very useful. Just over a fifth of respondents who had seen the information have regularly or occasionally changed the trains they get as a result of the information published.

To complement the survey, passenger loading data collected using automatic passenger counters were analysed to see if any changes in loading could be detected. Stations further out, like Basingstoke and Farnborough, have experienced some changes in passenger behaviour e.g. passengers changing to less crowded services, although it was not possible to say with certainty whether this was as a result of viewing the information in the capacity posters. Stations closer to London may be relatively insensitive to the effects of posters, with shorter journey times into London suggesting that a “turn-up-and-go” approach is preferred by passengers.

In a related development, similar initiatives undertaken by London Midland were praised by the DfT in August 2013 (Watt 2013):

“Norman Baker is urging train companies to follow the example of London Midland..., which colour-codes trains using a red, amber and green ‘traffic light’ system. ‘Publication of train-by-train crowding information is, in the short term, an important tool for allowing passengers to make informed choices about which trains to travel on, and convincing those passengers who can change their travel patterns to do so... The innovative approach taken by London Midland is helping to smooth the peaks in demand for their services and is making the most of the investment going into rail services in their area. I am keen to see the rail industry working together to follow London Midland's example.”

### **2.3.6.3 Inter-train crowding information via a smartphone app and posters**

C2C, the train operator of the Essex Thameside franchise, provides information on the expected level of crowding on different trains in the peak period for some stations through both their ‘C2C-Live’ smartphone app and also through posters (C2C n.d.). This includes three levels of occupancy: ‘Seats normally available’ (green); ‘Few seats normally available’ (amber); and ‘Standing room only’ (red). The app was created by IBM using historical APC data on commuting patterns.

Smartphones are a good channel for such inter-train crowding information, because it is possible to search for different stations and train services. As with the C2C-Live app, such information can be provided along with other sources of information, such as

station facilities and delay information, which can encourage higher uptake compared to a standalone app.

#### **2.3.6.4 Various – Expected levels of crowding from user-generated data**

A number of projects have been identified that share user-generated (i.e. 'crowdsourced') data on the typical levels of crowding on particular services.

The 'Moovit' app is an example of using crowdsourcing to collect and share information on passenger comfort to help users identify the least crowded services (Crawford 2013). A previous version of the Moovit app had a 'Report' functionality, which enabled users to share information on delays, accidents and levels of crowding, although this functionality is no longer available in the most recent version of the app. The commercially available app has been tested in various locations, including the San Francisco Bay Area Rapid Transport system. As of June 2014 there had been over one million downloads worldwide.

A campaign by the Green Party in Melbourne collected crowdsourced data provided by members of the public via a special smartphone app or an online survey (Barber 2012). People reported on how many people there were in the same coach as them and also gave qualitative comments on crowding. The data was then used to provide feedback to users on typical historical levels of crowding on a network map.

#### **Summary**

The purpose of the review of solutions was to understand what concepts have been proposed or piloted and to explore the extent to which there exists evaluation evidence on the impact of these interventions.

A review of the current use of systems for reducing uneven loadings within trains has identified a number of examples, some of which are now commercially available and being installed on new trains. Examples have been identified for both metro type operations and longer distance inter-city type trains.

The most sophisticated example found was that of the Dutch pilot at Den Bosch station, which provided real-time information on coach-by-coach occupancy. This pilot favoured infra-red door sensors to provide automatic passenger counts and then transmitted this information to passengers, initially via a smartphone app, and then later by an LED display along the full length of the platform. A number of similar

solutions were also identified, although mainly at the concept stage. There is good evidence that the full range of technologies required to provide a fully automated real-time coach-by-coach system already exists and is, in various ways, already appearing in the market.

There were also examples of more low-tech static solutions to address the same problem, i.e. where historical loading data is used to identify patterns and trends in coach-by-coach crowding. In some circumstances static solutions may be sufficient, either through information on posters and signs to direct people towards the least busy carriages/train services, or through passive mitigations such as design features to subconsciously encourage different passenger distributions in stations.

A number of different systems were identified that provide train-by-train information, suggesting that there is also potential demand for this type of system. In many situations, such systems can be operated using data derived from historical loading data, whether this is automatically collected (door sensors or weight sensors) or in some cases user-generated (i.e. 'crowdsourced'). Examples were found that presented such information via different information channels, including posters, websites and smartphone apps.

However, while the review identified a number of approaches that could in principle be adopted, very little evidence was found of the impact of such measures, or their costs. A simulation study by the Korean Transport Institute concluded that a 15% reduction in dwell times could be achieved by giving passengers real-time information on where seats are most likely to be available, however no equivalent figure has been identified from an in-service trial.

## **2.4 Techniques**

This section gives an introduction to relevant data mining and machine learning techniques, as defined in §2.4.1. An overview of particular techniques is given in §2.4.2 and some of these techniques were applied to the study dataset. Finally, in §2.4.3 there is a brief review of such examples in the context of transportation research.

### **2.4.1 Definitions of data mining and machine learning**

Bramer (2007) defined the knowledge discovery process as “the non-trivial extraction of implicit, previously unknown and potentially useful information from data... it is a

process of which data mining forms just one part, albeit a central one... prepared data is... passed to a data mining algorithm, which produces an output in the form of rules or some other kind of patterns... applications [of data mining] can be divided into four main types: classification, numerical prediction, association and clustering". Bramer defined two types of data as follows:

*"For the first type there is a specially designated attribute and the aim is to use the data given to predict the value of that attribute for instances that have not yet been seen. Data of this kind is called labelled. Data mining using labelled data is known as supervised learning. If the designated attribute is categorical, the task is called classification. If the designated attribute is numerical, e.g. the expected sale price of a house, the task is called regression. Data that does not have any specially designated attribute is called unlabelled. Data mining of unlabelled data is known as unsupervised learning. Here the aim is simply to extract the most information we can from the data available."*

Zaki and Meira (2014) have published a free textbook called 'Data mining and analysis: fundamental concepts and algorithms', in which they defined data mining as "the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data". The textbook gives a detailed review of four areas of data mining in turn: exploratory data analysis, frequent pattern mining, clustering, and classification.

A commonly quoted definition of machine learning is quite broadly seeking to answer the question, "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?" (Mitchell 2006). More specifically,

*"We say that a machine learns with respect to a particular task  $T$ , performance metric  $P$ , and type of experience  $E$ , if the system reliably improves its performance  $P$  at task  $T$ , following experience  $E$ . Depending on how we specify  $T$ ,  $P$ , and  $E$ , the learning task might also be called by names such as data mining, autonomous discovery, database updating, programming by example, etc." (Mitchell 2006)*

Ng (2013) gave four example applications of machine learning, one of which included data mining:

- "Database mining – large datasets from growth of automation / web, e.g. web click data, medical records, biology, engineering etc;

- *Applications we can't program by hand – e.g. autonomous helicopters, handwriting recognition, most Natural Language Processing (NLP), Computer vision;*
- *Self-customising programs – e.g. Amazon, Netflix, product recommendations;*
- *Understanding human learning – brain, real AI.”*

Brownlee (2013) gave a good introduction and overview to different types of machine learning algorithms. He defined four different types of learning style, the two main ones relevant to data mining being supervised and unsupervised, similar as above. The two other types defined were semi-supervised learning where there is only a small amount of labelled data and also reinforcement learning which is focused on training robots with punishment and reward. Brownlee then proposed 12 categories of algorithm, although admitted that arguably particular algorithms may fit in more than one category, as reproduced in Table 2.

**Table 2 – Categories of machine learning algorithms (reproduced from Brownlee 2013)**

<b>Type</b>	<b>Category and description</b>	<b>Example algorithms</b>
Supervised	<b>Instance based</b> learning model a decision problem with instances or examples of training data that are deemed important or required to the model. Such methods typically build up a database of example data and compare new data to the database using a similarity measure in order to find the best match and make a prediction. For this reason, instance-based methods are also called winner-take all methods and memory-based learning.	<ul style="list-style-type: none"> <li>• <i>k</i>-Nearest Neighbour (kNN)</li> <li>• Learning Vector Quantisation (LVQ)</li> <li>• Self-Organising Map (SOM)</li> </ul>
Supervised	<b>Bayesian</b> methods are those that are explicitly apply Bayes' Theorem for problems such as classification and regression.	<ul style="list-style-type: none"> <li>• Naive Bayes</li> <li>• Averaged One-Dependence Estimators (AODE)</li> <li>• Bayesian Belief Network (BBN)</li> </ul>
Supervised	<b>Decision tree</b> methods construct a model of decisions made based on actual values of attributes in the data. Decisions fork in tree structures until a prediction decision is made for a given record. Decision trees are trained on	<ul style="list-style-type: none"> <li>• CART</li> <li>• ID3</li> <li>• C4.5</li> <li>• CHAID</li> <li>• Decision Stump</li> </ul>

Type	Category and description	Example algorithms
	data for classification and regression problems.	<ul style="list-style-type: none"> <li>• Random Forest</li> <li>• MARS</li> <li>• GBM</li> </ul>
Supervised	<b>Kernel Methods</b> are best known for the popular method Support Vector Machines which is really a constellation of methods in and of itself. Kernel Methods are concerned with mapping input data into a higher dimensional vector space where some classification or regression problems are easier to model.	<ul style="list-style-type: none"> <li>• Support Vector Machines (SVM)</li> <li>• Radial Basis Function (RBF)</li> <li>• Linear Discriminate Analysis (LDA)</li> </ul>
Supervised	<b>Regression</b> is concerned with modelling the relationship between variables that is iteratively refined using a measure of error in the predictions made by the model. Regression methods are a work horse of statistics and have been co-opted into statistical machine learning. This may be confusing because we can use regression to refer to the class of problem and the class of algorithm. Really, regression is a process.	<ul style="list-style-type: none"> <li>• Ordinary Least Squares</li> <li>• Logistic Regression</li> <li>• Stepwise Regression</li> <li>• Multivariate Adaptive Regression Splines (MARS)</li> <li>• Locally Estimated Scatterplot Smoothing (LOESS)</li> </ul>
Supervised	<b>Regularisation methods</b> are popular, powerful and generally simple modifications made to other methods. An extension made to another method (typically regression methods) that penalises models based on their complexity, favouring simpler models that are also better at generalising.	<ul style="list-style-type: none"> <li>• Ridge Regression</li> <li>• Least Absolute Shrinkage and Selection Operator (LASSO)</li> <li>• Elastic Net</li> </ul>
Supervised	<b>Artificial Neural Networks</b> are models that are inspired by the structure and/or function of biological neural networks. They are a class of pattern matching that are commonly used for regression and classification problems but are really an enormous subfield comprised of hundreds of algorithms and variations for all manner of problem types.	<ul style="list-style-type: none"> <li>• Perceptron</li> <li>• Back-Propagation</li> <li>• Hopfield Network</li> <li>• Self-Organising Map (SOM)</li> <li>• Learning Vector Quantisation (LVQ)</li> </ul>
Semi-supervised	<b>Deep Learning</b> methods are a modern update to Artificial Neural Networks that exploit abundant cheap computation. They are	<ul style="list-style-type: none"> <li>• Restricted Boltzmann Machine (RBM)</li> <li>• Deep Belief Networks</li> </ul>

Type	Category and description	Example algorithms
	concerned with building much larger and more complex neural networks, and as commented above, many methods are concerned with semi-supervised learning problems where large datasets contain very little labelled data.	(DBN) <ul style="list-style-type: none"> <li>• Convolutional Network</li> <li>• Stacked Auto-encoders</li> </ul>
Unsupervised	<b>Clustering</b> , like regression describes the class of problem and the class of methods. Clustering methods are typically organised by the modelling approaches such as centroid-based and hierarchal. All methods are concerned with using the inherent structures in the data to best organise the data into groups of maximum commonality.	<ul style="list-style-type: none"> <li>• k-Means</li> <li>• Expectation Maximisation (EM)</li> </ul>
Unsupervised	<b>Association rule</b> learning are methods that extract rules that best explain observed relationships between variables in data. These rules can discover important and commercially useful associations in large multidimensional datasets that can be exploited by an organisation.	<ul style="list-style-type: none"> <li>• Apriori algorithm</li> <li>• Eclat algorithm</li> </ul>
Unsupervised	Like clustering methods, <b>Dimensionality Reduction</b> seek and exploit the inherent structure in the data, but in this case in an unsupervised manner or order to summarise or describe data using less information. This can be useful to visualise dimensional data or to simplify data which can then be used in a supervised learning method.	<ul style="list-style-type: none"> <li>• Principal Component Analysis (PCA)</li> <li>• Partial Least Squares Regression (PLS)</li> <li>• Sammon Mapping</li> <li>• Multidimensional Scaling (MDS)</li> <li>• Projection Pursuit</li> </ul>
Ensemble	<b>Ensemble methods</b> are models composed of multiple weaker models that are independently trained and whose predictions are combined in some way to make the overall prediction. Much effort is put into what types of weak learners to combine and the ways in which to combine them. This is a very powerful class of techniques and as such is very popular.	<ul style="list-style-type: none"> <li>• Boosting</li> <li>• Bootstrapped Aggregation (Bagging)</li> <li>• AdaBoost</li> <li>• Stacked Generalisation (blending)</li> <li>• Gradient Boosting Machines (GBM)</li> <li>• Random Forest</li> </ul>



As defined above, supervised learning involves situations with labelled data, i.e. where there is a specially designated attribute and the aim is to use other attributes of the data given to predict the value of the labelled data for instances that have not yet been seen. Classification techniques are used when the labelled data is categorical, i.e. either: nominal (un-ordered categories, e.g. 'Green', 'Brown'...); ordinal (ordered categories, e.g. 'Low', 'Medium', 'High'...); or Boolean (0 or 1). Furthermore, the classes need to be mutually exhaustive and exclusive, i.e. each instance is assigned to precisely one class. Typically the data is split into a training dataset and a test dataset. The training dataset is used to build a predictive model and this model is then applied to the test dataset. The predictions for the test dataset are then compared against the truth data to assess how accurate the predictive model was.

## 2.4.2 Overview of particular techniques

### 2.4.2.1 Naïve Bayes

The 'Naïve Bayes' algorithm is a classification technique that can be applied when the attributes used in the prediction are categorical and independent. Bramer (2007) presents a good worked example of how to apply the technique and describes it formally as follows:

*Given a set of  $k$  mutually exclusive and exhaustive classifications  $c_1, c_2 \dots c_k$ , which have prior probabilities  $P(c_1), P(c_2), \dots, P(c_k)$ , respectively, and  $n$  attributes  $a_1, a_2, \dots, a_n$  which for a given instance have values  $v_1, v_2 \dots, v_n$  respectively, the posterior probability of class  $c_i$  occurring for the specified instance can be shown to be proportional to*

$$P(c_i) \times P(a_1 = v_1 \text{ and } a_2 = v_2 \dots \text{ and } a_n = v_n \mid c_i)$$

*Making the assumption that the attributes are independent, the value of this expression can be calculated using the product*

$$P(c_i) \times P(a_1 = v_1 \mid c_i) \times P(a_2 = v_2 \mid c_i) \times \dots \times P(a_n = v_n \mid c_i)$$

*We calculate this product for each value of  $i$  from 1 to  $k$  and choose the classification that has the largest value.*

Essentially, for a particular class,  $i$ , the technique uses the training dataset to calculate the conditional probabilities for all the independent variables, given that the class =  $i$ .



Then for a new instance where the class is not yet known, the product of the prior probability and all the relevant conditional probabilities generates a score for class,  $i$ . This score is calculated for all other classes and the class with the highest score is selected as the predicted class.

An example where this approach may be applicable is in predicting the level of crowding on a particular train in the future, where the dependent variable is categorical, e.g. 'Low', 'Medium' or 'High'. Categorical predictor variables, such as scheduled departure time, day of week, number of coaches and so on could be used to predict for unknown cases the expected level of crowding. Such predictions could feed in to passenger information services, such as those identified in §2.3.6.1.

#### 2.4.2.2 Decision trees

Bramer (2007) gives an introduction to decision trees, defining terminology as below

*“A **decision tree** is created by a process known as **splitting on the value of attributes** (or just splitting on attributes), i.e. testing the value of an attribute... and then creating a branch for each of its possible values... One commonly used method [for choosing which attribute to split on] is to select the attribute that minimises the value of **entropy**, thus maximising the **information gain**.*

*Entropy is an information-theoretic measure of the ‘uncertainty’ contained in a training set, due to the presence of more than one possible classification.*

$$E = - \sum_{i=1}^K p_i \log_2 p_i$$

*E is positive or zero for all training sets and takes its minimum value (zero) if and only if all the instances have the same classification, in which case there is only one non-empty class, for which the probability is 1...*

*The process of decision tree generation by repeatedly splitting on attributes is equivalent to partitioning the initial training set into smaller training sets repeatedly, until the entropy of each of these subsets is zero (i.e. each one has instances drawn from only a single class). At any stage of this process, splitting on any attribute has the property that the average entropy of the resulting subsets will be less than (or occasionally equal to) that of the previous training set.”*

### 2.4.2.3 Other techniques

There is substantial literature on the various other algorithms listed in Table 2. Wu et al. (2008) summarised the ‘top ten’ data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) in December 2006. These were listed as: C4.5; k-Means; SVM; Apriori; EM; PageRank; AdaBoost; kNN; Naive Bayes; CART. The review paper gives an overview of how to use each technique in turn along with discussions on the implications and limitations. Kotsiantis (2007) is another good review paper focused on various classification techniques.

An industry workshop was held by the Knowledge Transfer Network (KTN) and the Smith Institute in 2014 to explore the mathematical foundations of data science (Knowledge Transfer Network & Smith Institute 2014). Feedback from the workshop delegates was that there is “a shift towards the use of more advanced and adaptable methods that have the ability to process structured or unstructured data, of changing sizes and dimensions, that is discrete or continuous, and with varying levels of complexity”. The following list of techniques were identified as having been successfully applied to data science challenges: probability theory; hidden Markov models; evolving and multiplex networks; deep learning; Bayesian analysis; classification and clustering; data mining; graphical models; topological data analysis; tropical geometry; dynamical systems; machine learning; sparse tensor methods; stochastic optimisation tools; large-scale linear algebra. The workshop notes provide links to reference material for each of these areas of study.

### 2.4.3 Applications of data mining techniques in transportation

The research papers first reviewed in §2.1 to §2.3 were revisited to review any data mining techniques used.

Lathia and Capra (2011) built a personalised recommendation system to predict the type of ticket users should purchase (see §2.2.3). They applied four different classifiers as listed below.

- Baseline – This simple classifier returned the most frequent class in the training set.
- Naïve Bayes – this classifier, based on Bayes’ theorem, assumed that each feature of the users’ profiles was independent from the others; geographic features were transposed into binary variables.

- k-Nearest Neighbours – this technique operated by finding, for each test profile, the k most similar profiles; the predicted class was then selected to be the most frequent class that appeared in the neighbour set.
- Decision Trees – the C4.5 algorithm was used to generate a decision tree to classify test instances; it did so by recursively partitioning the data on a single attribute, according to the measured information gain of each split.

Ceapa et al. 2012 used Oyster card ticket gate data at London Underground stations to investigate patterns in station crowding (see §2.2.3.1). They then built and evaluated three prediction models and investigated the effect of several parameters on the accuracy of the results as follows:

- 'Historic value' – This was the most basic predictor, taking the one corresponding value in the training dataset for the corresponding day of week and time of day;
- 'Historic mean' – This was similar to the 'Historic value' predictor, but used the mean of all values in the training dataset for the corresponding day of week and time of day;
- 'Historic trend' – This attempted to improve on the other two predictors by taking into consideration crowding level for the current time.

Two further predictors, one based on linear regression with ordinary least squares to estimate parameters, and one based on Kalman Filters were also attempted although offered no improvement over the three techniques above.

Stenneth et al. 2011 used GPS and accelerometer data off smartphones to predict mode of transport (see §2.1.2.3). To determine the most accurate classifier they compared the precision and recall accuracy of five distinct classification models: Naive Bayes; Bayesian Network; Decision Trees; Random Forest; Multilayer Perceptron. To evaluate the different classification models on transportation mode detection, the Weka machine learning tool set was used. The results indicated that the Random Forest technique produced the best model, with an average precision accuracy of 93.7% and recall of 93.8%.

## Summary

Supervised learning involves situations with labelled data, i.e. where there is a specially designated attribute and the aim is to use other attributes of the data given to predict the value of the labelled data for instances that have not yet been seen. Classification techniques are used when the labelled data is categorical. A brief review of supervised learning classification techniques was undertaken, in particular the Naïve Bayes and decision tree techniques.

Such techniques may be applicable in predicting the level of crowding on a particular train in the future, where the dependent variable is categorical, e.g. 'Low', 'Medium' or 'High'. Categorical predictor variables, such as scheduled departure time, day of week, number of coaches and so on could be used to predict for unknown cases the expected level of crowding. Examples of the use of such techniques in transport research have been reviewed.

### 3 Research focus, aims and questions

The literature review identified relevant research in loading diversity and related areas; some of the key points that informed the research focus are summarised again in §3.1. The research aims arise from this review, namely in attempting to address some of the gaps and opportunities that were identified, see §3.2. This then helped to formulate specific research questions, listed in §3.3, which informed the main analysis tasks.

#### 3.1 Research focus

There is research to suggest that conditions on trains that are near or over capacity can lead to lower customer satisfaction, discourage passengers from using the train, result in negative economic impacts and have been found to be a factor in a number of health and safety hazards. Various organisations have raised concerns over the level of rail passenger crowding on some trains in the UK, highlighting how this is an important issue for rail passengers and train operators alike (§1.1).

Over recent years, in many different domains, there has been a proliferation of new and larger datasets and this has been accompanied by the emergence of data science as a discipline with new techniques and software. Through this thesis I sought to address the issue of rail passenger crowding in the UK from a data analytics standpoint. The primary focus was on extracting more value from existing ‘found data’ by reusing air pressure data from suspension systems as a measure of carriage loadings along a train, to help assess the case for investing in initiatives to optimise passenger loadings (§1.2).

The application of data-intensive methods to different aspects of transportation is an emerging area and §2.1 gave an overview of several examples that situate the thesis in the wider context of related research. This included a discussion on the fields of Intelligent Transport Systems (ITS) and the Internet of Things (IoT), which are common in that they both involve data flows from source to consumption. Passenger information systems are a public transport example of ITS (§2.1.1). There exists an emerging field of research around ‘found data’ and the ‘digital footprint’ from ubiquitous computing. Examples discussed included: mobility data and mobility profiling; investigating how users interact with cycle hire schemes; using data from smartphones to detect the mode of transport of travellers; and the mining of transport-related tweets as a passive data source (§2.1.2). Crowdsourcing is another area of research where data is actively

collected from the public. In a transportation context there have been several research projects that have successfully applied the crowdsourcing concept to different interest groups (§2.1.3). Examples of research projects were identified where ICT has been used to encourage use of more sustainable transport; some studies found that providing better information on available alternatives was effective in encouraging behavioural change (§2.1.4).

The DfT conducts an annual survey of rail passenger crowding, which is increasingly based on passenger counts from APC systems. Two measures of rail crowding are used: 'passengers in excess of capacity' and 'standard class load factor'; some limitations of these measures are acknowledged by the DfT, namely that they do not reflect coach-by-coach variations, nor do they reflect variations across the peak period (§2.2.1). The two main types of APC sensors used in the UK rail industry are door sensors and weight sensors, with around 40% of the fleet equipped as of 2010. The literature review did not find any academic research on the use of APC data in rail in the UK, which may be because the data is owned by the train operators and is commercially sensitive. This perhaps suggests that there is a gap for novel research if it were possible to gain access to this data source. Some academic research was found into the use of APC data for estimating dwell times of buses, but this was not directly relevant (§2.2.2). There is a substantial amount of academic research into AFC data across a range of topic areas, some of which has focused on estimating levels of crowding for the London Underground. The structure of AFC data is typically that of 'touch-ins' and 'touch-outs' and so does not provide a direct link to crowding and occupancy data, unlike APC data. Nevertheless, it has been shown that if the variability is low, simple predictors can give quite accurate predictions of crowding, even with a relatively small amount of data (§2.2.3).

The TRB have proposed a three-part definition of 'loading diversity' to describe the distribution of passengers: i) near doors within a coach; ii) from coach to coach; iii) between trains across the peak period. The metrics associated with these definitions are not readily applicable to heavy rail, although they do provide a useful starting point to explore alternative measures of uneven occupancy on trains. The first type of loading diversity relates to carriage design and was not the main focus of the thesis. The second type of loading diversity was of interest, although the current TRB metric is only a relative measure and does not include information on the capacity. The third type of loading diversity was also of interest, although again the current metric may not be appropriate for use in heavy rail because of larger headways between trains

(§2.2.4.1). The review identified only a few research papers in the area of loading diversity. One paper found that about three-quarters of subway users reported choosing a specific coach intentionally and of these when asked to explain their motivation there were a variety of reasons, although 70% said “to minimise the walking distance to exit when they disembarked at a destination station”. This paper went some way to investigate the causes of uneven loadings, but it should be noted that the survey was conducted for a subway during the weekday morning peak and so may not be representative of other scenarios (§2.2.4.2).

Existing research studies have found that the mean alighting and boarding time per passenger is typically around one second, although this increases or decreases depending on various conditions, such as passenger mobility, platform design, vehicle design and crowding effects. The relationship between loading diversity and boarding times is likely two-fold. Firstly, the boarding time for a whole train is equal to the maximum boarding time of each individual door, thus if a disproportionate number of passengers board and alight from one particular door this will result in higher overall boarding time. Secondly, there is research from multiple studies to suggest that once the vestibule reaches a certain threshold of crowding, the flow rate at which passengers can board reduces. Thus it is intuitive that optimising boarding patterns along the platform will result in lower boarding times, although this relationship is likely non-linear if there is an interactive effect of crowding in vestibules (§2.2.5).

A review of solutions identified a number of interventions for reducing uneven loadings on trains, in both metro operations and longer distance inter-city trains. The purpose of this review was to understand what concepts have been proposed or piloted and to explore the extent to which there exists evaluation evidence on their impact (§2.3). The most sophisticated example found was that of the Dutch pilot at Den Bosch station, which provided real-time information on coach-by-coach occupancy. This pilot favoured infra-red door sensors to provide automatic passenger counts and then transmitted this information to passengers, initially via a smartphone app, and then later by an LED display along the full length of the platform. A number of similar solutions were also identified from other sources, although mainly at the concept stage. There is good evidence that the full range of technologies required to provide a fully automated real-time coach-by-coach system already exists and is, in various ways, already appearing in the market (§2.3.1). There were also examples of more low-tech static solutions to address the same problem, i.e. where historical loading data is used to identify patterns and trends in coach-by-coach crowding. In some circumstances static solutions may be



sufficient, either through information on posters and signs to direct people towards the least busy carriages/train services (§2.3.3), or through passive mitigations such as design features to subconsciously encourage different passenger distributions in stations (§2.3.4).

A number of different systems were identified that provide train-by-train information, suggesting that there is also potential demand for this type of system. In many situations, such systems can be operated using data derived from historical loading data, whether this is automatically collected (door sensors or weight sensors) or in some cases user-generated (i.e. 'crowdsourced'). Examples were found that presented such information via different information channels, including posters, websites and smartphone apps (§2.3.6).

However, while the review identified a number of approaches that could in principle be adopted, very little evidence was found of the impact of such measures, or their costs. A simulation study by the Korean Transport Institute concluded that a 15% reduction in dwell times could be achieved by giving passengers real-time information on where seats are most likely to be available, however no equivalent figure has been identified from an in-service trial.

A review of supervised learning techniques identified approaches that may be applicable to the study (§2.4). Supervised learning involves situations with labelled data, i.e. where there is a specially designated attribute and the aim is to use other attributes of the data given to predict the value of the labelled data for instances that have not yet been seen. Classification techniques are used when the labelled data is categorical. A brief review of supervised learning classification techniques was undertaken, in particular focusing on the Naïve Bayes and decision tree techniques. Such techniques may be applicable in predicting the level of crowding on a particular train in the future, where the dependent variable is categorical, e.g. 'Low', 'Medium' or 'High'. Categorical predictor variables, such as scheduled departure time, day of week, number of coaches and so on could be used to predict for unknown cases the expected level of crowding. Examples of the use of such techniques in transport research have been reviewed.

### 3.2 Research aims

The literature review identified several gaps in the knowledge:

- The literature review did not find any academic research on the use of APC data in rail in the UK, which may be because the data is owned by the train operators and is commercially sensitive.
- Existing UK measures for rail crowding do not reflect coach-by-coach variations, nor do they reflect variations across the peak period; furthermore, American metrics for 'loading diversity' are not readily applicable to heavy rail, although they are a good starting point.
- There is evidence to suggest that optimising the distribution of passengers on the train may lead to dwell time savings, but there is only limited research on the extent of these savings.
- Although a wide range of solutions have been proposed that would seek to optimise loading diversity, only limited evaluation evidence is available on the impact of such systems.

The literature review also identified several opportunities:

- APC data has the potential to be used to investigate a wide range of questions surrounding on-train crowding.
- The use of APC door sensors and weight sensors on trains is becoming increasingly widespread in the UK, with it being specified for installation on all new trains and already over 40% of existing trains equipped.
- There are many different data mining techniques that may be readily applicable to help realise the potential offered by APC data.
- Crowding on trains in the UK is an important problem from both a passenger and operator perspective. As such, investigating measures to ease crowding would likely yield benefits to both passengers and operators.

An 11-month sample of data was obtained from a UK train operating company, which included air suspension pressure, the times the doors lock/unlock and timetable information. This dataset was manipulated with a view to addressing the gaps and

exploiting the opportunities identified above. Specifically the following hypothesis was proposed:

- Hypothesis: Automatic passenger counting (APC) data has the potential to deliver more even loadings on trains through the provision of new real-time and static solutions; furthermore, such solutions have benefits of reduced dwell times and reduced crowding and are of value to rail passengers and train operators.

### 3.3 Main research questions

The following five main research questions were defined in the context of the research focus and research aims; these questions were primarily targeted at exploring the potential of weight-based APC data in understanding and predicting loading diversity.

#### Main research questions on loading diversity

##### RQ 1. Quantification of inter-coach loading diversity

- (a) Is it possible to identify coach-by-coach occupancy using air suspension data?
- (b) Which train services and stations had the most uneven inter-coach loadings?
- (c) For what proportion of train departures were there available seats in at least one coach, while passengers were standing elsewhere on the train?
- (d) Which stations had the highest proportion of passengers who were standing, but could have been sitting down?
- (e) What metrics are appropriate to quantify the inter-coach loading diversity?

##### RQ 2. Link between inter-coach loading diversity and dwell times

- (a) What was the distribution of dwell times and estimated boarding times?
- (b) To what extent were longer dwell times related to uneven coach-by-coach occupancy?

##### RQ 3. Quantification of inter-train loading diversity

- (a) How did load factor vary by station, time of day and other attributes?
- (b) What metrics are appropriate to quantify the inter-train loading diversity?

##### RQ 4. Prediction of inter-train loading diversity

- (a) What techniques can be applied to predict load factor, based on historic data?
- (b) What was the accuracy, precision and recall of these predictors?

##### RQ 5. Perceived causes and effects of loading diversity

- (a) From the perspective of frontline staff, which services had high levels of inter-coach local crowding?
- (b) From the perspective of frontline staff, which services had high levels of inter-train local crowding?
- (c) What were the perceived causes of inter-coach local crowding?
- (d) What were the perceived causes of inter-train local crowding?
- (e) What were the perceived effects of crowding, both inter-coach and inter-train?

**Rationale:** The initial focus of the study was to use the APC data to quantify the loading diversity, both for inter-coach (RQ 1) and inter-train (RQ 3). The purpose of this was to investigate whether it was possible to use the data in this way and also to understand any patterns and trends.

Several different solutions that aim to optimise inter-coach loading diversity were identified in the literature review, although a gap in the knowledge was the lack of evaluation evidence on the impacts of these solutions. Reduced boarding times would be a key benefit to train operators and so questions on the relationship between boarding time and inter-coach loading diversity were defined (RQ 2); these questions aimed to identify the scale of the possible dwell time savings that could be achieved if all trains were evenly loaded.

Examples of inter-train solutions were identified whereby some train operators use APC data to provide information on the level of crowding for a particular service through a variety of information channels (§2.3.6). However, only limited information is published on the method for generating these predictions and their accuracy. The other main focus of the study was therefore defined to investigate how the APC data could be used to generate predictions for the level of crowding on a particular train (RQ 4) and thus address this gap in the research. More specifically, the questions on techniques were appropriate in understanding whether predictive information in such systems can be sufficiently accurate to be of value to rail passengers.

Better understanding of the root causes of local crowding may help to identify how these issues could be overcome, whereas better understanding of the effects of local crowding sets the context for the scale of the issue. As such, questions were set to aim to gain a better understanding on the perceived causes and effects of uneven loadings (RQ 5).

A gap identified from the literature review was the lack of evaluation evidence on the impact of local crowding solutions. It was unfeasible to conduct full-scale pilots within the scope of the MPhil; however the following additional questions were partially addressed through the literature review and in the discussion.

- How can the solutions be categorised, both inter-coach and inter-train?
- How would passengers interact with the solutions, both inter-coach and inter-train?
- What would be the impacts of the solutions, both inter-coach and inter-train?
- What are the most promising solutions, both inter-coach and inter-train?

## 4 Methodology

This section outlines the methodology of the tasks that were undertaken in order to answer the research questions. Figure 10 outlines the approach of the study, with tasks relating to ‘inter-coach’ loading diversity above the dotted line and those relating to ‘inter-train’ loading diversity below the dotted line.

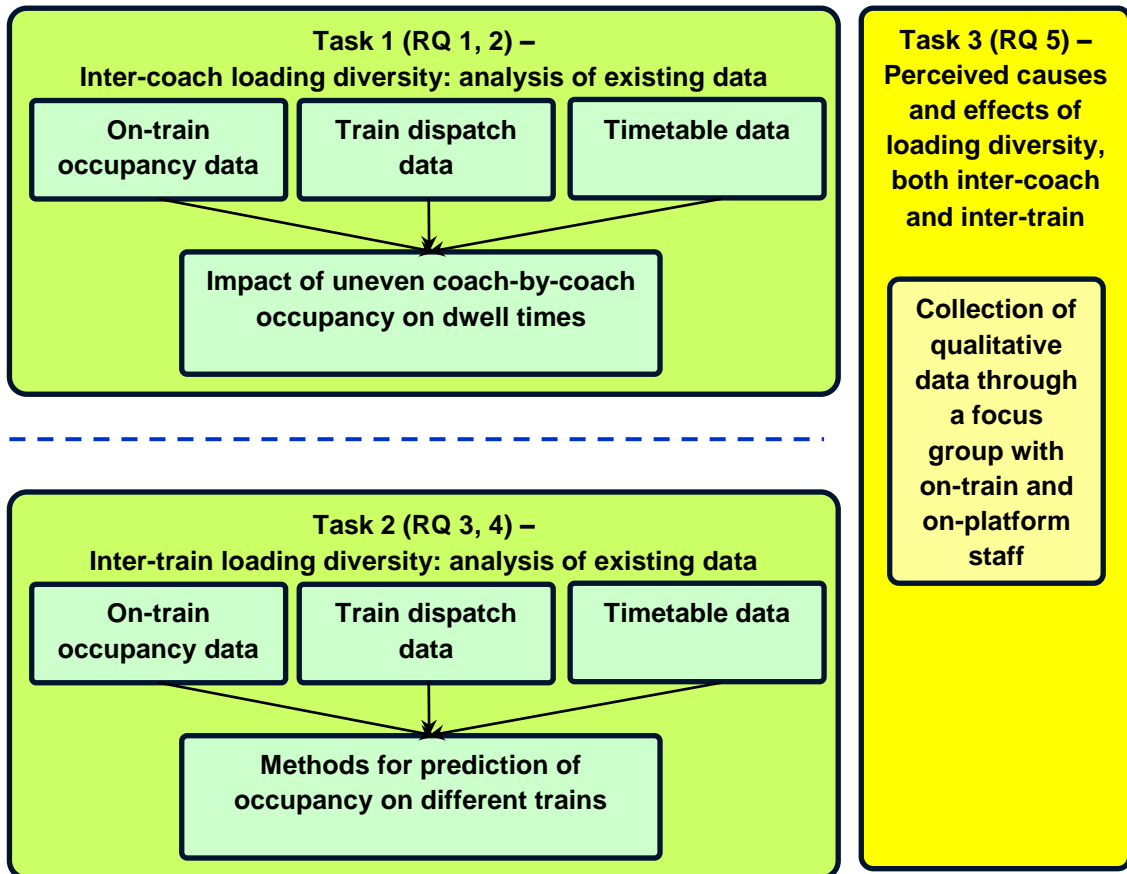


Figure 10 – Overview of tasks

### 4.1 Task 1 – Inter-coach loading diversity: analysis of existing data

Task 1 was designed to answer the following research questions:

- **RQ 1 – Quantification of inter-coach loading diversity**
  - (a) *Is it possible to identify coach-by-coach occupancy using air suspension data?*
  - (b) *Which train services and stations had the most uneven inter-coach loadings?*
  - (c) *For what proportion of train departures were there available seats in at least*

*one coach, while passengers were standing elsewhere on the train?*

*(d) Which stations had the highest proportion of passengers who were standing, but could have been sitting down?*

*(e) What metrics are appropriate to quantify the inter-coach loading diversity?*

▪ **RQ 2 – Link between inter-coach loading diversity and dwell times**

*(a) What was the distribution of dwell times and estimated boarding times?*

*(b) To what extent were longer dwell times related to uneven coach-by-coach occupancy?*

The literature review identified gaps and opportunities in this area, specifically in the opportunity for using APC data to evaluate the impact of uneven passenger loadings on train departures. An 11-month sample of data was obtained from a UK train operating company, covering a large proportion of trains in the fleet for the period April 2013 to February 2014. The three main datasets were as follows:

- Air suspension pressure – Data is automatically collected on the air pressure in the suspension, which gives an approximate passenger loading for each train. A large historic sample of this data was analysed to assess to what extent the occupancy on-board trains is uneven by time of day and along each route. This was compared against the theoretical capacity (as defined in §2.2.1) and each coach was classified based on the occupancy levels. This data was analysed to identify patterns for particular trains and sections of routes. These patterns were summarised to produce a shortlist of stations where departing trains have the most uneven loadings. New metrics were explored to describe loading diversity, using the TRB metrics as a starting point.
- Times the doors lock/unlock – Data is collected for the arrival and departure time for each train service at each station through the times that the doors unlock and lock. This was used to generate a measure for estimated boarding times at each station for each train service.
- Timetable information – Data on scheduled arrival and departure times for each station was also obtained; this included information on service changes, for example on bank holidays and during engineering works. The timetable data was combined with the door locking data to determine for each service whether it arrived and/or departed late.



All three datasets were combined and analysed, in order to assess which stations have higher than average delay for late-running services. In particular this was correlated with the coach-by-coach occupancy data to identify if there was a relationship between uneven occupancy and longer than average dwell times.

**Rationale:** The datasets obtained were comprehensive in their coverage of most services and most trains over an 11-month period. This had advantages over traditional survey approaches (similar to those in §2.2.4.2) of giving a complete picture of trends rather than a sample for a few stations over a much shorter time period. Another advantage of this approach was the negligible cost of data collection, in that it is automatically produced as a by-product ‘exhaust’ from the suspension system. A disadvantage though was perhaps lower level of detail (e.g. no boarding and alighting flows) and poorer accuracy, compared to door sensor data; however, no door sensors were installed on the train operator’s fleet.

## 4.2 Task 2 – Inter-train loading diversity: analysis of existing data

Task 2 was designed to answer the following research questions:

- **RQ 3 – Quantification of inter-train loading diversity**
  - (a) *How did load factor vary by station, time of day and other attributes?*
  - (b) *What metrics are appropriate to quantify the inter-train loading diversity?*
- **RQ 4 – Prediction of inter-train loading diversity**
  - (a) *What techniques can be applied to predict load factor, based on historic data?*
  - (b) *What was the accuracy, precision and recall of these predictors?*

The analysis in Task 1 considered the extent of coach-by-coach crowding. Task 2 built on this previous analysis through analysing the variability and predictability of the trends for inter-train loading diversity across different predictor variables.

The review of solutions identified that some train operators use APC data to provide information on the level of crowding for a particular service through a variety of information channels (§2.3.6). The Naïve Bayes technique was implemented on a training dataset for different groups of predictor variables and the accuracy, precision and recall of these classifiers was measured through applying the models to a test dataset. An initial exploration with decision trees was also attempted for one station.

**Rationale:** Similar to Task 1, the benefits for a data analytics approach were mainly in the coverage and low cost of the data. Specifically, for the prediction task there was a good sample size of around 45 weeks of data.

As identified in the literature review in §2.4, there are many techniques that may be applicable to this type of problem. The Naïve Bayes algorithm was chosen, because it was relatively easy to implement and is regarded in the Literature as a good first step for such classification tasks. A requirement of the theorem is that the predictor variables are independent; however, this condition was likely not met. Nevertheless, Naïve Bayes has been shown to produce reasonable results even when the independence condition is not met. In particular running several different Naïve Bayes models is considered a good method for identifying suitable predictor variables.

The review of machine learning techniques identified decision trees as another popular approach for classification tasks. There exist several off-the-shelf software packages to implement the many variants of decision tree approaches, some of which are listed in Table 2. As an initial investigation, a decision tree was implemented from first principles for one station only; the order in which attributes were split was determined by maximising the information gain at each split, i.e. calculating the entropy for each option and choosing the feature that resulted in the largest reduction in entropy. This was not intended as a robust analysis; rather it was purely a preliminary investigation into how this approach works.

### 4.3 Task 3 – Perceived causes and effects of loading diversity, both inter-coach and inter-train

This task was designed to answer the following research questions:

- **RQ 5 – Perceived causes and effects of loading diversity**
  - (a) *From the perspective of frontline staff, which services had high levels of inter-coach local crowding?*
  - (b) *From the perspective of frontline staff, which services had high levels of inter-train local crowding?*
  - (c) *What were the perceived causes of inter-coach local crowding?*
  - (d) *What were the perceived causes of inter-train local crowding?*
  - (e) *What were the perceived effects of crowding, both inter-coach and inter-train?*

A focus group was held with nine of the train operator's frontline staff, seven were on-train staff and two were platform staff. This provided qualitative evidence on the causes and effects of local crowding, both inter-coach and inter-train. A secondary purpose of the focus group in Task 3 was to give a qualitative validation of the scale and location of crowding in relation to the more quantitative findings from Tasks 1 and 2.

The topic guide for the focus group was as follows:

- Welcome and introduction (10 minutes) – personal introduction from everyone.
- Experience of local crowding on trains (10 minutes) – the extent to which during their day-to-day work that some parts of trains are crowded while there are seats available elsewhere and also the variability they see in crowding across different times of day.
- Causes (20 minutes) – the main factors that lead to uneven occupancy, both coach-by-coach and across the peak period.
- Effects (20 minutes) – who is affected and how for different levels of crowding on trains.
- Suggestions for improvement (15 minutes) – ideas for solutions, working in two groups then report back to the whole group.

**Rationale:** Frontline staff use the trains every day and so have first-hand experience of the factors surrounding crowding. A focus group was considered a good method for drawing on this experience, because it would provoke the sharing of experiences and discussions from different perspectives. This method was preferred to a paper-based survey, because it was felt more appropriate in exploring and understanding relevant topics in depth.

### Summary

Three tasks were defined in order to answer the five main research questions. The methodology for each of these tasks was set out and justified. Task 1 (RQ 1 and 2) and Task 2 (RQ 3 and 4) used a data analytics approach to manipulate three large datasets on air suspension, door locking and timetables. Task 3 (RQ 5) comprised a more qualitative approach through conducting a focus group with nine frontline staff.

## 5 Findings: occupancy data

This section presents the findings for the first two tasks outlined in the methodology section:

- Task 1 – Inter-coach loading diversity: analysis of existing data (§5.1)
- Task 2 – Inter-train loading diversity: analysis of existing data (§5.2)

Lower-level section headings are used for each of the main research questions that were posed in §3.3, each with a summary at the end of the section.

### 5.1 Task 1 – Inter-coach loading diversity: analysis of existing data

Task 1 answered two groups of research questions; as discussed in §4.1, this was achieved through analysis of data on air suspension pressure, the times the doors locked/unlocked and timetable information.

- RQ 1 – Quantification of inter-coach loading diversity (§5.1.1)
- RQ 2 – Link between inter-coach loading diversity and dwell times (§5.1.2)

#### 5.1.1 RQ 1 – Quantification of inter-coach loading diversity

##### 5.1.1.1 (1a) Is it possible to identify coach-by-coach occupancy using air suspension data?

##### *The air suspension dataset*

The air suspension dataset was automatically collected for the 11-month period, April 2013 to February 2014. This consisted of two readings of air pressure for each coach, taken just prior to the doors closing on departing each station. Each instance also included: the vehicle ID; the number of coaches; the train service route ID; the time stamp to the nearest second; and a record of the station for which it was departing. Data was available for a large proportion of the train operator's fleet. There were two lengths of equipped trains, which for the purpose of this thesis are referred to as long and short.

### ***Cleansing the air suspension data***

Initial cleansing involved removal of instances with an invalid number of coaches. The reason for these errors was unknown, although only a small proportion of the dataset was affected (0.03%).

Instances were also removed where the previous station was 'NULL'. This represented a relatively large proportion of the raw dataset (10.87%). However, these records were not actual observations; for example they represented instances at the depot etc.

After the two cleansing steps above, the dataset was used to generate estimates of passenger counts, as described below. This version of the dataset is referred to as 'Sample 1'. As will be discussed later, the air suspension data was fused with timetable data and door locking data. The combined dataset of air suspension data and timetable data is referred to as 'Sample 2', whereas the combined dataset of air suspension data, timetable data and door locking data is referred to as 'Sample 3'.

### ***Generating the passenger counts***

As discussed above, two pressure sensors were present for each coach. The distribution of air pressure for each sensor was plotted, primarily to calculate the minimum air pressure observed for each coach. Specifically, the 2<sup>nd</sup> percentile, 25<sup>th</sup> percentile, 75<sup>th</sup> percentile and 98<sup>th</sup> percentile were plotted for each coach.

The 2<sup>nd</sup> percentile was used as a proxy for the population minimum in order to reduce the influence of erroneous data. It is known that some services run with very low numbers of passengers, for example some late night services. As such, it was reasonable to assume that the 2<sup>nd</sup> percentile would be close to the population minimum; i.e. that in at least 2% of instances a coach is either empty or very close to empty. The two air pressure sensors within a coach do not necessarily give the same reading, although for some coaches the two 2<sup>nd</sup> percentile values were similar. The train operator recommended taking the average of the two sensors in each coach to get one reading per coach.

The front and rear coaches were the heaviest and as expected had the highest air pressure. More generally, the ratio of the tare weight (i.e. the weight with no passengers on board) to the 2<sup>nd</sup> percentile values were relatively similar for all coaches, suggesting that the minimum pressure estimates were consistent. Although

the estimated minimum values in each coach were similar across the fleet, they were calculated for each individual vehicle to account for subtle variations in vehicle weights.

To generate the estimated passenger counts, the ratio was taken of the air pressure on departure to the minimum air pressure for each coach on each vehicle. A linear relationship between weight and pressure was assumed; i.e. the percentage increase in air pressure was multiplied by the tare coach weight to generate an estimated increase in weight as a result of the passengers on board. This was in turn divided by the average weight of a passenger with luggage to generate an estimated passenger count.

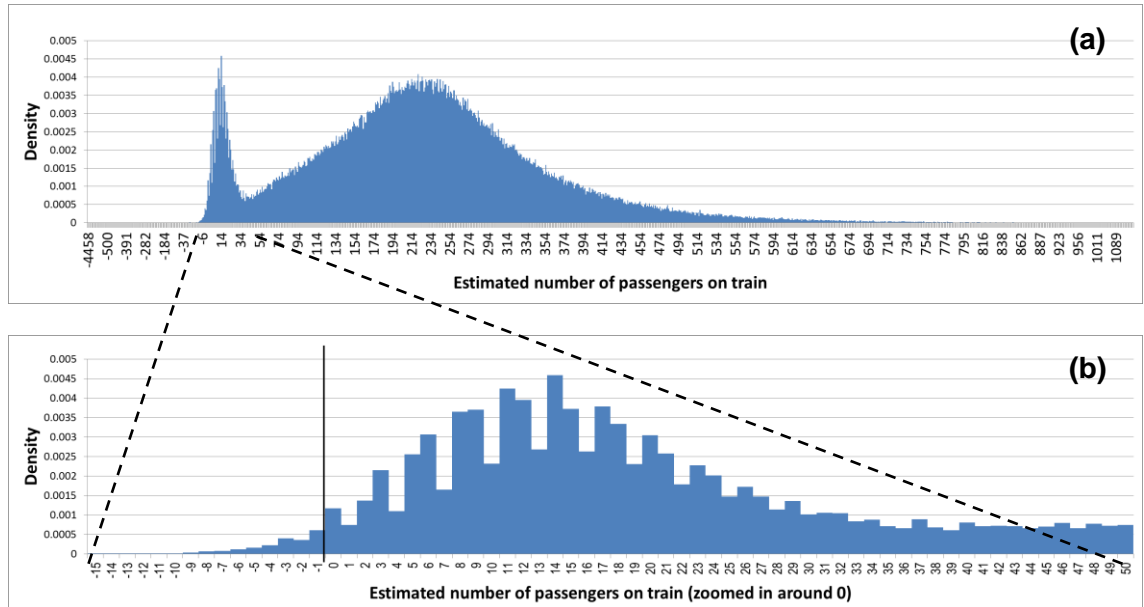
It was assumed that the average weight of a passenger with luggage was 80kg. This value was suggested by the train operator as the value they had used in the past. The standard value used by the airline industry for a passenger with luggage, both hand luggage and in the hold, is 100kg (Stockholm Environment Institute n.d.). It may be expected that an airline passenger would have more luggage than a rail passenger and thus this likely forms an upper bound. Other estimates by the ONS (2010) suggested that for England and Wales, the average weight of men and women was 83.6kg and 70.2kg, respectively. A passenger train would carry a mixture of adults and children, with varying amounts of luggage and so without further quantitative information 80kg seems to be a reasonable assumption in light of these other estimates.

The 80kg assumption has a large influence on the absolute accuracy of the estimated passenger counts. However, the primary focus of this research was to investigate localised crowding, both inter-coach and inter-train and as such the relative levels of crowding were of greater interest than the absolute levels of crowding.

The air pressure reading is recorded a few seconds before the doors close and so this provides a snapshot of the distribution of coach-by-coach loading on departure from each station. Another consideration regarding accuracy is that some passengers move along a train between stations, meaning that crowding in a coach on departure may not translate directly into crowding during a journey.

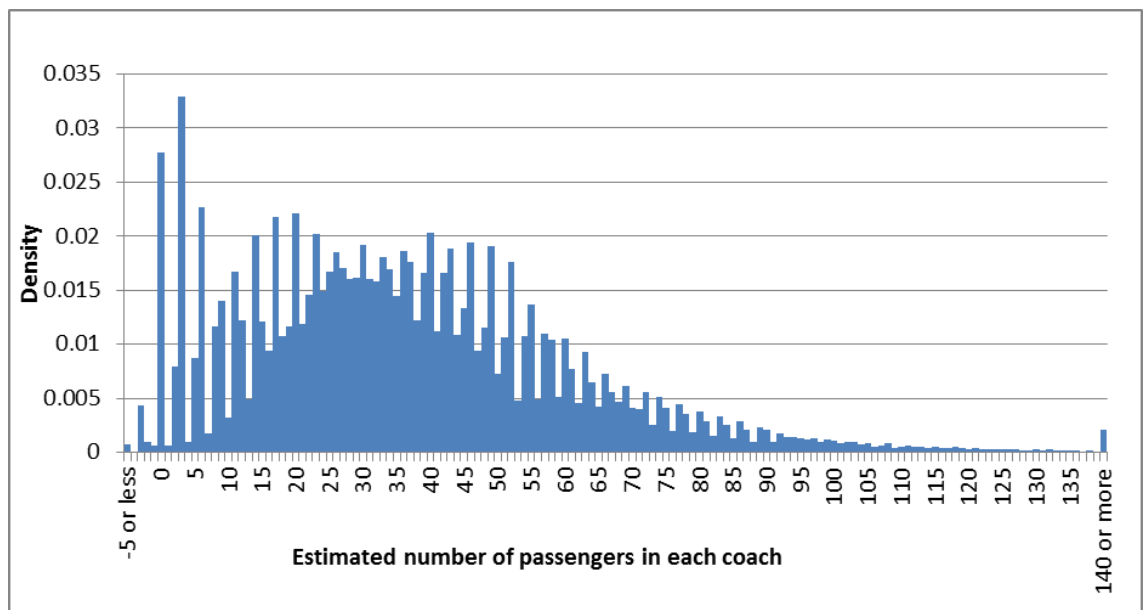
### ***Distribution of passenger counts***

Figure 11a shows the distribution of the estimated number of passengers on all trains. There were two peaks, one around the median value of 222 and a small peak around zero. Only a small proportion of observations gave negative estimates (0.25%); Figure 11b zooms in around the Origin to illustrate this.



**Figure 11 – Distribution of estimated number of passengers on train upon departure (Sample 1)**

Figure 12 shows the distribution of the estimated number of passengers in a particular coach for the whole sample. There were at least two peaks, one around the median value of 34 and another peak between approximately 0 and 20. Some coaches had higher capacity than other coaches and thus there were perhaps more than two peaks, because all coaches were plotted in the same graph. The shape of the distribution had large jumps between subsequent numbers, which was explained by the relatively low level of precision (i.e. the rounding) of the air pressure data. Only a small proportion of observations gave negative estimates (0.67%); these were set to be zero.



**Figure 12 – Distribution of estimated number of passengers in each coach upon departure (Sample 1)**



For the purpose of this study, the capacity of each coach was taken to be the number of seats, i.e. no standing allowance (see §2.2.1). The estimated passenger count was divided by the capacity for each coach, thus giving a ‘coach load factor’. Figure 13 shows the distribution of the estimated coach load factor (pooled for all coaches), banded to the nearest 5%. There were two peaks, one around the median, which fell into the 55%-60% class, and another peak around zero. This now takes into account the fact that different coaches had different capacities, unlike in Figure 12.

The observations to the right of the black line represent instances where a particular coach had more passengers than seats were available; this was the case for 14.7% of instances. As discussed in §2.2.1, the formal definition of capacity does not include coach-by-coach variations. These observations were therefore not necessarily ‘over capacity’, in that other coaches on the same train may have had spare capacity; this is explored in §5.1.1.2.

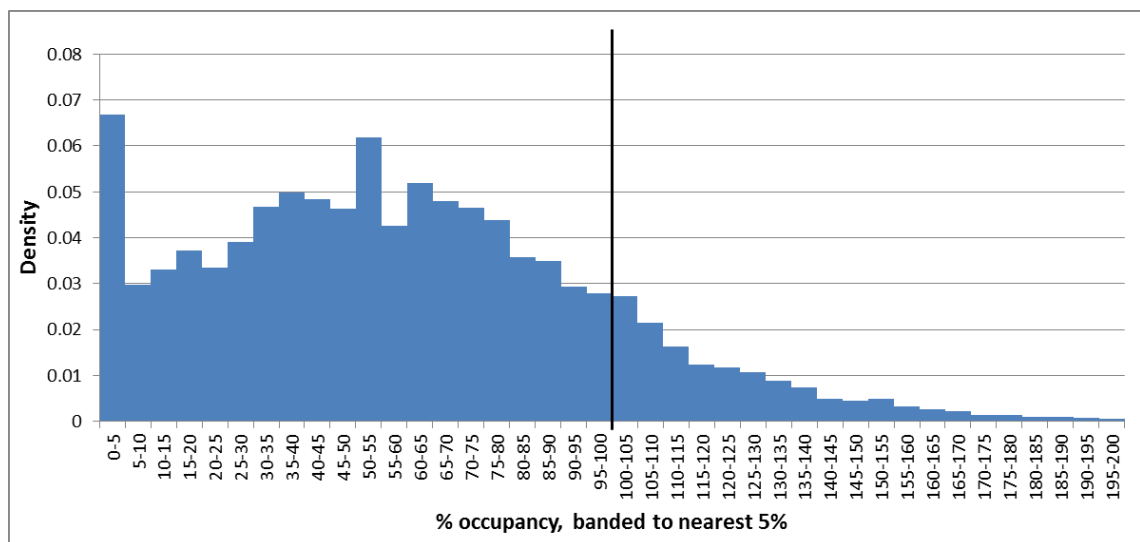


Figure 13 – Distribution of occupancy in each coach upon departure (Sample 1)

### Thresholds

For later use in visualisations and predictive models, it was decided to partition the coach-by-coach sample into a small number of classes. The quantitative definitions of crowding in §2.2.1 would suggest a binary classification of ‘0% to 100%: Within capacity’ and ‘>100%: Over capacity’ would be sufficient.

As was discussed in §1.1, a literature review by RSSB revealed that while some definitions related to objective elements such as density and the available space, other definitions focused on more subjective elements on the perception of crowding. Qualitative research by RSSB (2005) suggested that four classes of perceived

crowding were required to distinguish the different effects of crowding. In a similar manner Cox et al. (2006) suggested that perceptions of crowding were perhaps more complicated than a binary system when they stated that, “Passenger density may not be simply and linearly related to perceptions of crowding”.

As discussed in §2.3.1, the Dutch pilot provided real-time crowding information on current coach-by-coach levels of crowding on the approaching train through a ‘Red-Amber-Green’ system. Similarly as discussed in §2.3.6, SBB, C2C and SWT each provide passenger information on crowding levels for different trains through a variety of channels (online, smartphone, posters), each also distinguishing three levels of crowding.

This thesis is focused primarily on passenger counting data and therefore predominantly the more quantitative aspects of crowding and ‘passenger density’. However as discussed above, various sources from the literature review suggest that three or perhaps four levels are required to sufficiently describe the nature of crowding.

It was decided to split the ‘0% to 100%: Within capacity’ class down further to define two sub-classes: ‘Plenty of seats’ and ‘Nearing capacity’. Both of these classes are still technically ‘Within capacity’, although the distinction is that for ‘Nearing capacity’ although all passengers have an available seat, it may ‘feel more crowded’. The threshold of 75% was taken to distinguish between ‘Plenty of seats’ and ‘Nearing capacity’. Using the National Transport Model as a relative comparison, although not directly applicable, cars on links in conditions above 80% of theoretical capacity are defined as travelling in “very congested conditions” (DfT 2013d).

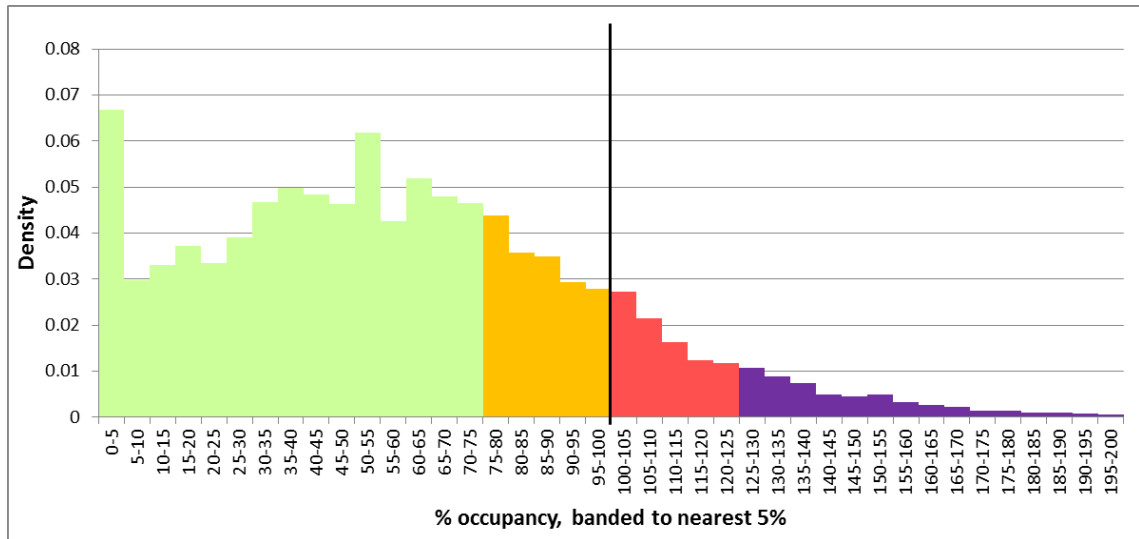
It was decided to split the ‘>100%: Over capacity’ class down further to define two sub-classes: ‘Some people standing’ and ‘Many people standing’. Both of these classes are still technically ‘Over capacity’, although the purpose was to go some way to distinguish between different severities of crowding. The value of 125% was chosen as the threshold.

For illustration purposes, these four classes for coach-by-coach occupancy were each assigned a colour and a qualitative name:

- 0% to 75%: (green) ‘Plenty of seats’ – 68.1% of the sample
- 75% to 100%: (amber) ‘Nearing capacity’ – 17.1% of the sample
- 100% to 125%: (red) ‘Some people standing’ – 8.9% of the sample

- Over 125%: (purple) 'Many people standing' – 5.8% of the sample.

The sample as in Figure 13 is shown again in Figure 14, but colour-coded and partitioned into the four classes of observation. From visual inspection, this suggests that the proposed thresholds give an appropriate split of the sample in order to illustrate the variability in crowding across the network through various visualisations, as will be explored in §5.1.1.2.



**Figure 14 – Distribution of occupancy in each coach upon departure, with four classes (Sample 1)**

These four proposed classes were compared with other classification schemes in Table 3; limited information was available for the quantitative thresholds used by the other schemes, although this comparison was attempted using the qualitative descriptions.

**Table 3 – Comparison of crowding classifications**

(RSSB 2005) (on-train)	“No crowding – Available seats, available luggage space, unrestricted movement throughout train.		Low to moderate crowding – Standing for short periods of time (less than 20 minutes), available handholds, luggage space taken but not obstructing movement, some restricted movement in moving down train.	Severe crowding – Standing for moderate periods (20 minutes to 1 hour), small amounts of luggage restricting movement, cannot get catering trolley through train, train delayed at station due to alighting and boarding times.	Unacceptable crowding – Standing for long periods (>1 hour), standing for short periods without access to handhold, luggage obstructing movement throughout train, passengers or staff cannot move through train, cannot shut train doors.”
(RSSB 2003) (on-platform)	'No crowding – All of body visible'		'Low to moderate crowding – Only body and head visible'	'Severe crowding – Only shoulder and head visible'	'Unacceptable crowding – Only head visible'
(DfT 2013a)	0% to 100%: Within capacity		>100%: Over capacity		
(SBB n.d.)	Low/medium occupancy expected (Level 1)		High occupancy expected (Level 2)	Very high occupancy expected (Level 3)	
(SWT 2012)	Seats available (green)		All seats taken (yellow)	Limited standing (red)	
(C2C n.d.)	Seats normally available (green)	Few seats normally available (amber)	Standing room only (red)		
Proposed classification	0% to 75%: (green) 'Plenty of seats'	75% to 100%: (amber) 'Nearing capacity'	100% to 125%: (red) 'Some people standing'	Over 125%: (purple) 'Many people standing'	

The proposed four-part classification relates to occupancy in a particular coach, whereas most of the other classifications relate to occupancy on the whole train; however, the proposed classification is perhaps closest to the C2C classification, in that amber relates to 'few seats normally available'. On the contrary the SWT classification takes the amber class to mean 'All seats taken'; likewise SBB combine 'low' and 'medium' occupancy into the first level. The variability between the different classifications is perhaps because different train services have different definitions of capacity, i.e. due to the amount of 'standing allowance', as discussed in §2.2.1.

The RSSB qualitative classifications focus on the higher end of the spectrum for 'unacceptable crowding'. Other Red-Amber-Green classifications were observed in §2.3.1, although no information was available on either the quantitative or qualitative thresholds used and so they were not used in the comparison.

#### ***Data fusion with timetable data: 'Sample 2'***

A limitation of the above analysis was that it combined both directions of travel<sup>5</sup>. To overcome this limitation and break down the results by particular routes, the data was combined with timetable data. This fused dataset is referred to as 'Sample 2'.

The timetable dataset consisted of one row for all of the train operator's departures spanning the same 11-month period<sup>6</sup>. Information was available for each train service on arrival and departure times for all calling points. This included changes to the published timetables, due to bank holidays and engineering works etc. A unique ID was created by combining the train service route ID, the date and the departure station.

Of 'Sample 1', it was possible to join 75.18% of the air suspension data with the timetable data. A further 2.88% consisted of multiple observations, whereby there was more than one instance for each departure; for these, only the most recent observation was retained. Data was also available on services that were travelling in 'reverse', which represented a further 1.42% of observations; these reversed services were also removed from Sample 2. Although these data cleansing steps removed a relatively

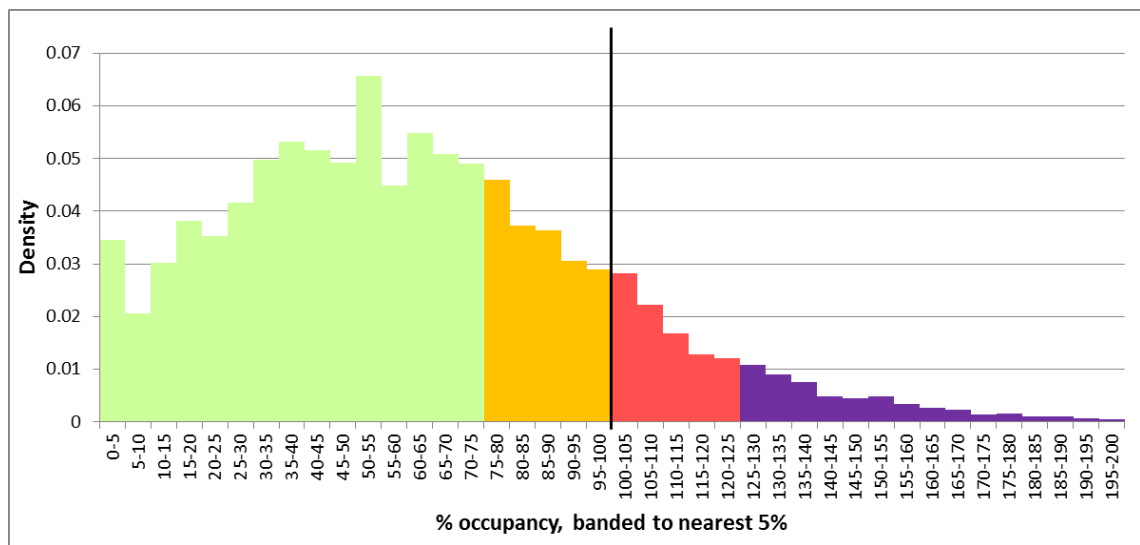
---

<sup>5</sup> With the exception of terminus stations, where there was only one direction for departing trains

<sup>6</sup> Although timetable data was not available for a two-week period in June and an eight week period from mid-September to mid-November

large proportion of the dataset, given that there was initially an 11-month sample this still represented a sufficiently large dataset.

It was understood that the train service ID in the air suspension dataset is input by the driver and so it was likely that the unmatched records were because of a manual input error. It was expected therefore that these manual errors were randomly distributed. The multiple observations of the same departure were predominantly in the early morning, suggesting that this error may have been related to initialising the system in the morning, in which case they would be genuinely erroneous data. The trains travelled in reverse on occasion for operational reasons; it was understood that these instances were randomly distributed. The sample as in Figure 14 is shown again in Figure 15, but for the fused dataset, Sample 2.



**Figure 15 – Distribution of occupancy in each coach upon departure, with four classes (Sample 2)**

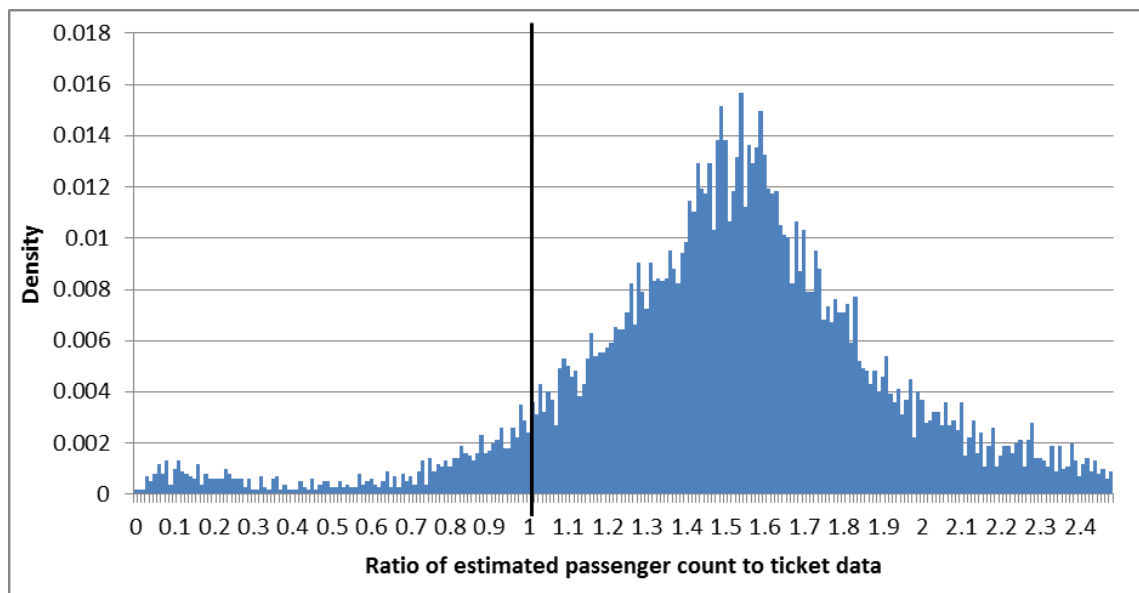
The distribution of Sample 2 was very similar to Sample 1, with the median being within the 55%-60% class. The main difference was that the peak around 0 was less pronounced; this was as a result of the removal of the multiple observations in the early morning. The proportion of the sample in each of the four classes was as follows:

- 0% to 75%: (green) 'Plenty of seats' – 67.0% of the sample
- 75% to 100%: (amber) 'Nearing capacity' – 17.9% of the sample
- 100% to 125%: (red) 'Some people standing' – 9.2% of the sample
- Over 125%: (purple) 'Many people standing' – 5.9% of the sample.

### Validation

The estimated passenger counts above gave ‘sensible’ results, in that only a small proportion of estimates were negative. This suggested that the estimates for the minimum air pressure for each coach worked well.

It was not possible to conduct manual calibration / validation passenger counts within the scope of this study. However, counts from ticket data were available for some services in February 2014, although it is understood that these ticket counts did not contain all ticket types. The ratio of the estimated passenger counts against the ticket data is shown in Figure 16, i.e. instances greater than 1 (to the right of the black line) represent situations where the estimated passenger count was higher than the ticket data.



**Figure 16 – Comparison of estimated passenger counts against ticket data (subset of Sample 2, where it was possible to match to ticket counts; n=9,964)**

The median was 1.53, i.e. suggesting that on average the estimated passenger counts were 53% higher than the ticket counts. This result was as expected, because the ticket data did not include all types of available ticket. This did not fully validate the passenger counts, although it did go some way to suggest that the estimates were reasonable. As discussed previously, the primary focus of this research was to investigate localised crowding, both inter-coach and inter-train. As such the relative levels of crowding were of greater interest than the absolute levels of crowding and so the validation against ticket counts gave sufficient confidence in the data quality.

For future work, manual validation counts could be collected to give greater confidence in the findings from the air suspension occupancy data. This could be achieved through



surveyors collecting passenger counts while walking along the length of the train. An alternative method could be to collect cordon counts of the doors and vestibules, either manually from CCTV data or automatically from door sensors where these are installed.

#### **5.1.1.2 (1b) Which train services and stations had the most uneven inter-coach loadings?**

##### ***Methods for viewing the counts: 1. Cumulative distribution of occupancy***

The cumulative distribution of occupancy for all stations and all times of day was plotted for each coach, separately for short and long trains. The cumulative distribution graphs provided a good overview, although it was required to break this down to investigate variations for different stations and different times of day.

A macro spreadsheet utility was developed to automatically generate PowerPoint animations showing how the cumulative distribution of each coach varied throughout the day at a particular station. This tool combined both directions of travel and binned the data to the nearest hour of actual departure. The slideshow visualisations consisted of 24 graphs for each station, one for each hour of the day. These were a useful method for illustrating how the occupancy varied throughout the day, in particular the relative occupancy between different coaches.

The visualisations showed that for particular stations, it was often the same coaches that had the highest occupancy throughout the day. These trends were station-specific, i.e. for some stations it was usually one end of the train that was busiest, whereas for other stations it was the middle of the train where this was the case. The times of day and sections of network that would be expected to be busy proved to be so, suggesting that the data gave sensible results.

##### ***Methods for viewing the counts: 2. Proportion of observations with occupancy > 100%, by time of day for a particular station***

From the above visualisations, it became apparent that by taking a cross-section of the cumulative distribution it was possible to condense the information. Specifically, taking the proportion of observations with occupancy greater than 100% (i.e. at least some people standing in a coach) was sufficient to capture the key information on the inter-coach loading diversity. The spreadsheet utility was adapted to generate one graph for

each station, with a collection of bar charts for each hour of the day. This gave the same insights as before, although in a more concise format.

***Methods for viewing the counts: 3. Proportion of observations with occupancy > 100%, for all times of day combined***

From the above analysis it was clear that there were similar trends in the relative occupancy of coaches throughout the day, although in the peak periods this was amplified with more instances of higher occupancy in all coaches. As such it was possible to combine results for all times of day to summarise the inter-coach patterns for each station.

The preliminary analysis suggested that there was uneven occupancy on many train services and that these trends were station-specific. In summary, from visual inspection:

- Two stations were identified where on departure typically the middle of the train was more often over capacity compared to other parts of the train.
- Eight stations were identified where on departure typically one end of the train was more often over capacity compared to other parts of the train.
- All other stations had relatively fewer instances where coaches had occupancy over capacity.

These trends were more pronounced for long trains, although also present to a lesser extent for short trains.

***Methods for viewing the counts: 4. 'Green-Amber-Red-Purple' average occupancy 'Heat Map' visualisations, by train service and direction***

As discussed in §2.2.3.3, Itoh et al. (n.d.) proposed the concept of a 'Heat Map view' in which colours are used to show whether occupancy is higher or lower than historic trends; the horizontal axis represents time and the vertical axis represents the different lines and stations. This concept was applied to the train operator's datasets and modified to represent inter-coach loading, so that the vertical axis still represented the stations, but the horizontal axis represented the different coaches.

A macro spreadsheet utility was developed to automatically generate visualisations for the average weekday occupancy coach-by-coach. Green indicated occupancy between

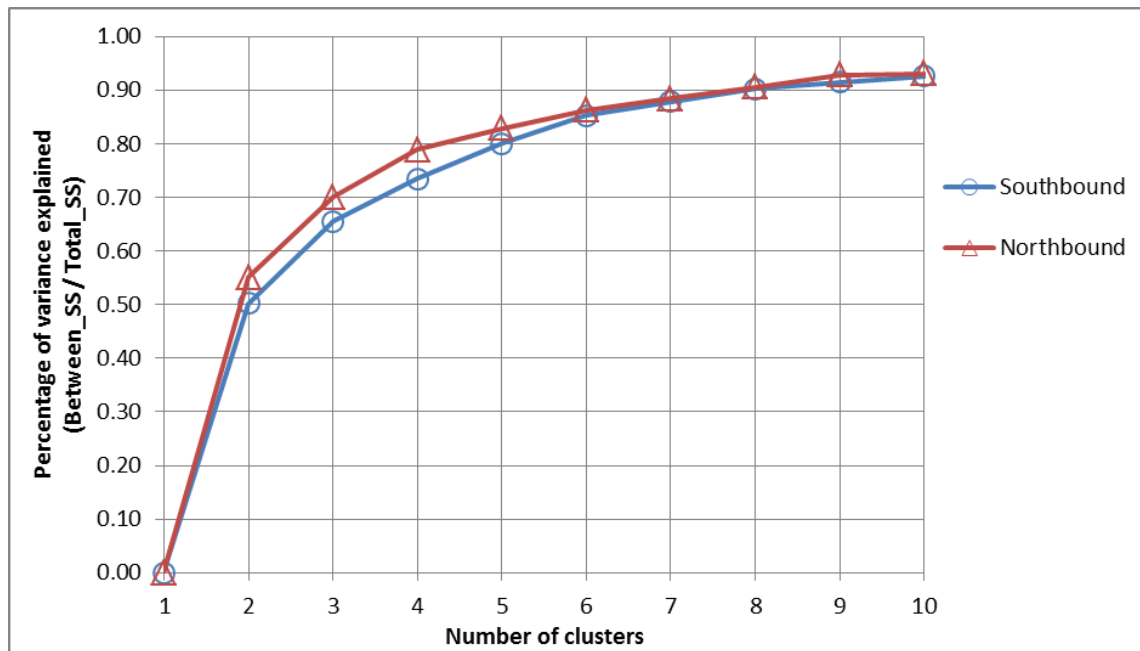
0 and 75%; orange indicated occupancy between 75 and 100%; red indicated occupancy between 100 and 125%; and purple indicated occupancy greater than 125%. Such outputs were generated for all routes and services on 'Sample 2', distinguishing the direction of travel. The benefit of this approach was to consider the inter-coach loading diversity in the context of preceding and subsequent stations.

It was found from visual inspection of the heat maps that patterns of inter-coach loading diversity were often similar for neighbouring stations. A more robust cluster analysis was also undertaken, as discussed below.

### ***Cluster analysis***

A cluster analysis was undertaken on the average coach-by-coach occupancy data to identify stations with similar patterns in a robust manner. This was done for long trains for southbound and northbound services separately.

The k-means method was chosen for the cluster analysis and was implemented multiple times with the number of clusters varying from  $k=2$  to 10. For each value of  $k$ , the algorithm was run ten times, with a randomly selected set of starting centroids each time. For each run, the percentage of variance explained (i.e. the ratio of the between-group variance to the total variance) was captured. The average percentage of variance explained across the ten repetitions is presented in Figure 17; this suggested that six clusters in the southbound direction gave sensible results, i.e. with a slight 'elbow' at  $k=6$ . For the northbound direction, the 'elbow' was not as pronounced, although after visual inspection of the cluster output (for  $k=4, 5$  and  $6$ ), five clusters was chosen as the most suitable partition. The results for some clusters changed slightly depending on the initial centroids; the cluster selection that gave the highest percentage of variance explained was chosen as the final partition.



**Figure 17 – Percentage of variance explained for each number of clusters tested (average across all ten repetitions)**

In the southbound direction three of the six clusters had some coaches with average occupancy over 70% with other coaches below 55%, while for northbound this was true for two of the five clusters; 15 stations were assigned to these clusters for southbound services and only seven stations for northbound services, which suggested that uneven loadings may be a more common issue for southbound trains, compared to northbound trains.

For southbound services, Cluster A1 comprised stations where on average the middle of the train was quite busy, while the rear of the train was quiet; the opposite was true for Clusters A2 and A3, with varying levels of average occupancy.

- Cluster A1 had average occupancy around 75% towards the middle of the train and 50% towards the rear of the train;
- Cluster A2 had average occupancy around 80% towards the rear of the train and 50% towards the middle of the train;
- Cluster A3 had average occupancy around 75% towards the rear of the train and 30% towards the middle of the train, which was the greatest range (i.e. 'most uneven') of all the clusters;
- Cluster A4 had a similar profile to Clusters A2 and A3, although was typically less busy with average occupancy around 55% towards the rear of the train and 30% towards the middle of the train.

- Clusters A5 and A6 comprised services with relatively even average occupancy across all coaches, with those stations in Cluster A5 being typically busier than those in Cluster A6.

For northbound services, Cluster B1 comprised stations where on average the middle of the train was quite busy, while other parts of the train were quiet; the opposite was true for Cluster B2. There were often similar stations in the respective southbound and northbound clusters, which suggested that for some but not all stations there were similar patterns in both directions. This result is perhaps surprising; possible causes of such patterns are discussed later in §6.1.1.3.

- Cluster B1 had average occupancy around 70% towards the middle of the train and 50% on other parts of the train;
- Cluster B2 had average occupancy around 80% towards the front of the train and 50% towards the middle of the train;
- Cluster B3 had average occupancy around 50-60% along the length of the train with relatively even occupancy across all coaches;
- Cluster B4 had average occupancy around 50% towards the front of the train and 30% towards the middle of the train.
- Cluster B5 comprised services with relatively low average occupancy across all coaches.

The 15 stations within Clusters A1, A2 and A3 and the seven stations within Clusters B1 and B2 were similar to those identified from visual inspection of previous analysis, although here the findings were based on average occupancy, rather than the proportion of observations with occupancy over 100%.

The stations within particular clusters were often neighbouring stations on the same route. This finding that neighbouring stations had similar loading profiles was not a surprise, because in a situation where the net number of passengers alighting is relatively low, if a coach is busy on arrival at a station it will likely also be busy on departure. An important thing to note from this is that if a station has particularly uneven loadings, initiatives at that station in isolation may not be sufficient; i.e. if the problem is being caused by behaviours at a preceding station, it would be better to tackle the root cause of the problem 'upstream'. In other words prioritising effort to

provide solutions at those stations with uneven loadings and high boarding flows may also alleviate crowding at several subsequent stations ‘downstream’.

A limitation of the analysis was that no information was available on boarding and alighting flows. Without information on boarding and alighting flows, it is challenging to identify which stations were the root cause, i.e. at which stations the effort should be targeted for solutions to smooth the crowding.

#### 5.1.1.3 (1c) For what proportion of train departures were there available seats in at least one coach, while passengers were standing elsewhere on the train?

##### *All services*

All departing train services were partitioned into one of three categories:

- ‘Some coaches had standing passengers while other coaches had spare seats’;
- ‘All coaches had only seated passengers’;
- ‘All coaches had at least some standing passengers’.

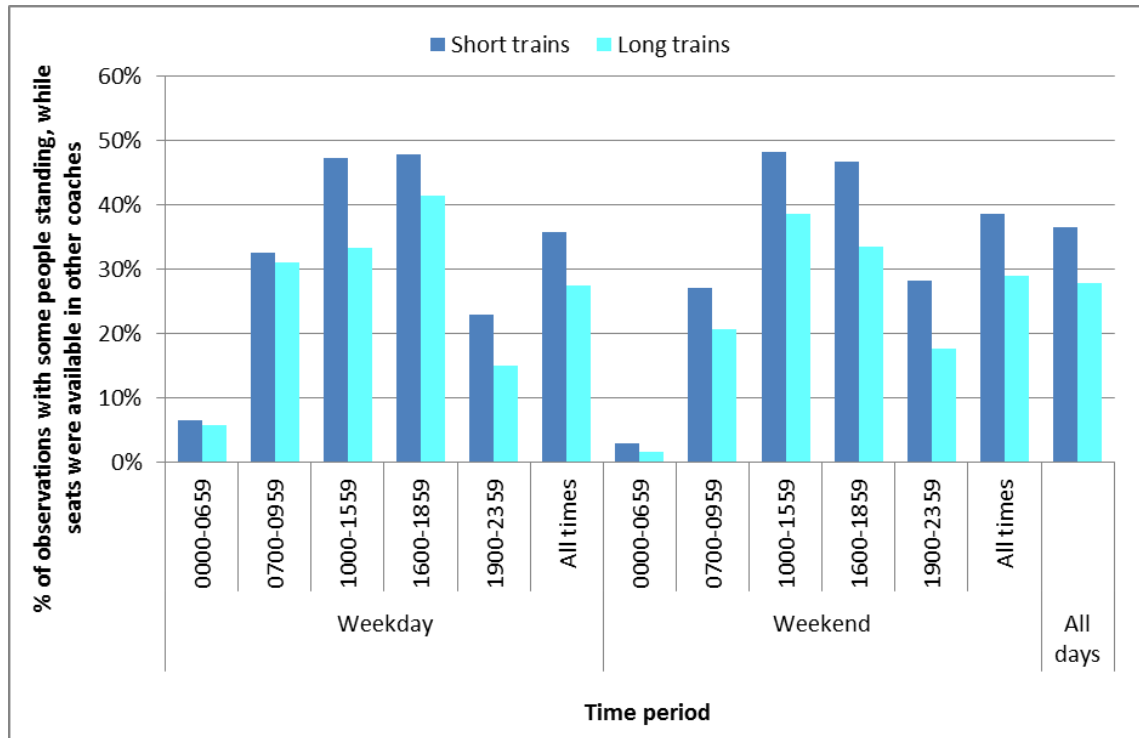
The first category was of most interest, because this was a useful headline figure to highlight the extent to which uneven inter-coach loading was a problem.

As shown in Table 4, 36% of short train services and 28% of long train services had available seats in at least one coach, while passengers were standing in at least one coach elsewhere on the train. A headline figure therefore was that on around a third of services there were at least some passengers standing who could have been sitting down.

**Table 4 – Proportion of observations where some coaches had standing passengers while other coaches had spare seats (Sample 2)**

Classification	Short trains	Long trains	Short trains	Long trains
Some coaches had standing passengers while other coaches had spare seats	36.5%	27.9%	31,911	38,708
All coaches had only seated passengers	56.2%	70.2%	49,208	97,300
All coaches had at least some standing passengers	7.3%	1.9%	6,391	2,686
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>87,510</b>	<b>138,694</b>

This analysis was broken down by time of day and whether weekday or weekend, as shown in Figure 18. This suggested these instances were more often in the 10:00-15:59 and 16:00-18:59 periods.



**Figure 18 – Proportion of observations where some coaches had standing passengers while other coaches had spare seats, by time of day (Sample 2)**

### ***Busy services***

Because uneven boarding has little impact on the more lightly loaded trains, more detailed analysis was carried out on 'busy' services. Busy services were defined as being those services departing with load factor over 75%. These 'busy services' represented 48% of short train services and 21% of long train services.

In summary, 70% of busy short train services and 83% of busy long train services had available seats in at least one coach while passengers were standing in at least one coach elsewhere on the train.



**Table 5 – Proportion of observations for ‘busy’ services where some coaches had standing passengers while other coaches had spare seats (Sample 2)**

Classification	Short trains	Long trains	Short trains	Long trains
Some coaches had standing passengers while other coaches had spare seats	70.1%	83.4%	29,174	24,073
All coaches had only seated passengers	14.5%	7.2%	6,047	2,089
All coaches had at least some standing passengers	15.4%	9.3%	6,391	2,686
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>41,612</b>	<b>28,848</b>

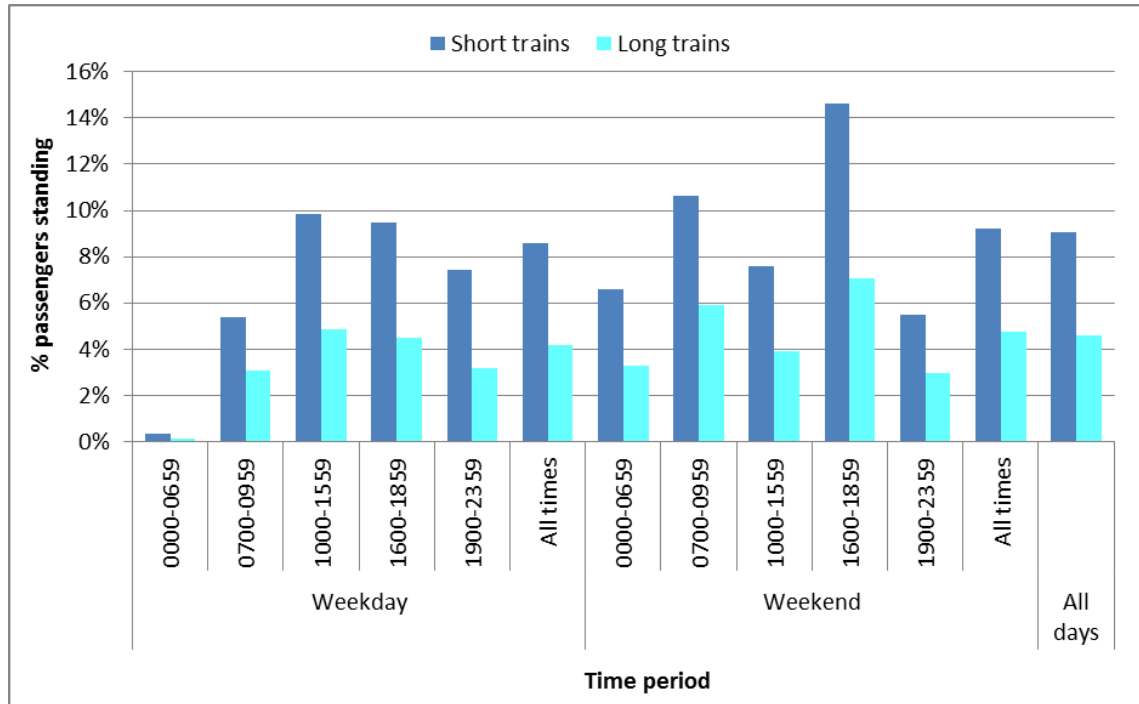
#### 5.1.1.4 (1d) Which stations had the highest proportion of passengers who were standing, but could have been sitting down?

##### *All services*

Further headline measures were defined as follows:

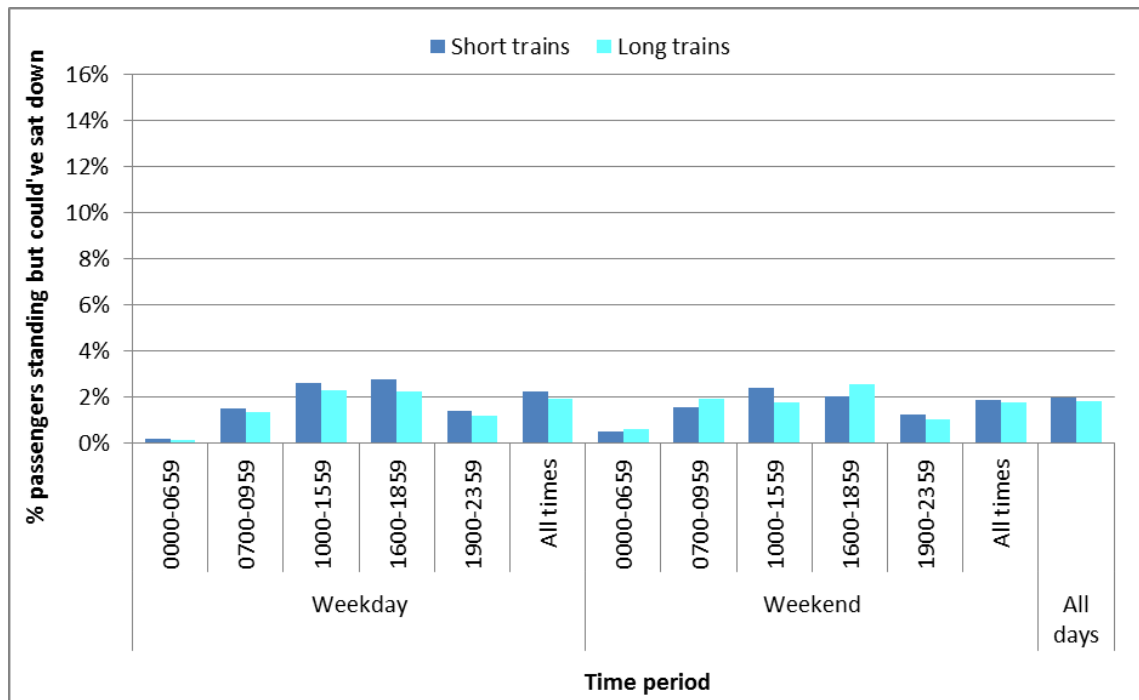
- **The proportion of passengers who were standing.** The estimated passenger counts were compared against the number of seats to calculate the number of standing passengers or number of spare seats in each coach;
- **The proportion of passengers who were standing but could have been sitting down.** This was computed by comparing the number of people standing and number of spare seats across all coaches, until all unused seats were filled.

Overall 9.1% of passengers on short trains and 4.6% of passengers on long trains were standing on departure. It is possible that some of these passengers may have moved along the train to find a seat, although data on this was not available. Figure 19 shows the proportion of passengers who were standing for different times of day; this shows that this was highest for the 1600-1859 period on weekend services.



**Figure 19 – Proportion of passengers who were standing, by time of day (Sample 2)**

Overall for all stations and times of day, 2.0% of passengers on short trains and 1.8% of passengers on long trains were standing on departure, but there were sufficient seats elsewhere on the train for them to sit down. Figure 20 shows that this measure was consistently around 2% for different times of day, with the exception of early morning and late evening services, when this measure was lower.



**Figure 20 – Proportion of passengers who were standing but could have been sitting down, by time of day (Sample 2)**

### ***Busy services***

More detailed analysis was carried out on 'busy' services, i.e. those services departing with load factor over 75%.

For busy services, 13.7% of passengers on short trains and 11.9% of passengers on long trains were standing on departure.

For busy services, 2.9% of passengers on short trains and 4.1% of passengers on long trains were standing on departure, but could have been sitting down. This measure was as high as 9% of passengers for a particular station and 7% for two other stations.

#### **5.1.1.5 (1e) What metrics are appropriate to quantify the inter-coach loading diversity?**

### ***Application of existing metrics***

As identified in the literature review in §2.2.4.1, the TRB (2003) issued guidance that introduced the concept of 'loading diversity', of which they defined three types: "1. Loading diversity within a car"; "2. Loading diversity among cars of a train"; and "3. Unevenness of passenger demand during the peak hour".

For the second type of loading diversity, where there are more passengers in some coaches than other coaches, they defined a metric called 'ratio of car occupancy to train average'. This is a metric where a value equal to 1 represents an individual coach load equal to the average load of all cars in the train. Figure 3 in §2.2.4.1 was an example of this metric for the Toronto subway.

The 'ratio of car occupancy to train average' was computed across all stations. This showed that for short trains typically one particular coach had occupancy 20% lower than the train average. Similarly for long trains, two coaches had occupancy 22% and 14% lower than the train average.

When looking at different time periods throughout the day, the 0000-0659 time period had the largest range of values (from 0.86 to 1.21 for short trains and 0.77 to 1.39 for long trains), i.e. according to this metric the 'most uneven' train services. The value of this metric is somewhat debatable for this time period, because average occupancy was particularly low (around 35% as will be discussed in §5.2.1.1).

More generally, it is the case that a limitation of the existing TRB metric is that it only explains the occupancy relative to other coaches and ignores the absolute occupancy; for example some quiet services would be classified as having very high loading diversity, even though there may be no coaches at or near capacity. Because of this limitation, two new metrics to describe inter-coach loading diversity were proposed.

***New metric 1: ‘Two busiest and two quietest coaches’ inter-coach loading diversity classification***

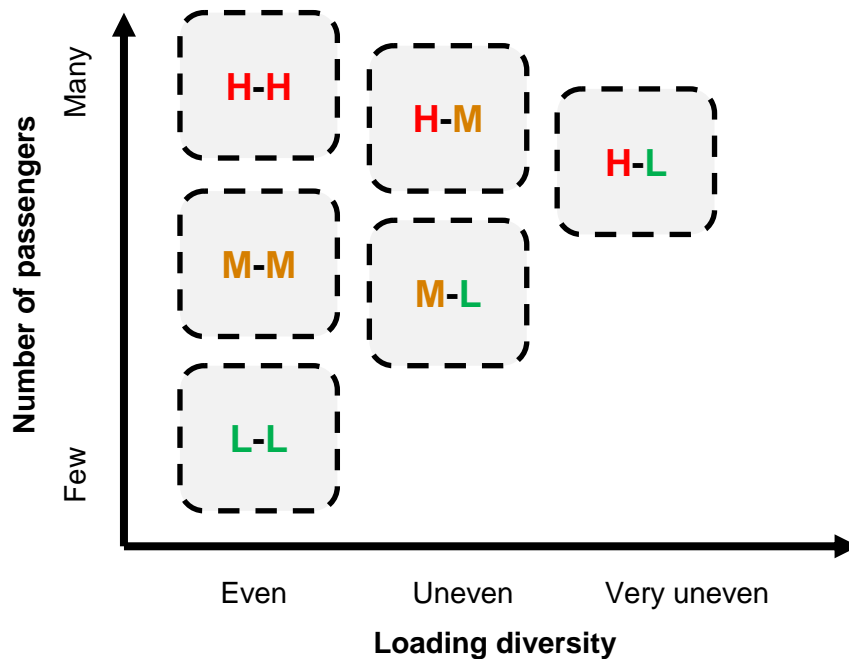
For the first metric, thresholds of occupancy were set as follows:

- Low (L) – 0% to 75%;
- Medium (M) – 75% to 100%;
- High (H) – Over 100%.

Six classes were proposed meeting the conditions as defined in Table 6. Each of the six classes were given a name, such as “X-X”, whereby the first letter related to the two busiest coaches and the second letter related to the two quietest coaches. For example, the most uneven class had at least two coaches with occupancy greater than 100% and at least two coaches with occupancy less than 75% and was named “H-L”. These six classes are also illustrated in Figure 21 with the loading diversity on the horizontal axis and number of passengers on the vertical axis. This proposed new metric overcame the limitation of the existing TRB metric in that it included information relative to the capacity.

**Table 6 – Proposed new metric of ‘Two busiest and two quietest coaches’: definition of six classifications to describe inter-coach loading diversity**

<b>Loading diversity</b>	<b>Name</b>	<b>Number of coaches &lt;75% (L)</b>	<b>Number of coaches between 75 and 100% (M)</b>	<b>Number of coaches &gt;100% (H)</b>
‘Very uneven’	‘H-L’	At least 2	No condition	At least 2
‘Uneven’	‘H-M’	At most 1	At least 2	At least 2
	‘M-L’	At least 2	At least 2	At most 1
‘Even’	‘H-H’	At most 1	At most 1	At least 2
	‘M-M’	At most 1	At least 2	At most 1
	‘L-L’	At least 2	At most 1	At most 1



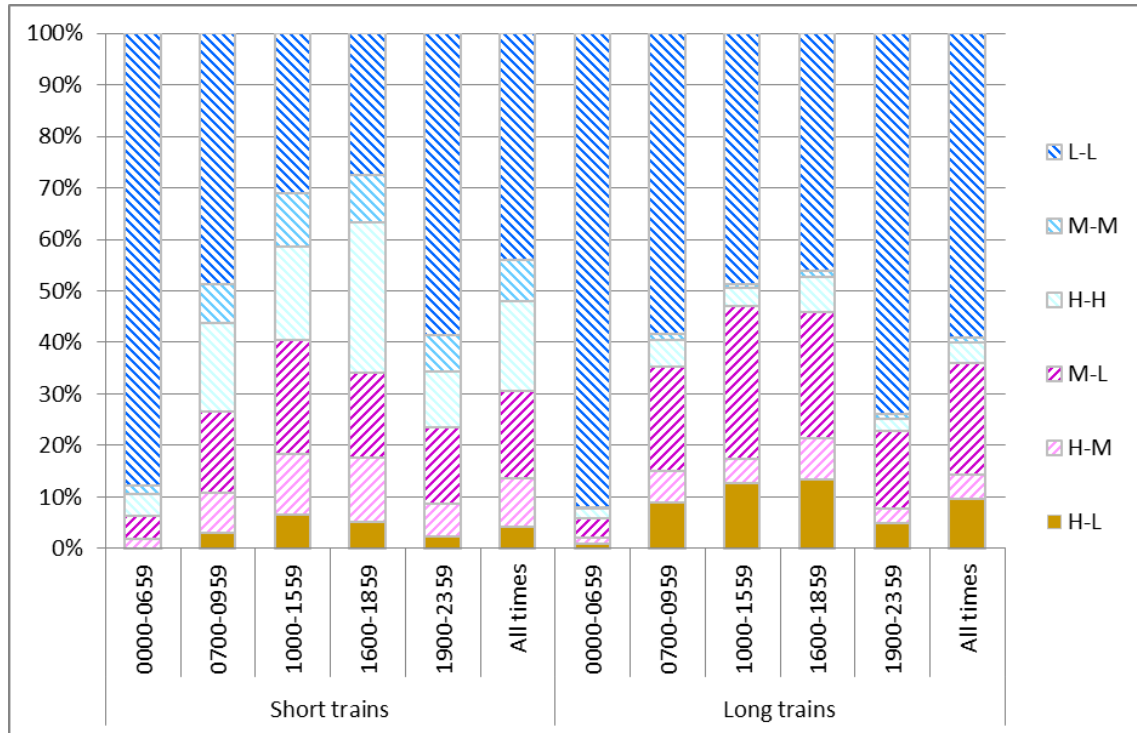
**Figure 21 – Proposed new metric of ‘Two busiest and two quietest coaches’: illustration**

Table 7 gives the partition of the sample into the six classes as defined above. This shows that 9.5% of long train departures were classified as ‘very uneven’ (‘H-L’), i.e. observations where at least two coaches had occupancy less than 75% and at least two coaches had occupancy greater than 100%. This partition will be used again in §5.1.2 when considering average boarding times.

**Table 7 – Partition by ‘Two busiest and two quietest coaches’ metric (Sample 2)**

Classification		Short trains	Long trains	Short trains	Long trains
‘Very unevenly loaded’	H-L	4.3%	9.5%	3722	13147
	H-M	9.3%	4.7%	8142	6548
‘Unevenly loaded’	M-L	17.0%	21.8%	14886	30212
	H-H	17.3%	4.0%	15135	5548
‘Evenly loaded’	M-M	8.2%	0.8%	7191	1141
	L-L	43.9%	59.2%	38434	82098
All		100%	100%	87510	138694

Figure 22 presents the same analysis split by time of day and shows that the 1000-1559 time period had the highest proportion of observations that were classified as either ‘very uneven’ (‘H-L’) or ‘uneven’ (‘H-M’ or ‘M-L’). This was in contrast to the TRB metric which suggested the 0000-0659 time period had the most unevenly loaded trains.



**Figure 22 – Partition by ‘Two busiest and two quietest coaches’ metric, by time (Sample 2)**

The same analysis was conducted for different routes and this showed that trains from a particular station typically had the highest proportion of observations that were classified as either ‘very uneven’ (‘H-L’) or ‘uneven’ (‘H-M’ or ‘M-L’). This was consistent with the initial analysis conducted in §5.1.1.2.

Table 8 presents the same analysis, filtered for ‘busy’ services, which were defined as being those services departing with load factor greater than 75%. Compared with the previous analysis in Table 7, this shows that busy services were more prone to be affected by uneven occupancy, in particular for long trains.

**Table 8 – Partition by ‘Two busiest and two quietest coaches’ metric for ‘busy’ services (Sample 2)**

Classification		Short trains	Long trains	Short trains	Long trains
‘Very unevenly loaded’	H-L	8.6%	34.3%	3600	9901
	H-M	19.5%	22.7%	8141	6548
‘Unevenly loaded’	M-L	15.3%	19.3%	6381	5557
	H-H	36.4%	19.2%	15135	5548
‘Evenly loaded’	M-M	16.9%	3.9%	7047	1140
	L-L	3.1%	0.5%	1308	154
All		100%	100%	41612	28848

This new metric was successful in classifying services by how evenly they were loaded and furthermore overcame the limitations of the existing metric that were discussed in the previous section.

***New metric 2: 'Rear-Middle-Front' inter-coach loading diversity classification***

A limitation of the proposed new metric above is that it does not contain information on where along the train the crowded coaches are. In order to address this limitation, a second new metric was also proposed.

The train was split into 'Rear', 'Middle' and 'Front' and then all the ordered permutations of occupancy were considered using the same thresholds as above: Low (L), 0% to 75%; Medium (M), 75% to 100%; High (H), Over 100%. This yielded 27 possible classes: 'L-L-L', 'L-L-M', 'L-L-H', 'L-M-L', 'L-M-M' and so on, as defined in Table 9.

This metric was applied to departures for long trains from one station. This subset was chosen for illustrative purposes, because it was one of the scenarios identified in the previous section as being more susceptible to uneven loading. Figure 23 partitions the observations for this subset into the eight (out of 27) most common classes (shown in the columns) for each departure throughout the day (shown in the rows).



**Table 9 – Proposed new metric of ‘Rear-Middle-Front’ inter-coach loading diversity classification: definition of 27 ordered classes**

Name	Rear	Middle	Front
L-L-L	0% to 75% (L)	0% to 75% (L)	0% to 75% (L)
L-L-M	0% to 75% (L)	0% to 75% (L)	75% to 100% (M)
L-L-H	0% to 75% (L)	0% to 75% (L)	>100% (H)
L-M-L	0% to 75% (L)	75% to 100% (M)	0% to 75% (L)
L-M-M	0% to 75% (L)	75% to 100% (M)	75% to 100% (M)
L-M-H	0% to 75% (L)	75% to 100% (M)	>100% (H)
L-H-L	0% to 75% (L)	>100% (H)	0% to 75% (L)
L-H-M	0% to 75% (L)	>100% (H)	75% to 100% (M)
L-H-H	0% to 75% (L)	>100% (H)	>100% (H)
M-L-L	75% to 100% (M)	0% to 75% (L)	0% to 75% (L)
M-L-M	75% to 100% (M)	0% to 75% (L)	75% to 100% (M)
M-L-H	75% to 100% (M)	0% to 75% (L)	>100% (H)
M-M-L	75% to 100% (M)	75% to 100% (M)	0% to 75% (L)
M-M-M	75% to 100% (M)	75% to 100% (M)	75% to 100% (M)
M-M-H	75% to 100% (M)	75% to 100% (M)	>100% (H)
M-H-L	75% to 100% (M)	>100% (H)	0% to 75% (L)
M-H-M	75% to 100% (M)	>100% (H)	75% to 100% (M)
M-H-H	75% to 100% (M)	>100% (H)	>100% (H)
H-L-L	>100% (H)	0% to 75% (L)	0% to 75% (L)
H-L-M	>100% (H)	0% to 75% (L)	75% to 100% (M)
H-L-H	>100% (H)	0% to 75% (L)	>100% (H)
H-M-L	>100% (H)	75% to 100% (M)	0% to 75% (L)
H-M-M	>100% (H)	75% to 100% (M)	75% to 100% (M)
H-M-H	>100% (H)	75% to 100% (M)	>100% (H)
H-H-L	>100% (H)	>100% (H)	0% to 75% (L)
H-H-M	>100% (H)	>100% (H)	75% to 100% (M)
H-H-H	>100% (H)	>100% (H)	>100% (H)

For this particular station in the morning the most common classification was ‘L-L-L’, i.e. all three parts of the train had average occupancy less than 75%. From around 1000 to midday the most common classification was ‘M-L-L’, suggesting that often the rear of the train was starting to approach capacity. In the afternoon and evening there were a large proportion of trains that departed classified as ‘H-H-L’, ‘H-M-L’ or ‘H-L-L’, i.e. with coaches over capacity towards the rear of the train, but with spare capacity towards the front of the train.

These trends were consistent with the analysis in §5.1.1.2. This metric was therefore successful in quantifying at what point along the train the crowding occurred, i.e. for this example the rear of the train often being over capacity while the middle and front had spare capacity.

Time of day	H_H_H	H_H_M	H_H_L	H_M_L	H_L_L	M_M_L	M_L_L	L_L_L	Other
0600-0959	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	99.2%	0.8%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	12.5%	87.5%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.6%	98.4%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	2.0%	3.0%	95.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	8.9%	87.3%	2.5%
	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	8.8%	90.0%	0.0%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9.5%	90.5%	0.0%
	6.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	93.3%	0.0%
	0.0%	0.0%	0.9%	0.0%	0.0%	0.9%	16.8%	81.3%	0.0%
	5.0%	0.0%	0.0%	0.0%	0.0%	5.0%	5.0%	80.0%	5.0%
	0.9%	0.0%	0.0%	0.9%	0.0%	3.4%	15.4%	79.5%	0.0%
	3.5%	1.2%	0.0%	0.0%	0.0%	1.2%	4.7%	89.5%	0.0%
	0.0%	0.0%	0.0%	5.3%	1.8%	1.8%	19.3%	70.2%	1.8%
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	95.0%	5.0%
	4.0%	4.8%	4.0%	16.9%	2.4%	32.3%	21.8%	7.3%	6.5%
	2.3%	2.3%	0.0%	4.7%	2.3%	4.7%	23.3%	60.5%	0.0%
1000-1559	0.0%	0.0%	1.4%	10.8%	4.1%	9.5%	59.5%	8.1%	6.8%
	4.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	96.0%	0.0%
	0.0%	0.9%	4.7%	3.7%	8.4%	7.5%	43.9%	28.0%	2.8%
	0.0%	0.0%	4.2%	3.2%	1.1%	4.2%	30.5%	56.8%	0.0%
	0.0%	0.0%	8.1%	17.6%	13.5%	6.8%	37.8%	16.2%	0.0%
	0.0%	1.4%	1.4%	4.3%	20.3%	0.0%	55.1%	15.9%	1.4%
	0.0%	0.0%	1.9%	3.8%	4.7%	5.7%	44.3%	38.7%	0.9%
	0.9%	0.0%	2.7%	6.3%	13.5%	4.5%	45.9%	25.2%	0.9%
	0.8%	0.8%	1.5%	5.4%	15.4%	5.4%	62.3%	7.7%	0.8%
	0.0%	0.0%	0.0%	2.7%	4.1%	0.0%	39.7%	53.4%	0.0%
	0.0%	0.0%	1.2%	9.8%	12.2%	2.4%	57.3%	17.1%	0.0%
	1.0%	1.0%	2.0%	13.7%	6.9%	6.9%	50.0%	14.7%	3.9%
	0.0%	0.0%	2.1%	5.2%	12.4%	6.2%	50.5%	20.6%	3.1%
	0.0%	0.0%	3.5%	7.0%	17.5%	5.3%	40.4%	26.3%	0.0%
	0.0%	2.6%	3.9%	11.8%	7.9%	5.3%	48.7%	18.4%	1.3%
	0.0%	0.0%	0.0%	6.9%	14.9%	1.0%	58.4%	18.8%	0.0%
	0.0%	0.0%	4.0%	16.0%	20.0%	2.0%	46.0%	10.0%	2.0%
	1.4%	1.4%	8.7%	15.9%	11.6%	7.2%	44.9%	8.7%	0.0%
	0.0%	0.0%	8.1%	22.2%	10.1%	7.1%	36.4%	12.1%	4.0%
1600-1859	0.0%	2.2%	22.5%	30.3%	11.2%	7.9%	20.2%	3.4%	2.2%
	4.2%	5.2%	30.2%	24.0%	7.3%	11.5%	12.5%	3.1%	2.1%
	5.1%	16.9%	24.3%	15.4%	8.1%	5.9%	12.5%	8.1%	3.7%
	4.2%	4.2%	39.6%	25.0%	12.5%	0.0%	12.5%	2.1%	0.0%
	23.6%	38.2%	11.4%	9.8%	4.9%	1.6%	7.3%	1.6%	1.6%
	2.7%	3.6%	50.0%	11.6%	4.5%	0.0%	13.4%	12.5%	1.8%
	2.6%	0.0%	7.9%	18.4%	28.9%	7.9%	21.1%	10.5%	2.6%
	0.0%	1.8%	34.5%	41.8%	5.5%	3.6%	7.3%	5.5%	0.0%
	0.0%	2.6%	6.4%	19.2%	21.8%	3.8%	32.1%	12.8%	1.3%
1900-2359	1.3%	1.3%	12.8%	21.8%	15.4%	14.1%	25.6%	7.7%	0.0%
	10.5%	0.0%	26.3%	15.8%	21.1%	5.3%	10.5%	10.5%	0.0%
	0.0%	0.0%	8.7%	21.7%	30.4%	17.4%	13.0%	8.7%	0.0%
	0.0%	0.0%	16.9%	20.8%	16.9%	10.4%	27.3%	7.8%	0.0%
	12.5%	12.5%	25.0%	12.5%	12.5%	25.0%	0.0%	0.0%	0.0%
	12.5%	0.0%	12.5%	12.5%	0.0%	0.0%	37.5%	25.0%	0.0%
	5.3%	0.0%	5.3%	10.5%	36.8%	5.3%	31.6%	5.3%	0.0%
	0.0%	0.0%	4.3%	10.9%	19.6%	4.3%	41.3%	19.6%	0.0%
All	1.7%	2.5%	7.0%	8.9%	7.3%	4.7%	25.2%	41.5%	1.3%

Figure 23 – ‘Rear-Middle-Front’ inter-coach loading diversity classification applied to departures from one station (Sample 2)

## Summary

The air suspension dataset was automatically collected for the 11-month period, April 2013 to February 2014. This consisted of two readings of air pressure for each coach, taken just prior to the doors closing on departing each station. To generate the estimated passenger counts, the ratio was taken of the air pressure on departure to the minimum air pressure for each coach on each vehicle. A linear relationship between weight and pressure was assumed; i.e. the percentage increase in air pressure was multiplied by the tare coach weight to generate an estimated increase in weight as a result of the passengers on board. This was in turn divided by the average weight of a passenger with luggage to generate an estimated passenger count.

For illustration purposes, these four classes for coach-by-coach occupancy were each assigned a colour and a qualitative name: 0% to 75%, (green) 'Plenty of seats'; 75% to 100%, (amber) 'Nearing capacity'; 100% to 125%, (red) 'Some people standing'; Over 125%, (purple) 'Many people standing'.

The air suspension data was combined with timetable data to determine direction of travel and the dataset was cleansed, which included removing trains that departed in reverse formation. It was not possible to conduct manual calibration / validation passenger counts within the scope of this study. However, counts from ticket data were available for some services. This did not fully validate the passenger counts, although it did go some way to suggest that the estimates were reasonable. As discussed previously, the primary focus of this research was to investigate localised crowding, both inter-coach and inter-train. As such the relative levels of crowding were of greater interest than the absolute levels of crowding and so the validation against ticket counts gave sufficient confidence in the data quality for the purposes of the study.

Four different methods for viewing the counts were applied: coach-by-coach cumulative distributions; proportion of observations with coach-by-coach occupancy over 100% by time of day; proportion of observations with coach-by-coach occupancy over 100% for all times of day combined; and average coach-by-coach occupancy 'heat maps' across the network. This analysis suggested that there was uneven inter-train occupancy on many train services and that these trends were station-specific.

A cluster analysis was undertaken on the average coach-by-coach occupancy data to identify stations with similar patterns in a more robust manner. This was done for long trains, with six clusters in the southbound direction and five clusters in the northbound

direction giving sensible results. In the southbound direction three of the six clusters had some coaches with average occupancy over 70% with other coaches below 55%, while for northbound this was true for two of the five clusters. There were often similar stations in the respective southbound and northbound clusters, which suggested that for some but not all stations there were similar patterns in both directions.

Specifically, for southbound services: Cluster A1 had average occupancy around 75% towards the middle of the train and 50% towards the rear of the train; Cluster A2 had average occupancy around 80% towards the rear of the train and 50% towards the middle of the train; Cluster A3 had average occupancy around 75% towards the rear of the train and 30% towards the middle of the train, which was the greatest range (i.e. 'most uneven') of all the clusters, these eight stations were predominantly on the route identified by other metrics as being most susceptible to uneven loadings.

For northbound services: Cluster B1 had average occupancy around 70% towards the middle of the train and 50% on other parts of the train; Cluster B2 had average occupancy around 80% towards the front of the train and 50% towards the middle of the train.

It was found from the cluster analysis that patterns of inter-coach occupancy were often similar for neighbouring stations; this was also observed from the average occupancy heat maps. The finding that neighbouring stations had similar loading profiles was not a surprise, because in a situation where the net number of passengers alighting is relatively low, if a coach is busy on arrival at a station it will likely also be busy on departure. An important thing to note from this is that if a station has particularly uneven loadings, initiatives at that station in isolation may not be sufficient; i.e. if the problem is being caused by behaviours at a preceding station, it would be better to tackle the root cause of the problem 'upstream'. In other words, prioritising effort to provide solutions at those stations with uneven loadings and high boarding flows may also alleviate crowding at several subsequent stations 'downstream'.

Several metrics were used to identify locations where uneven occupancy was most prevalent. Overall on around a third of services there were at least some passengers standing who could have been sitting down. Overall for all stations and times of day, 2.0% of passengers on short trains and 1.8% of passengers on long trains were standing on departure, but there were sufficient seats elsewhere on the train for them to sit down. For one station on 'busy' trains, this metric was 9%.

The 'ratio of car occupancy to train average' was computed across all stations. A limitation of this existing metric is that it only explains the occupancy relative to other coaches and ignores the absolute occupancy; for example some early morning quiet services were classified as having very high loading diversity, even though there were no coaches at or near capacity. Because of this limitation, two new metrics to describe inter-coach loading diversity were proposed.

Six classes were proposed based on the two busiest and two quietest coaches; around 10% of long train departures were classified as 'very uneven' ('H-L'), i.e. observations where at least two coaches had occupancy less than 75% and at least two coaches had occupancy greater than 100%. This metric was applied individually to each station and direction to identify in a systematic way the stations where the inter-coach loading diversity was high and was successful in overcoming the limitations of the existing metric.

A limitation of the proposed new metric above is that it does not contain information on where along the train the crowded coaches are; in order to address this, a second new metric was also proposed. The train was split into 'Rear', 'Middle' and 'Front' and then all the ordered permutations of occupancy were considered. This metric was applied to one station that was identified as having particularly uneven loadings. In the afternoon and evening there were a large proportion of trains that departed classified as 'H-H-L', 'H-M-L' or 'H-L-L', i.e. with coaches over capacity towards the rear of the train, but with spare capacity towards the front of the train. This metric was successful in quantifying at what point along the train the crowding occurred.

## **5.1.2 RQ 2 – Link between inter-coach loading diversity and dwell times**

### **5.1.2.1 (2a) What was the distribution of dwell times and estimated boarding times?**

#### ***Data fusion with door locking data: 'Sample 3'***

As discussed in §5.1.1.1, the air suspension data was combined with timetable data to determine the direction of travel. This combined dataset was used for most of the analysis in the previous sections and has been referred to as 'Sample 2'. A third dataset was also available from the door systems, which consisted of one row for each time the doors unlocked and locked. This included the vehicle unit ID, but no position

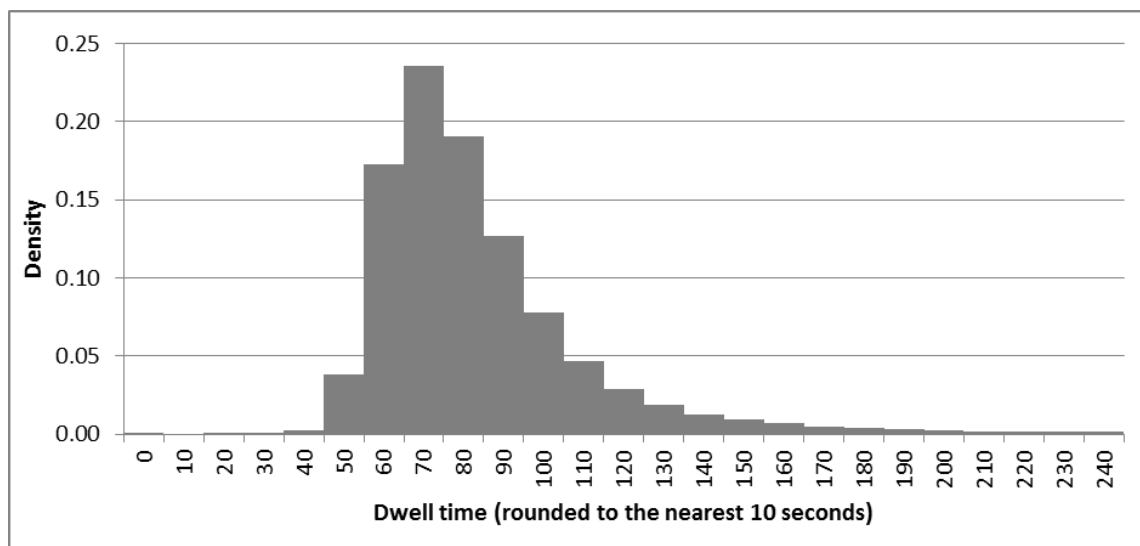
data or train service ID. Despite the lack of position data it was possible to combine the door locking data with Sample 2, using the timestamp from the air suspension dataset, which was usually between the doors unlocking and locking. This was a relatively complicated join and required careful sorting of the two datasets, rather than using a 'lookup'. Of 'Sample 2', it was possible to join 75.2% of the data and this is referred to as 'Sample 3'.

The time at which the doors locked was compared against the scheduled departure time from the timetable dataset as a measure of late-running.

### ***Generating estimated boarding times***

No data was directly available on boarding times, although the dwell times (defined for this study as the time between the doors unlocking and locking again at a station) were used as a proxy to estimate the boarding times for train services that arrived late. This was only done for trains that arrived late, because for trains that arrived early or on-time, it was likely that some of the dwell time would not include any boarding. In this study, 'trains that arrived late' was defined as instances in which a train arrived (i.e. the doors unlocked) at least one second after the scheduled departure time.

Figure 24 presents the distribution of the dwell times for trains that arrived late and shows that these 'estimated boarding times' were typically between 50 and 120 seconds. A small proportion of dwell times were over 180 seconds and these were excluded from the analysis in the remainder of this section, because they were considered likely to have been affected by reasons other than boarding.



**Figure 24 – Distribution of dwell time for services that arrived late (Sample 3)**

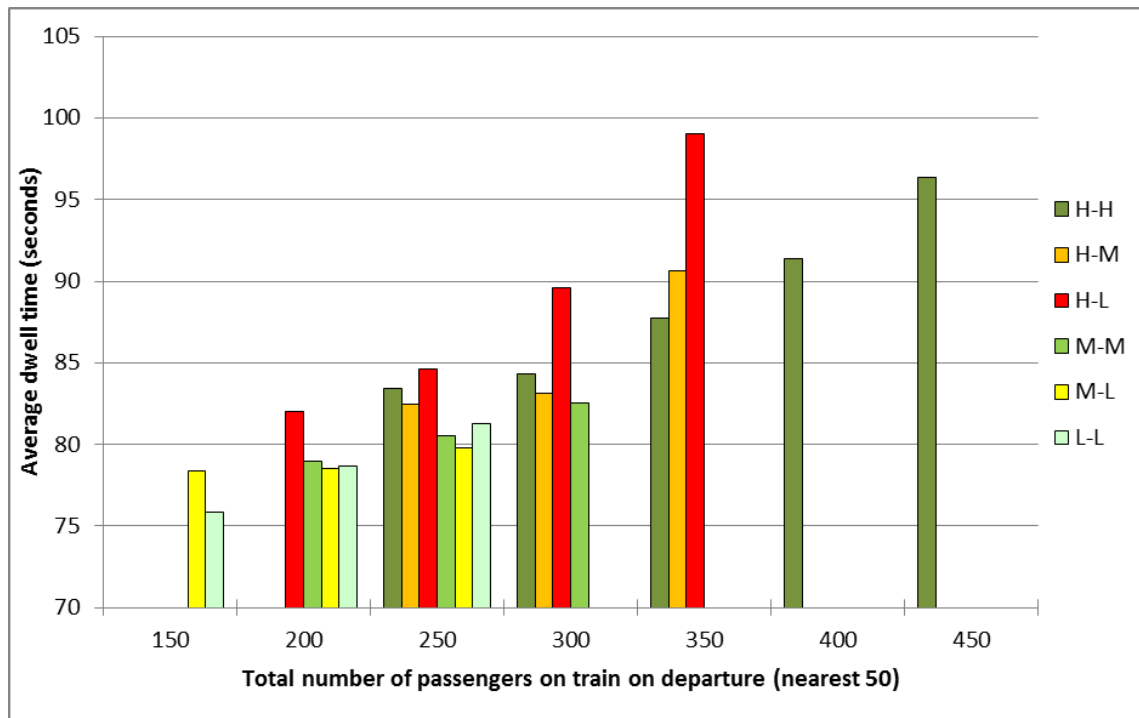


### 5.1.2.2 (2b) To what extent were longer dwell times related to uneven coach-by-coach occupancy?

#### *Average dwell times for late-running services, for different classes of inter-coach loading diversity*

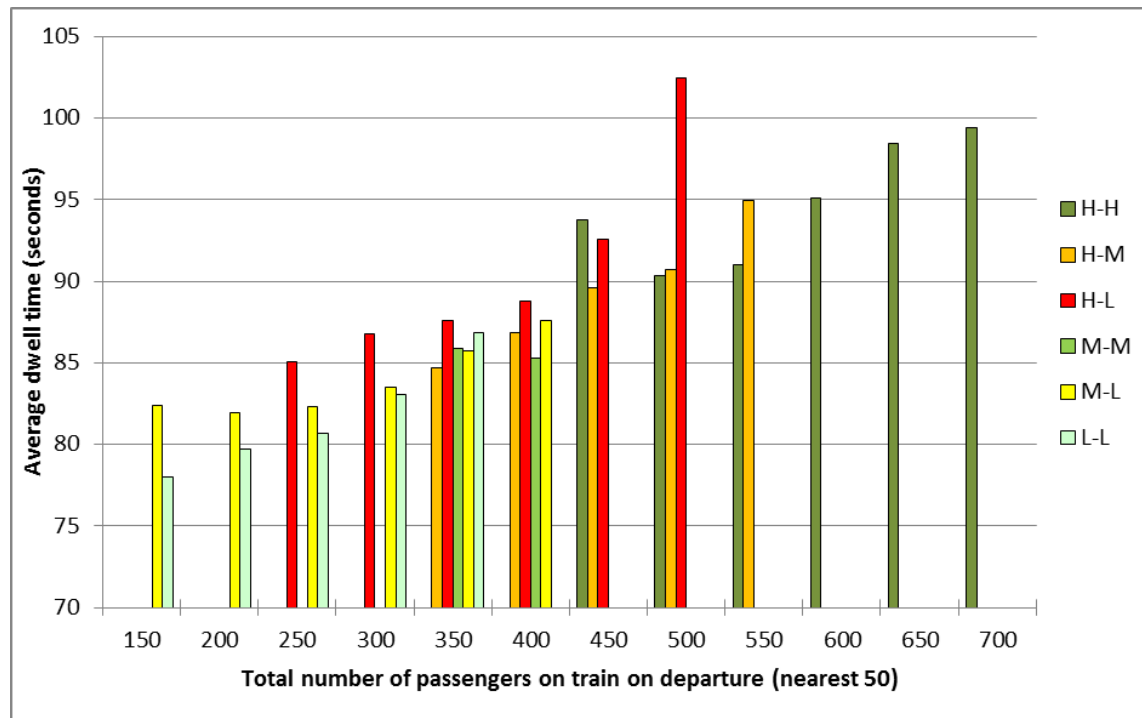
The average dwell times for train services that arrived late (from Figure 24) were correlated against the six classes of inter-coach loading diversity (from Figure 21). This is shown in Figure 25 and Figure 26 for short and long trains, respectively. Those coloured red and orange represent the observations classified as ‘very uneven’ (‘H-L’) and ‘uneven’ (‘H-M’ or ‘M-L’), whereas those coloured green represent observations classified as ‘even’ (‘H-H’, ‘M-M’ or ‘L-L’). The vertical axis shows the average dwell time and the horizontal axis shows the total number of passengers on-board, grouped to the nearest 50. As would be expected, the average dwell times were typically larger for the busier trains, ranging from around 80 seconds when there were 200 passengers on-board up to around 100 seconds when there were 700 passengers on-board.

The analysis suggested that there is a link between uneven loadings and dwell times. Trains classified as ‘very uneven’ (‘H-L’), i.e. with at least two coaches with occupancy less than 75% and at least two coaches with occupancy greater than 100%, typically had dwell times of approximately five to ten seconds greater than services that were classified as being ‘even’, with a similar total number of passengers on board.



**Figure 25 – Average dwell time for the six classes of inter-coach loading diversity, by total number of passengers on-board, short trains (Sample 3, trains that arrived late)**





**Figure 26 – Average dwell time for the six classes of inter-coach loading diversity, by total number of passengers on-board, long trains (Sample 3, trains that arrived late)**

## Summary

The occupancy data was combined with door locking data and timetable data. Dwell times were calculated as the time between the doors unlocking and locking again at a station. The time the doors unlocked was also compared with timetable data to identify train services that arrived late. No data was directly available on boarding times, although the dwell times were used as a proxy to estimate the boarding times for train services that arrived late.

These estimated boarding times were correlated against the new inter-coach loading diversity metric and the analysis suggested that there is a link between uneven loadings and dwell times. Trains classified as ‘very uneven’ (‘H-L’) on departure, i.e. with at least two coaches with occupancy less than 75% and at least two coaches with occupancy greater than 100%, typically had dwell times of approximately five to ten seconds greater than services that were classified as being ‘even’, with a similar total number of passengers on board.

## **5.2 Task 2 – Inter-train loading diversity: analysis of existing data**

Task 2 answered two groups of research questions; as discussed in §4.2, this was achieved through using the same dataset as in Task 1 ('Sample 2').

- RQ 3 – Quantification of inter-train loading diversity (§5.2.1)
- RQ 4 – Prediction of inter-train loading diversity (§5.2.2)

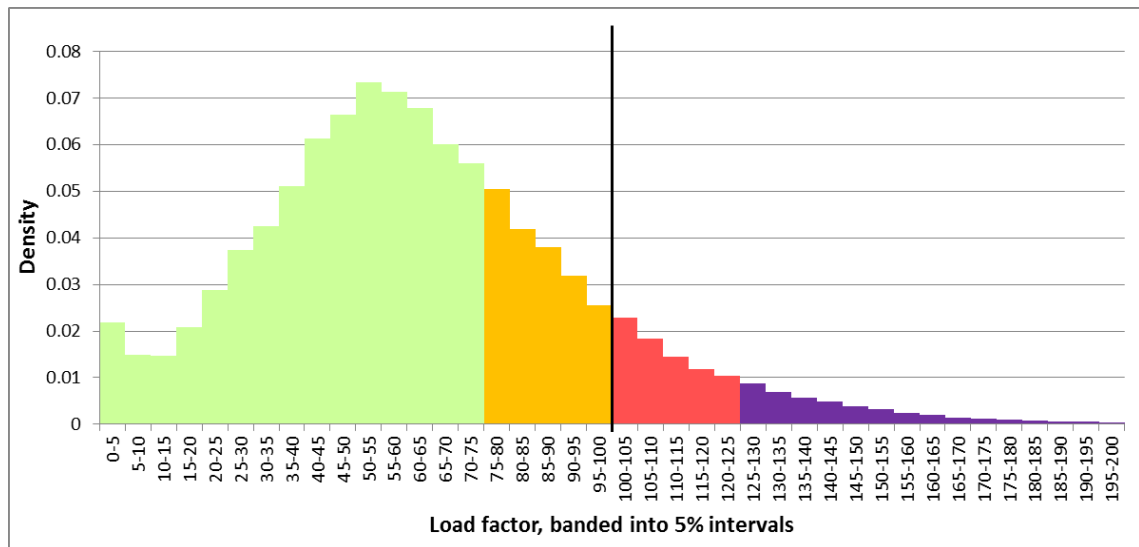
### **5.2.1 RQ 3 – Quantification of inter-train loading diversity**

#### **5.2.1.1 (3a) How did load factor vary by station, time of day and other attributes?**

##### ***Distribution***

The findings from the previous task demonstrated that it was possible to use air suspension data to detect coach-by-coach variations using the data cleansing approach described in §5.1.1.1. This same dataset was summed over all coaches to generate the estimated number of passengers on each departing train service. This estimate was divided by the capacity to generate the load factor. This was equivalent to the DfT definition that was introduced in §2.2.1 (DfT 2013b).

Figure 27 shows the distribution of the load factor, banded into 5% bins. The observations to the right of the black line represent instances where a particular train service had more passengers than seats were available, as such this corresponds with the official definition of 'over capacity', as described in §2.2.1. This was the case for 12.4% of observations, which as expected was lower in contrast to 15.1% of observations for individual coaches departing with occupancy over 100%, as shown previously in Figure 15.



**Figure 27 – Distribution of load factor (total passengers / capacity) upon departure (Sample 2)**

The proportion of the sample in each of the four classes was as follows:

- 0% to 75%: (green) – 68.9% of the sample
- 75% to 100%: (amber) – 18.8% of the sample
- 100% to 125%: (red) – 7.8% of the sample
- Over 125%: (purple) – 4.6% of the sample.

As was discussed previously the primary focus of this research was to investigate localised crowding, both inter-coach and inter-train. As such the relative levels of crowding were of greater interest than the absolute levels of crowding. Comparison against ticket data in §5.1.1.1 suggested sensible results sufficient for relative comparisons, although caution must be taken when making any conclusions on absolute levels of crowding, because no robust validation against manual counts has been undertaken.

In this section there follows a series of graphs showing the load factor, split by different variables. For each of these the top half presents the average load factor and the bottom half gives the partition of observations into the four classes listed above. The graphs give results separately for short trains (left) and long trains (right).

### ***Variability by station***

For short trains, the average load factor ranged from 94% at one station to 32% at another station. Of note, from preliminary analysis in §5.1.1.2, a particular station was identified as having some of the highest instances where there was occupancy greater than 100% in some coaches, whereas it is ranked 17<sup>th</sup> in terms of load factor compared to other stations. These results are consistent, because it was also identified as one of the stations most susceptible to uneven inter-coach loadings from the metrics defined in Task 1.

### ***Variability by scheduled departure time***

Figure 28 presents the load factor, split by time period. This shows that 1600-1859 was typically the busiest time period, with average load factor being 91% for short trains and 67% for long trains, likewise the proportion of observations classified as 'busy' (i.e. with load factor over 75% occupancy) being 65% for short trains and 34% for long trains. At the other end of the scale, 0000-0659 was the least busy time period with average occupancy being 34% and 27% and the proportion of services classified as 'busy' being 10% and 5%, for short and long trains respectively. Figure 29 suggests similar patterns, but presented for half-hour time periods, where for example '0730' represents all services scheduled to depart between 0730 and 0759<sup>7</sup>.

---

<sup>7</sup> There was a relatively low number of services scheduled to depart between 0100 and 0459 and these time periods were therefore excluded from the graph.

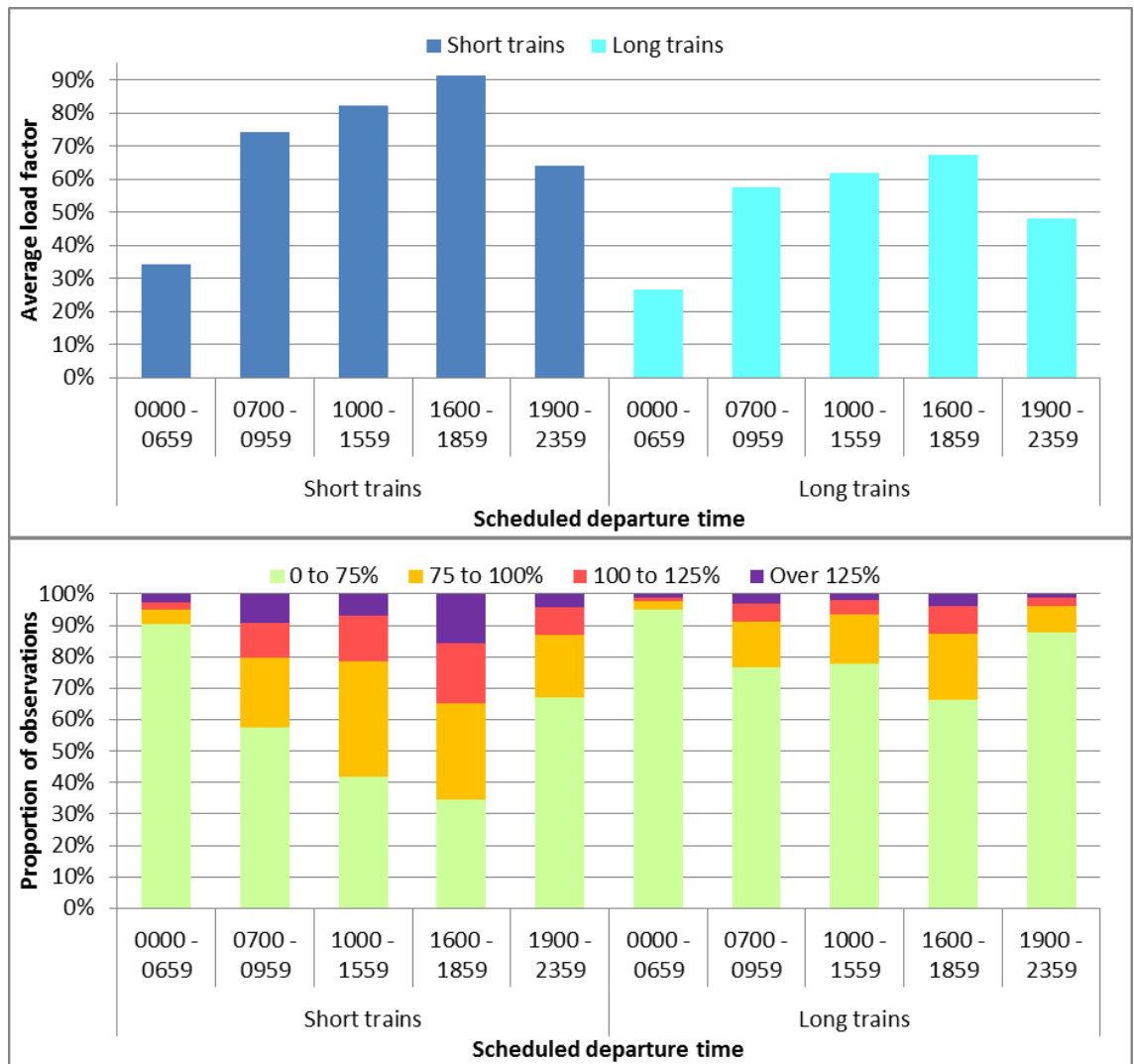


Figure 28 – Load factor average (top) and banded (bottom), by scheduled departure time period (Sample 2)

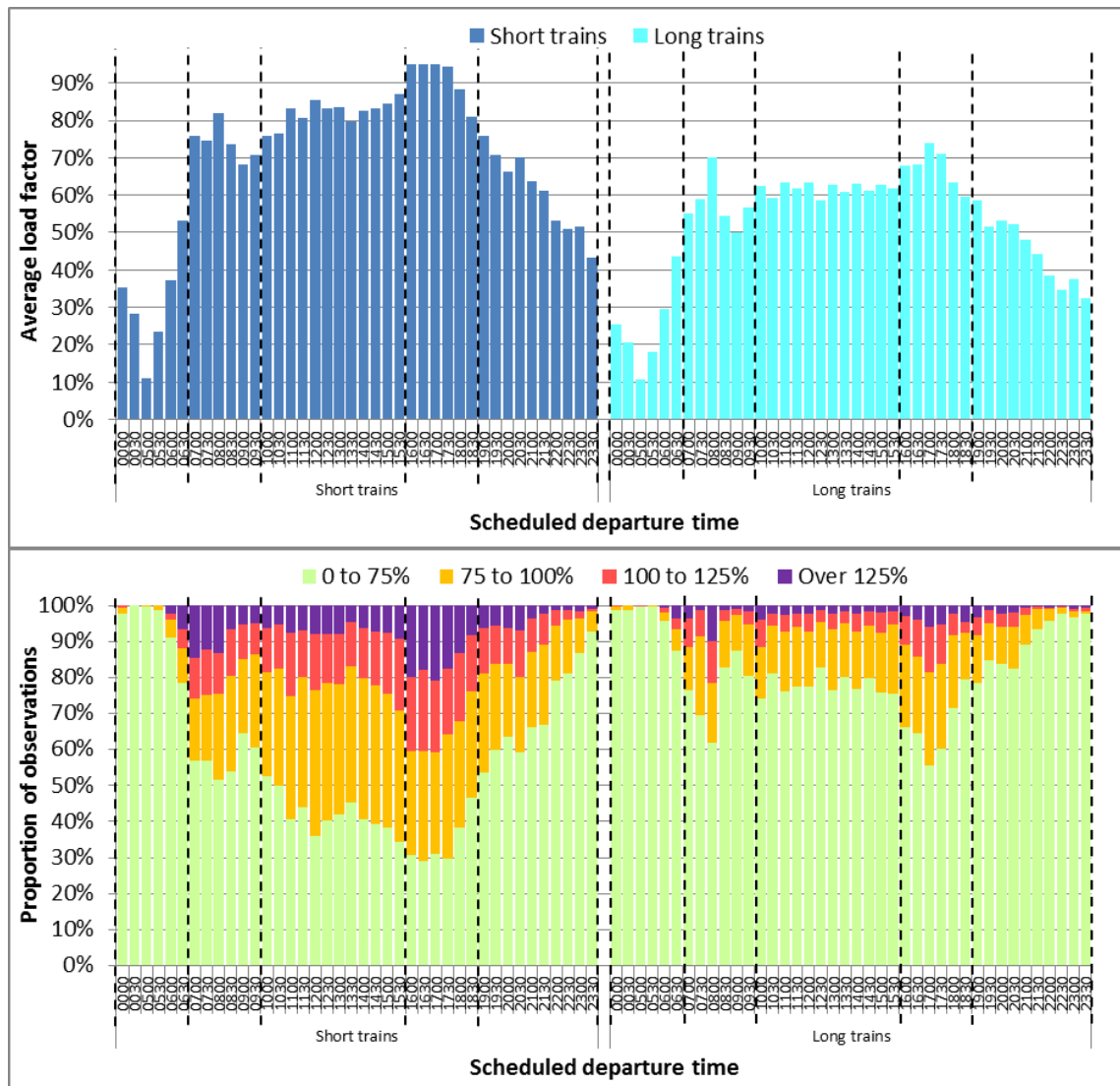


Figure 29 – Load factor average (top) and banded (bottom), by scheduled departure half-hour time period (Sample 2)

### Variability by day of week

Figure 30 presents the load factor, split by day of week. This shows that Fridays and Sundays were typically the busiest days, with average load factor being around 80% for short trains and 60% for long trains, likewise the proportion of observations classified as 'busy' being around 55% for short trains and 30% for long trains. Thursdays tended to be the next most busy day of the week, with Saturdays tending to be the least busy.

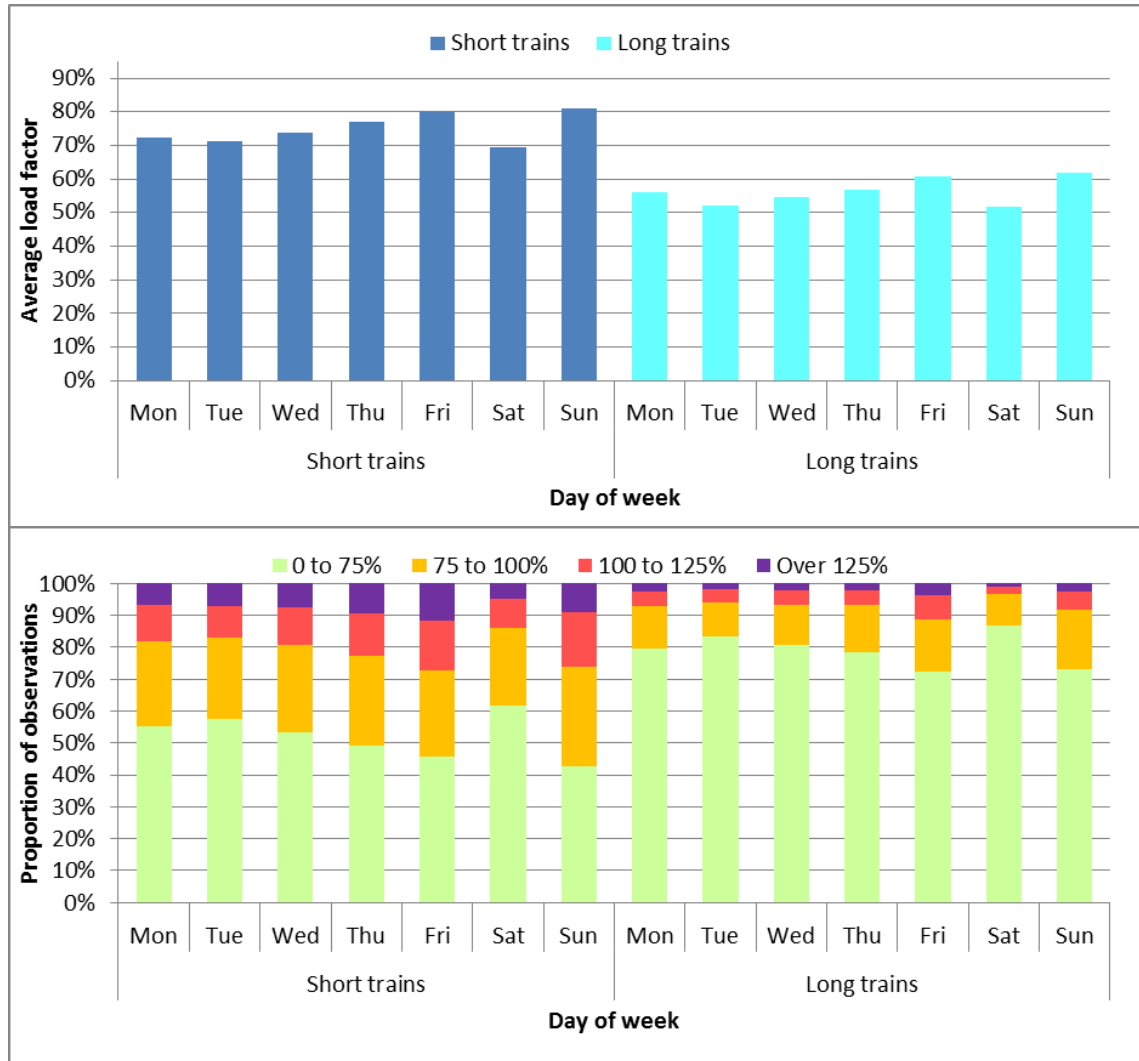


Figure 30 – Load factor average (top) and banded (bottom), by day of week (Sample 2)

### Variability by month

Figure 31 presents the load factor, split by month<sup>8</sup>. This suggests that November, December and February tended to be the busiest months, with average load factor being around 80% for short trains and 60% for long trains, likewise the proportion of observations classified as 'busy' being around 55% for short trains and 25% for long trains. All other months typically had a similar average load factor, around 70% for short trains and 55% for long trains, while the proportion of observations classified as 'busy' being around 45% for short trains and 20% for long trains. Each week across the time period was coded as being either term time or school holiday. As shown in Figure

<sup>8</sup> As discussed in §5.1.1.1, data was available for an 11-month period from the start of April 2013 to the end of February 2014, although timetable data was not available for October 2013



32, there were some differences between term time and school holiday, both in terms of average loading factor and the proportion of services classified as being busy.

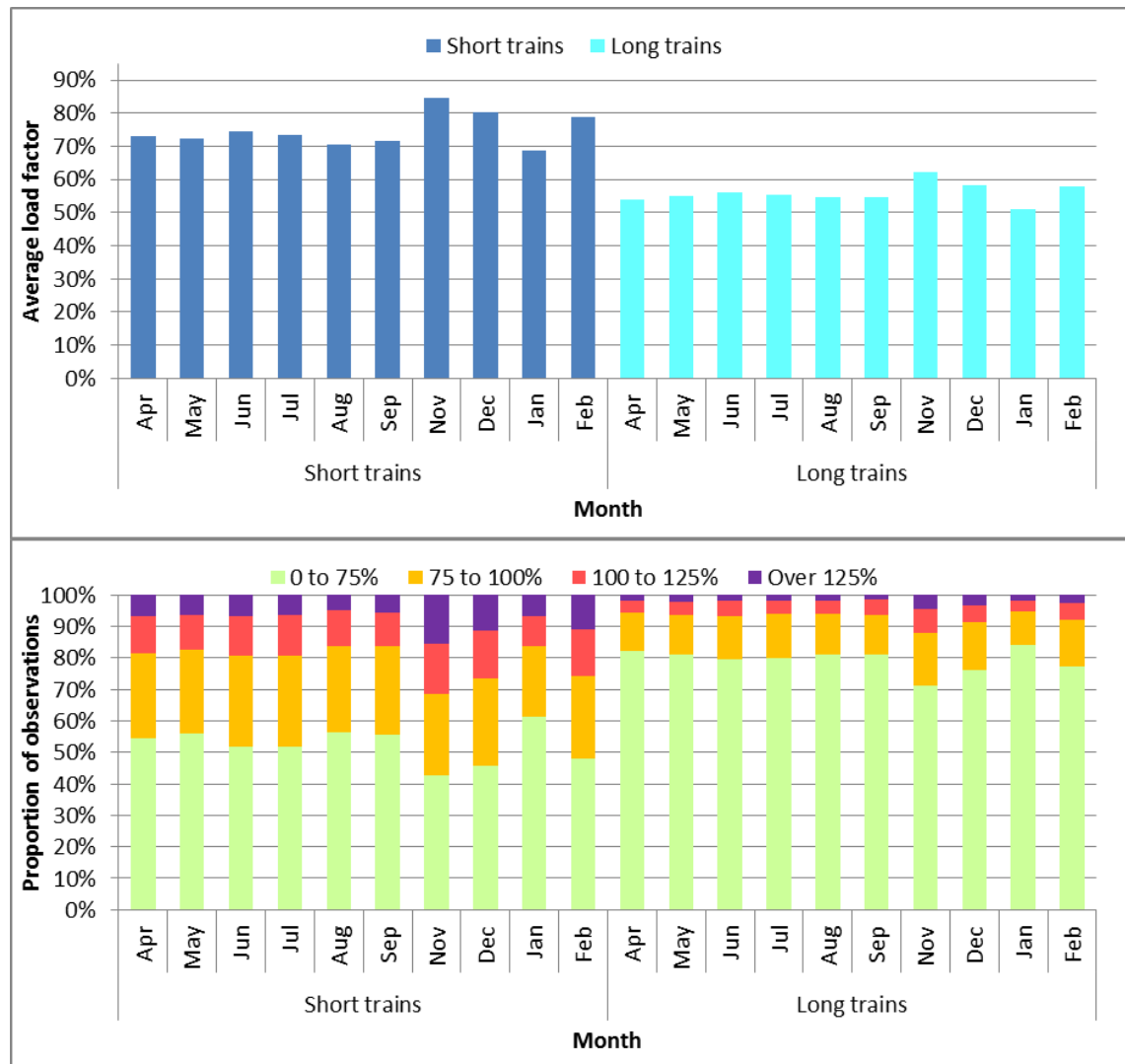


Figure 31 – Load factor average (top) and banded (bottom), by month (Sample 2)

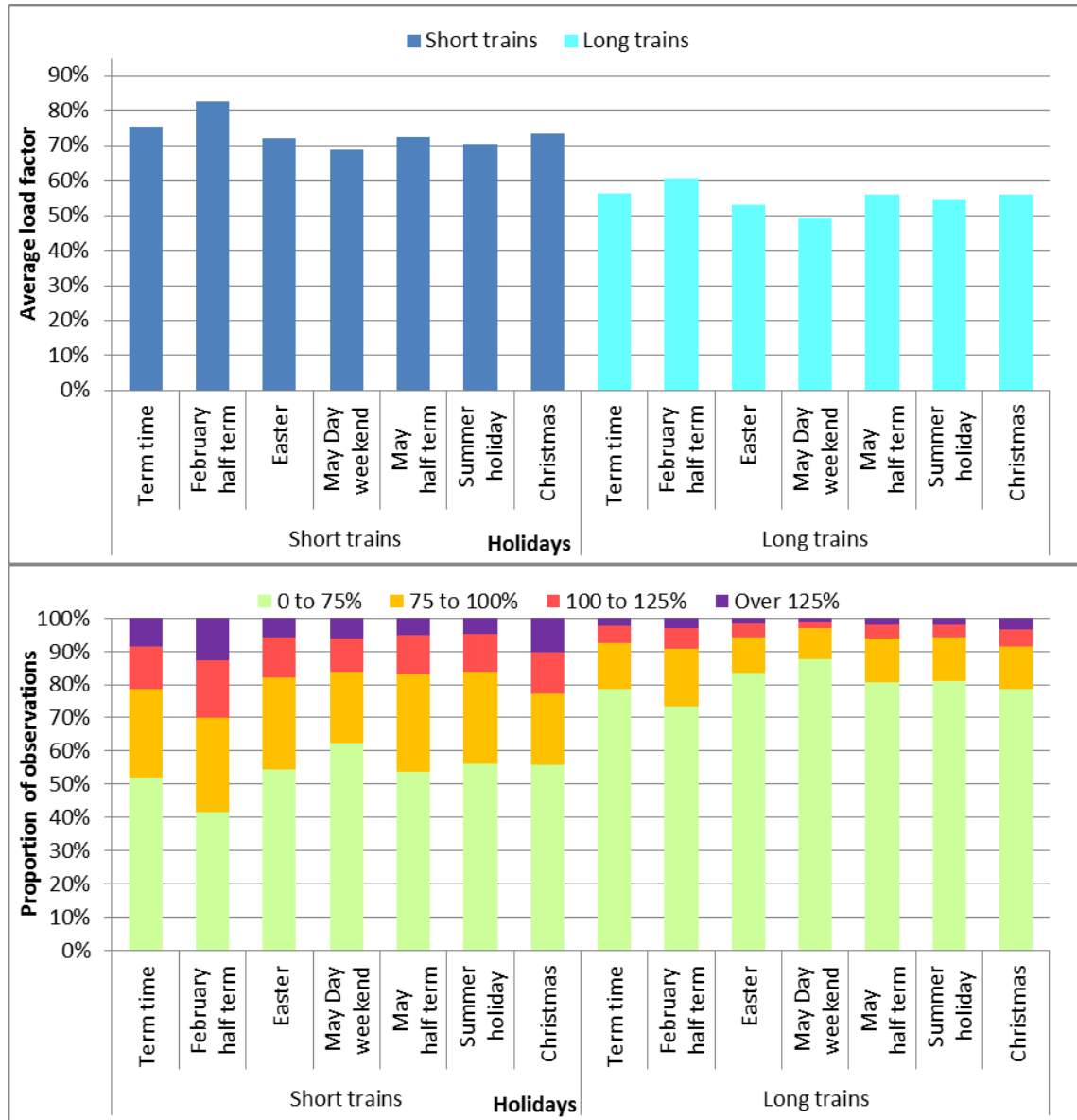


Figure 32 – Load factor average (top) and banded (bottom), by school holiday (Sample 2)

### Variability by late-running

Figure 33 presents the load factor, split by actual departure, banded into 'On time', '1 to 10 minutes late', '10 to 60 minutes late' and '60 to 120 minutes late'<sup>9</sup>. Here the timestamp for departure from the air suspension dataset was used rather than from the door locking dataset, i.e. 'Sample 2' rather than 'Sample 3'. This shows that late-running trains were typically busier than those running on time, with average load factor being around 80% for short trains and 60% for long trains, likewise the proportion of

<sup>9</sup> There were 347 observations with delay greater than 120 minutes, which have been filtered from the analysis.

observations classified as 'busy' being around 50% for short trains and 25% for long trains. One possible reason for this is that when trains run late, they start to carry passengers who were planning to catch the subsequent service. Another possible reason is that for trains that are busy anyway, boarding times may take longer and thus contribute to late-running; in this case there may be some inter-relationship with the other variables already discussed earlier in this section.

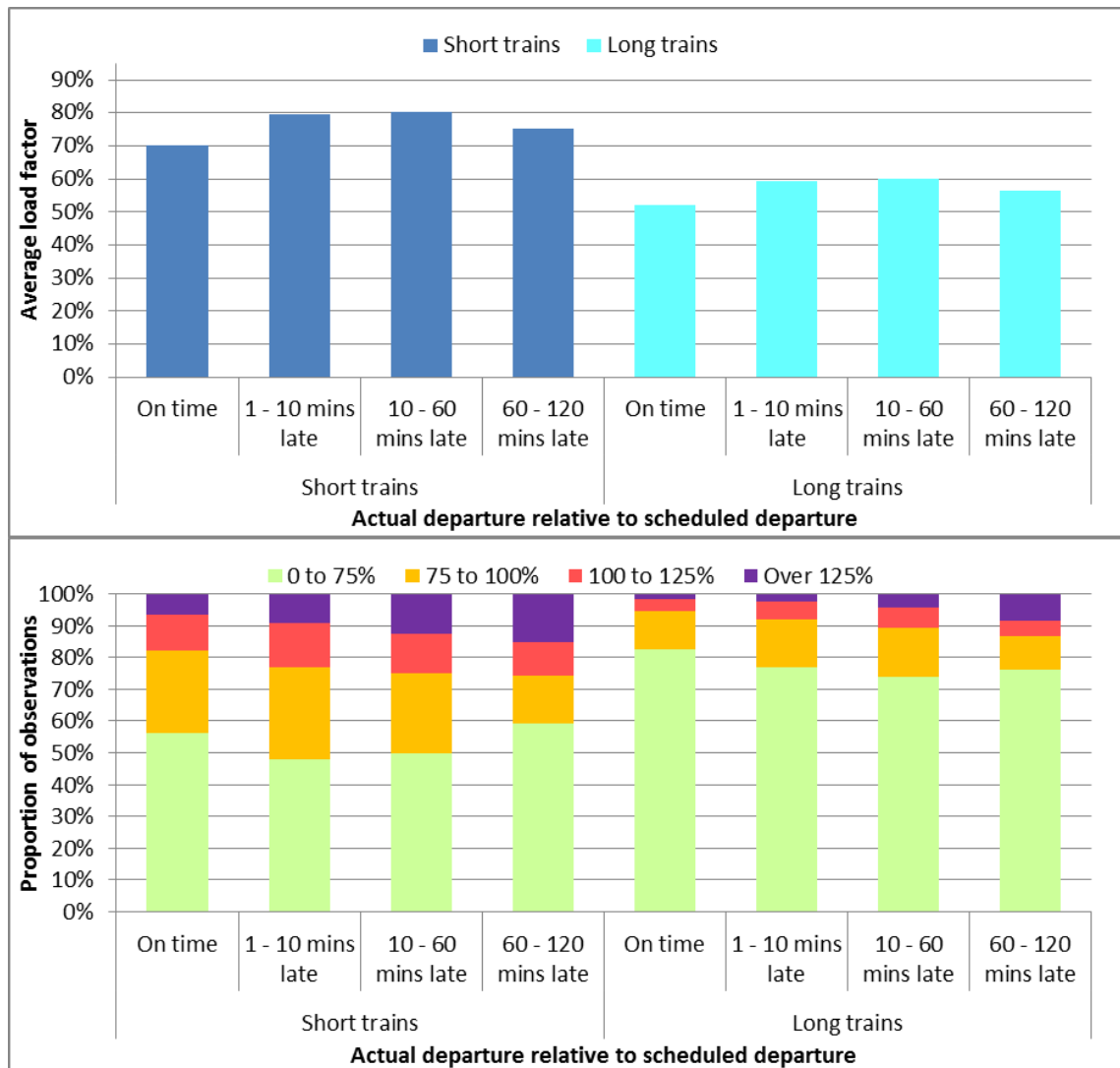


Figure 33 – Load factor average (top) and banded (bottom), by late-running (Sample 2)

### Standard deviation and coefficient of variation

The coefficient of variation is defined to be the average divided by the standard deviation and is used as a relative measure for variability. Most stations had coefficient of variation around 0.4, with similar values for both short trains and long trains, although a few stations had higher coefficient of variation, over 0.6.

The variability by time of day is explored below for two stations by plotting a time series of average load factor plus and minus one standard deviation for half-hour periods throughout the day; see Figure 34 and Figure 35. For these graphs, '0730' represents all scheduled departures from 0730 to 0759 and so on. These two stations had higher coefficient of variation than most other stations and their daily profile had particularly high peaks in the afternoon around 1600 to 1759, perhaps explaining the high coefficient of variation. The profile for one station oscillated from being relatively higher for the xx00-xx29 periods compared to the xx30-xx59 periods, which was explained by some trains terminating nearby while other trains continued through to a more distant destination.

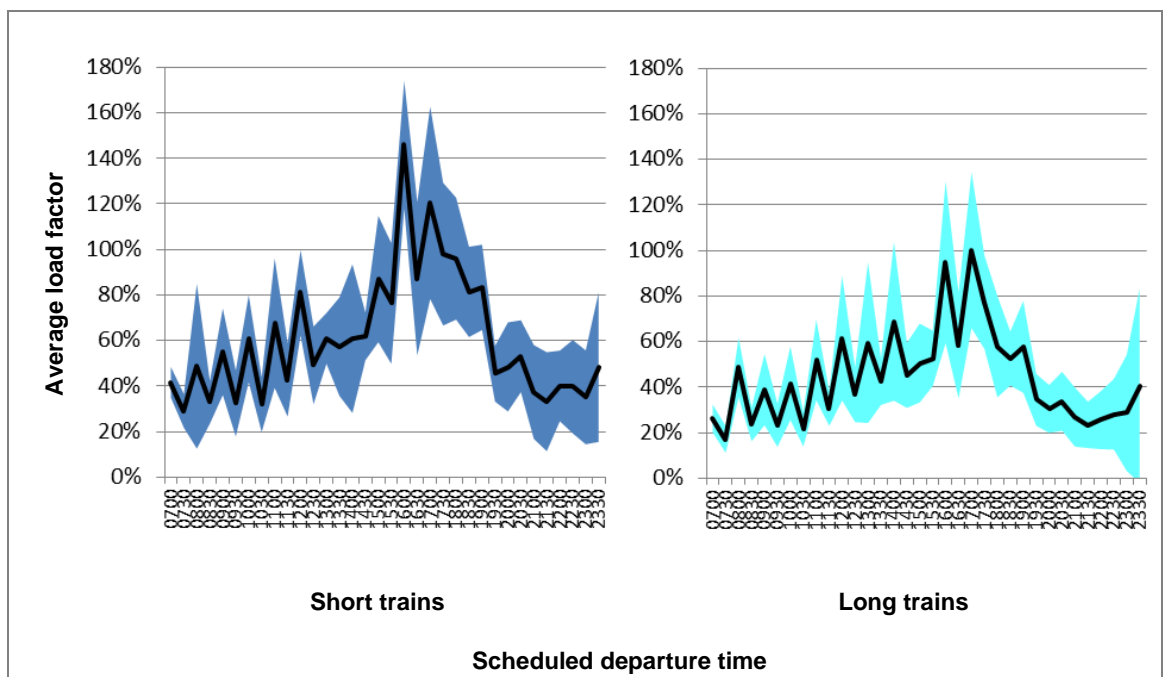


Figure 34 – Average load factor +/- standard deviation for one station (Sample 2)

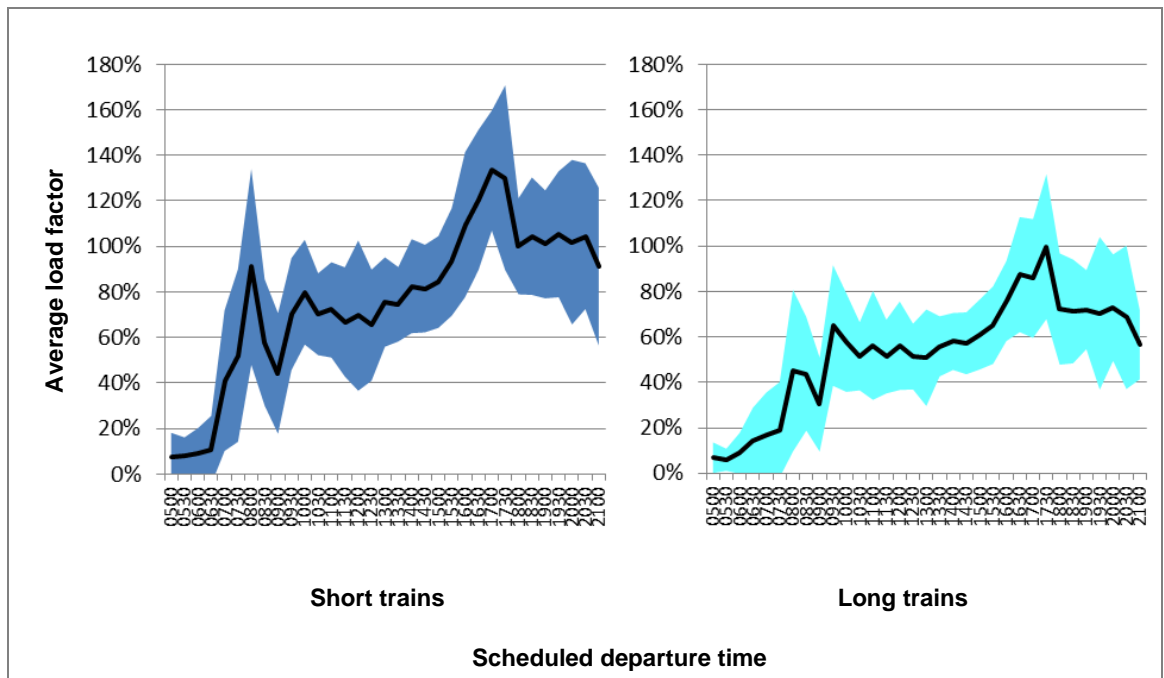


Figure 35 – Average load factor +/- standard deviation for another station (Sample 2)

#### 5.2.1.2 (3b) What metrics are appropriate to quantify the inter-train loading diversity?

##### *Application of existing metrics*

As identified in the literature review in §2.2.4.1, the TRB (2003) issued guidance that introduced the concept of 'loading diversity', of which they defined three types: "1. Loading diversity within a car"; "2. Loading diversity among cars of a train"; and "3. Unevenness of passenger demand during the peak hour". For the third type of loading diversity, where there are more passengers on some trains by time of day, they defined a metric called 'peak hour factor'.

*"Peak hour factor = (passenger volume in peak hour) / (4 \* passenger volume in peak 15 mins)"*

The peak hour factor gives a value between 0.25 and 1, where a higher value means that the passenger demand is more evenly spread over the peak hour. This measure was originally contrived for use with light rail services, which often have much higher frequency than heavy rail. As such this measure was adapted as follows:

*"Heavy rail peak hour factor = (passenger volume in two hour peak period) / (4 \* passenger volume in peak 30 mins)"*

Figure 29 showed that the 0700-0859 and 1600-1759 time periods were typically busy and thus the 'heavy rail peak hour factor' was proposed for these time periods. This identified the situations where a small shift earlier or later than the peak 30 minutes within peak two hour periods would result in a substantially quieter service. Of the stations for which this metric was calculated, the station with the profile shown in Figure 34 had the lowest value, likely explained by the oscillating pattern.

### Summary

The services with highest average load factor were identified across all stations, both by different routes and times of day. The variability of load factor according to different variables was investigated, with time of day yielding the greatest range, although train length, day of week, month and late-running also having an effect.

Time series graphs of average load factor with standard deviation were presented for selected stations, illustrating a range of profiles across the network. The TRB 'peak hour factor' metric was modified to identify the stations where a small shift of passengers taking earlier or later trains would yield the greatest smoothing of occupancy.

## 5.2.2 RQ 4 – Prediction of inter-train loading diversity

### 5.2.2.1 (4a) What techniques can be applied to predict load factor, based on historic data?

As discussed in §2.4, classification techniques are used for supervised learning problems where the dependent variable is categorical. Such techniques were applied in order to predict the (banded) load factor for a particular train in the future. As discussed in §4.2 the Naïve Bayes algorithm was implemented on a training dataset for different groups of predictor variables and the accuracy, precision and recall of this classifier was measured through applying the model to a test dataset.

#### *Thresholds: relationship between inter-coach and inter-train occupancy*

As discussed in §5.1.1.1, there are a variety of classifications in the literature for load factor of two, three or four classes, with different thresholds. In order to select a number of classes with appropriate thresholds for load factor, the relationship between inter-coach and inter-train occupancy was investigated.

Figure 36 shows the load factor banded in 5% intervals along the horizontal axis, with the lightest grey indicating observations where no coaches had occupancy over 100%, going through to the darkest grey where three or more coaches had occupancy over 100%. This showed that:

- For trains with load factor less than around 75%, typically most services had no coaches with occupancy over 100%;
- For trains with load factor between 75% and 100%, most had at least one coach with occupancy over 100%;
- And for trains with load factor over 100%, most had three or more coaches with occupancy over 100%.

This suggested that for the prediction task three classes would be sufficient, with thresholds being: '0 to 75%'; '75 to 100%'; and 'Over 100%' (indicated by the dotted line below).

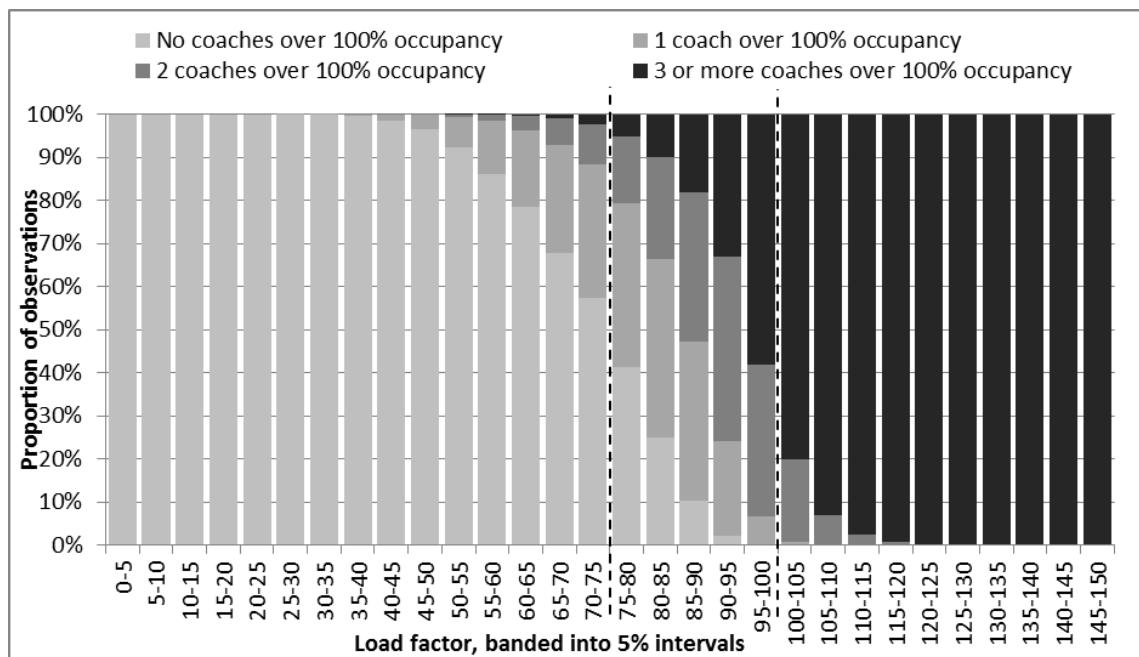


Figure 36 – Relationship between inter-coach and inter-train occupancy (Sample 2)

### Historic average: baseline model

A convention for supervised learning tasks is to split the dataset into a training dataset and a test dataset, with typical ratios being 2:1 or 1:1 (Bramer 2007). The 'Sample 2' dataset was split 50:50; weeks were numbered from 1 to 48, with 'odd' weeks forming the training dataset and 'even' weeks forming the test dataset.

As a baseline for the predictive model, the 'historic average' was computed for each station in each direction. The average load factor was computed for each half-hour period on the training dataset; this was essentially the same as Figure 29, but done for each station and each direction, although only on half of the dataset and for short and long trains combined. The average load factor was then used to classify each half-hour period for each station and direction as either '0 to 75%', '75 to 100%' or 'Over 100'.

The rationale for this choice of baseline model was that 'time of day' in half-hour periods showed the highest variability of all the predictor variables considered. The sample size for each station and each half-hour period was relatively high (typically ranging between 20 and 100), although if further variables (such as day of week) were added, in some instances this sample size would have become too small for robust results.

### ***Models tested***

The Naïve Bayes algorithm was implemented, as described in §2.4.2.1. The following categorical predictor variables were used:

- 1. Train length: Short; Long
- 2a. Time of day: 0000-0659; 0700-0959; 1000-1559; 1600-1859; 1900-2359
- 2b. Time of day (half-hour periods): 0000-0030; 0030-0059; ...; 2330-2359
- 3. Day of week: Mon; Tue; Wed; Thu; Fri; Sat; Sun
- 4a. Month (inc holiday weeks): Apr; May; Jun; Jul; Aug; Sep; Nov; Dec; Jan; Feb
- 4b. Month (exc holiday weeks): Apr; May; Jun; Jul; Sep; Nov; Dec; Jan; Feb
- 5. Delay: On time; 1-10 mins late; 10-60 mins late; 60-120 mins late

Six different Naïve Bayes models were implemented on the training dataset, with a mixture of predictor variables, as indicated in Table 10.



**Table 10 – Description of Naïve Bayes models**

Name	Technique	Description	Predictor variables included						
			1	2a	2b	3	4a	4b	5
Baseline	-	Historic average							
Model A	Naive Bayes	Four variables on train length, time of day, day of week and month							
Model B	Naive Bayes	A variation on Model A, using half-hour periods, rather than longer time periods							
Model C	Naive Bayes	A variation on Model B, excluding all weeks that were school holidays <sup>10</sup>							
Model D	Naive Bayes	A variation on Model B, excluding information on train length							
Model E	Naive Bayes	A variation on Model B, including information on train delays							
Model F	Naive Bayes	A variation on Model C, including information on train delays							

Model A used four predictor variables on train length, time of day (periods), day of week and month; from the univariate analysis, all of these predictor variables seemed to be relevant and so this was a good choice for a first model.

Model B was designed to be the same as Model A, with the exception of using half-hour periods rather than larger bins in order to understand the impact of this selection.

Model C was designed to be the same as Model B, but excluded all weeks that were school holidays, in order to understand the effect of excluding this part of the sample.

Model D was the same as Model B, although information on train length was excluded. The purpose of this model was that it is not always known far in advance whether a short or long train will be allocated to a particular service, so in practice it may not be possible to use the variable on length of train.

Model E was the same as Model B, but included information on train delays; similar to train length this variable would only be available in near-real time. Likewise, Model F was similar to Model C, having excluded all weeks that were school holidays, although

<sup>10</sup> This removed May Day weekend, May half term, Summer Holiday, Christmas/New Year and February half term. Easter was not included in the 11-month period and timetable data was not available for all of October. On filtering holiday weeks, all months had at least one 'odd' week and one 'even' week, with the exception of August.

it also included information on train delays to understand the effect of including this variable.

#### 5.2.2.2 (4b) What was the accuracy, precision and recall of these predictors?

The left half of Table 11 shows the actual data from the test dataset, whereas the right half presents the predictions from applying the models to the test dataset. The highlighted cells on the diagonals show instances where the models accurately predicted what actually happened, whereas the other two cells not highlighted in each row show instances where the model incorrectly predicted what happened.

For example in reading the top row, the Baseline Model accurately predicted for 57.14% of instances that they were classified as '0 to 75%', but for the other 10.78% of instances that actually were '0 to 75%', it gave incorrect predictions (10.23% being '75 to 100%' and 0.55% being 'Over 100%'). The final column is the model prediction recall (accuracy of the positive predictions) for that class, e.g. for the top row,  $57.14\%/67.9\%=84.13\%$ . The precision is also given for each model (fraction of the positives that the model identified).

The percentage in the bottom right corner is the sum of the diagonal, i.e. the proportion of instances that were accurately predicted for all classes combined. The sample size varied slightly for different models due to the different predictor variables; for example filtering school holidays ('4b') reduced the sample size compared to other model runs.

**Table 11 – Model predictions compared to actual observations (Sample 2, test dataset)**

Model	Actual class	Sample size	Prediction: 0 to 75%	Prediction: 75 to 100%	Prediction: Over 100%	Actual %	Recall
Baseline model {2b}	0 to 75%	73458	57.14%	10.23%	0.55%	67.9%	84.13%
	75 to 100%	20824	10.38%	8.03%	0.84%	19.3%	41.72%
	Over 100%	13866	4.26%	6.46%	2.10%	12.8%	16.38%
	All	108148	71.78%	24.72%	3.50%	100.0%	
	Precision		79.60%	32.50%	60.06%		67.28%
Model A [Naïve Bayes] {1,2a,3,4a}	0 to 75%	73531	64.68%	2.13%	1.13%	67.9%	95.20%
	75 to 100%	20831	13.94%	3.42%	1.88%	19.2%	17.77%
	Over 100%	13867	6.53%	2.42%	3.86%	12.8%	30.13%
	All	108229	85.15%	7.97%	6.87%	100.0%	
	Precision		75.95%	42.89%	56.17%		71.96%
Model B [Naïve Bayes] {1,2b,3,4a}	0 to 75%	73458	64.24%	2.73%	0.96%	67.9%	94.57%
	75 to 100%	20824	12.70%	4.61%	1.95%	19.3%	23.92%
	Over 100%	13866	5.01%	3.12%	4.69%	12.8%	36.60%
	All	108148	81.95%	10.45%	7.60%	100.0%	
	Precision		78.39%	44.08%	61.72%		73.54%
Model C [Naïve Bayes] {1,2b,3,4b}	0 to 75%	56142	63.82%	2.67%	1.02%	67.5%	94.53%
	75 to 100%	16088	12.49%	4.66%	2.19%	19.3%	24.10%
	Over 100%	10926	4.70%	3.13%	5.31%	13.1%	40.42%
	All	83156	81.02%	10.46%	8.52%	100.0%	
	Precision		78.77%	44.58%	62.32%		73.79%
Model D [Naïve Bayes] {2b,3,4a}	0 to 75%	73458	64.40%	2.23%	1.29%	67.9%	94.81%
	75 to 100%	20824	14.72%	2.54%	2.00%	19.3%	13.18%
	Over 100%	13866	6.75%	1.84%	4.22%	12.8%	32.94%
	All	108148	85.87%	6.62%	7.51%	100.0%	
	Precision		74.99%	38.35%	56.21%		71.16%
Model E [Naïve Bayes] {1,2b,3,4a,5}	0 to 75%	73328	64.10%	2.73%	1.08%	67.9%	94.39%
	75 to 100%	20804	12.51%	4.76%	1.99%	19.3%	24.72%
	Over 100%	13850	4.84%	3.18%	4.80%	12.8%	37.45%
	All	107982	81.45%	10.67%	7.88%	100.0%	
	Precision		78.69%	44.62%	60.98%		73.66%
Model F [Naïve Bayes] {1,2b,3,4b,5}	0 to 75%	56032	63.67%	2.68%	1.14%	67.5%	94.33%
	75 to 100%	16071	12.33%	4.80%	2.23%	19.4%	24.80%
	Over 100%	10910	4.51%	3.20%	5.43%	13.1%	41.32%
	All	83013	80.52%	10.69%	8.80%	100.0%	
	Precision		79.08%	44.93%	61.71%		73.90%

The baseline historic average model had overall accuracy of 67.3%, which although relatively high represented a poor performance in particular in predicting the 'Over 100%' class, correctly predicting 2.1% of instances as being in this class and incorrectly predicting 10.7% of instances as being in this class (i.e. a recall of 16.4%).

Model A gave improved accuracy over the Baseline Model, with overall accuracy 71.96% and better performance on the 'Over 100' class (recall of 30.1%), although poorer performance on the '75 to 100%' class (recall of 17.8%).

Model B (using half-hour periods rather than larger bins) gave a further improvement over Model A, with overall accuracy 73.54%. There were three variations on Model B:

- Model C (excluded all weeks that were school holidays) had improved accuracy compared to Model B from 73.54% to 73.79%.
- Model D (excluded information on train length) had reduced accuracy from 73.54% to 71.16%.
- Model E (included information on train delays) had marginally improved accuracy from 73.54% to 73.66%

Model F (included information on train delays) had marginally improved accuracy compared to Model C from 73.79% to 73.9%.

For the most part, the precision was higher than the recall, i.e. the models tended to predict crowded conditions only when there was relative confidence that this was the case.

The results for Model C were investigated for selected stations and it was found that the accuracy varied substantially across the network. Typically the model performed less well at the busier stations. However, for one station, although it achieved 95.69% accuracy overall, the model predicted all departures to be '0 to 75%' and so didn't detect any of the busy services from this station (recall of 0%).

### ***Preliminary investigation into decision trees***

A further model, Model G, used the same predictor variables as Model C, but using a decision tree (see §2.4.2.2) rather than the Naïve Bayes technique. Model C had the highest overall accuracy of the Models A-D and so the same variables were selected for Model G. (Model E and F had higher overall accuracy than Model C, but included the late-running of trains as a predictor variable, which in practice would not be available far in advance).

Model G was only built only for one station as an initial investigation into the decision tree approach. The order in which attributes were split was determined by maximising the information gain at each split, i.e. calculating the entropy for each option and choosing the feature that resulted in the largest reduction in entropy.

As shown in Table 12, the overall accuracy of the model was 82.57% (compared to 60.80% for Model C for this station). Also of note, Model G had particularly high precision for this station, making incorrect predictions in 0.5% of instances. Model G also offered substantially higher recall for this station, although offered no prediction in 16.9% of cases.

**Table 12 – Decision tree predictions compared to actual observations (Sample 2, test dataset)**

Model	Actual class	Sample size	Prediction: 0 to 75%	Prediction: 75 to 100%	Prediction: Over 100%	No prediction	Actual %	Recall
Model G [Decision trees] {1,2b,3,4b}	0 to 75%	717	45.79%	0.22%	0.00%	6.94%	52.95%	86.48%
	75 to 100%	368	0.22%	21.20%	0.00%	5.76%	27.18%	78.00%
	Over 100%	269	0.07%	0.00%	15.58%	4.21%	19.86%	78.45%
	All	1354	46.09%	21.42%	15.58%	16.91%	100.00%	
	Precision		99.37%	98.97%	100.00%			82.57%

The preliminary results of the decision tree model were promising, suggesting that this may warrant further investigation in a future study. The technique used was one of the simplest types of decision tree; there are many other variations as listed in Table 2, along with a range of other techniques that may also be applicable.

## Summary

Classification techniques are used for supervised learning problems where the dependent variable is categorical. Such techniques were applied in order to predict the (banded) load factor for a particular train in the future. The Naïve Bayes algorithm was implemented on a training dataset for different groups of predictor variables and the accuracy, precision and recall of this classifier was measured through applying the model to a test dataset.

From looking at thresholds used elsewhere, as well as investigating the relationship between inter-coach and inter-train occupancy, three classes of load factor were justified: '0 to 75%'; '75 to 100%'; and 'Over 100%'. Six variations of the Naïve Bayes model were tested with different selections of predictor variables. The model using train length, departure time (half-hour periods), day of week and month (excluding school holiday weeks) yielded overall accuracy of 73.8%, which was an improvement over the 'historic average' (half-hour periods) baseline model, with overall accuracy of 67.3%.

Adding information on delays only had a marginal improvement; however, this would not be known in advance, suggesting that models without the delay variable may be more appropriate. Ideally information on the train length would be included in the model

if this was available, because this had a relatively large impact on the accuracy of the results. The predictions would likely be improved with additional information not currently included in the models; e.g. including information on scheduled events that are expected to have an impact on occupancy.

A decision tree was also implemented for departures from one station and had a substantial improvement over the Naïve Bayes models for this station. This preliminary investigation suggested that decision trees and possibly other techniques may warrant further investigation in a future study.

## 6 Findings: staff workshop

This section presents the findings for the third task outlined in the methodology section:

- Task 3 – Perceived causes and effects of loading diversity, both inter-coach and inter-train (§6.1).

### 6.1 Task 3 – Perceived causes and effects of loading diversity, both inter-coach and inter-train

Task 3 answered one group of research questions: RQ 5 – Perceived causes and effects of loading diversity. As discussed in §4.3, this was achieved through a focus group with nine of the train operator's frontline staff. This was a relatively small task that provided supporting qualitative information to the primary data analysis tasks in §5, both in corroborating the findings relating to which trains were typically busy, but also in offering insights into the perceived causes and effects.

#### 6.1.1 RQ 5 – Perceived causes and effects of loading diversity

##### 6.1.1.1 (5a) From the perspective of frontline staff, which services had high levels of inter-coach local crowding?

Focus group participants were asked in turn to describe experiences of where some parts of trains are crowded, while there are seats available elsewhere.

It was noted that at two particular stations the middle of the train is often crowded, while coaches at the end of the train have spare capacity. A possible explanation was that a large proportion of passengers typically board near the stairs at these stations, which is by the middle of the train.

*“The stairs are right by the middle of the train, so people are going to get on those first doors... people do just hang around at the bottom of the stairs.”*

At a different terminus station, the coaches at the end of the train nearest the ticket gates were said to be typically the most crowded, with spare capacity further down the train. The reason for this was considered to be because passengers often take the first seat that they see.

*“For trains that aren't really busy, passengers approach the train and see that the first coach is busy, but then might see one spare seat in the second coach and will take that and so on. So you find that the one end is very busy, but the coaches further down the platform are completely empty.”*

The feedback from staff for these three stations was consistent with the patterns identified from the air suspension data in Task 1. This corroborates the findings and goes some way to suggest that primary data analysis task provided meaningful results.

#### **6.1.1.2 (5b) From the perspective of frontline staff, which services had high levels of inter-train local crowding?**

Participants were asked for their experiences on which were the most crowded services, by time of day, time of week and time of year. The consensus was that for ‘destination’ terminus stations, the last off-peak train before the evening peak and the first off-peak train after the evening peak are particularly busy, whereas for commuter stations, it is typically busy for boarding during the weekday morning services before 0900 and then most busy for people getting off the train in the evening.

Participants talked about large seasonal variations, with higher levels of crowding in the school holidays and in the week before Christmas. It was also noted that there are spikes in crowding due to local events near some stations, for example concerts or football matches.

Again, the feedback from staff on inter-train crowding was consistent with the patterns identified from the air suspension data in Task 2 and corroborated the findings on which trains every day were typically busy and also the variability due to events near some stations.

#### **6.1.1.3 (5c) What were the perceived causes of inter-coach local crowding?**

Participants were asked what they thought were the main factors that caused local crowding on trains. The consensus was that the cause of these situations was a mixture of several factors. For crowding on platforms, the following factors were discussed:

- Fixed:



- The location of the stairs and entrance/exit to the platform was noted as a key factor, with many passengers not walking down the platform when waiting for the train to arrive.
- Other platform features were also mentioned, such as the location of seating and the canopy (in wet weather).
- Station bottlenecks were considered to contribute to the problem.
- Variable:
  - Platform alterations can cause a 'last-minute panic', whereby passengers board at the first available door.
  - As discussed previously the time of day has a large influence, in particular with the busiest trains being the last off-peak train before evening peak and first off-peak train after evening peak.

For crowding on trains, the following factors were discussed:

- Fixed:
  - The design of the train was discussed as not being ideal, in particular there being a lack of luggage space, with some luggage space not obvious to passengers.
  - It was felt that the layout of the train makes it difficult to move from one coach to the next coach when there were some passengers standing.
- Variable:
  - Obstructions on the train make the situation worse by making it difficult to move along the train; this was perceived to be equally an issue in both the aisles and the vestibules: prams, sometimes quite large; luggage, in particular at certain times of year, such as summer and Christmas; and shopping. Bicycles were also mentioned, but these were considered to have a small impact. – *"The quantities of luggage and the prams stop people from getting through. You get one or two stop in an aisle and nobody else can get past them, so it is that obstruction factor."*

- On occasion the heating or ventilation is faulty in some coaches, which can lead to crowding in the coaches that are not faulty.
- The behaviour of different types of traveller also has a bearing: commuters often board the coach that they know will be closest to the exit on arrival and so be first to leave the station, in particular for commuting stations; leisure travellers / more infrequent travellers have different behaviour, often with more luggage and are less familiar with the train layout. – *“You have all your different types of passengers as well. For example, the commuters know exactly where to stand and they do every day of the week. Then there are also the ones who are just going out for the day.”*

#### **6.1.1.4 (5d) What were the perceived causes of inter-train local crowding?**

Participants were asked what they thought were the main factors that caused inter-train local crowding. It was felt to be obvious that the busiest times were those when most people wanted to travel, for example commuters wanting to travel in peak hours, spectators wanting to travel at the time of football matches and so on. One participant said that some passengers were unaware of ticketing restrictions; for example people with railcards tend to use off-peak trains, but often are not aware that they are entitled to use their discount on peak fare services. They suggested that crowding could be alleviated by encouraging these passengers to travel on less busy trains, in particular where there is a high student population.

#### **6.1.1.5 (5e) What were the perceived effects of crowding, both inter-coach and inter-train?**

Participants were asked for their opinions on the effects of crowding on trains. Various negative effects on passenger satisfaction, comfort and welfare were discussed. There were also several examples of where uneven boarding was exacerbating the problems of crowding and making it harder for the on-train staff to carry out their responsibilities and resulted in making their job more stressful. The effects of crowding on performance and revenue were also raised.

These impacts would be expected to lead to costs to train operators, for example by contributing to increased delay minutes and more indirect costs, such as reduced customer satisfaction and staff morale.

## Summary

A focus group was held with a mixture of train staff and station staff. Participants identified that there were some parts of the network that often had some coaches fully occupied, while other coaches had spare seats. For some stations it was the middle of the train that was often crowded and the end of the train had spare capacity, whereas for other stations the opposite was true. For terminus stations, the last off-peak trains before the evening peak and the first off peak-trains after the evening peak were identified as being particularly busy, whereas for commuter stations a tidal pattern was identified. Seasonal variations and crowding due to events were also discussed. Feedback from staff was consistent with the patterns identified from the air suspension data in Tasks 1 and 2.

The consensus was that the cause of these situations was a mixture of several factors, including the location of the entrance to the platform and the tendency for commuters to use the coach that was nearest to the exit at their destination station. Other factors were raised, such as luggage and buggies in the aisles and vestibules, which made the situation worse by making it difficult to move along the train.

Participants were asked for their opinions on the effects of crowding on trains. Their responses were grouped into three categories: impacts on passengers; impacts on staff; and impacts on revenue. These impacts would be expected to lead to costs to train operators, for example by contributing to increased delay minutes and more indirect costs, such as reduced customer satisfaction and staff morale.

## 7 Discussion

This section discusses the following:

- The implications of the key findings from §5 and §6 and how these relate to existing research (§7.1);
- Limitations of the datasets used (§7.2);
- Discussion on additional questions not covered in the study (§7.3).

### 7.1 Implications of findings

The implications of the findings for each of the research questions are discussed in turn.

#### 7.1.1 RQ 1 – Quantification of inter-coach loading diversity

The literature review identified the following gaps and opportunities:

- Gap: “The literature review did not find any academic research on the use of APC data in rail in the UK, which may be because the data is owned by the train operators and is commercially sensitive”.
- Gap: “Existing UK measures for rail crowding do not reflect coach-by-coach variations, nor do they reflect variations across the peak period; furthermore, American metrics for ‘loading diversity’ are not readily applicable to heavy rail, although they are a good starting point”.
- Opportunity: “APC data has the potential to be used to investigate a wide range of questions surrounding on-train crowding”.

Regarding the first and third bullet points, a large sample of APC data was obtained from a train operator, which enabled the possibility for novel research to be undertaken. The air suspension data was analysed to investigate patterns in coach-by-coach occupancy and to identify services where some coaches were over capacity while other coaches had spare capacity.

It was possible to use the air suspension data to generate estimated coach-by-coach occupancy on departure and this showed that there was uneven occupancy on many

train services and these trends were station-specific. For example, it was found that for trains departing some terminus stations there was clear evidence to suggest very uneven loading. There were very clear repeated trends, which were corroborated with qualitative evidence from a focus group with staff (§6.1.1.1). A particular station was identified where a real-time system may not be appropriate because of the repeated nature of the trends; better solutions at this station might be either static or variable message signs saying 'Please walk down platform'.

A range of several techniques were developed for visualising the data; this included the use of heat maps which were identified from the literature review in §2.2.3.3 (Itoh et al. n.d.) and were successfully adapted to investigate patterns in coach-by-coach occupancy. Such techniques would be applicable to similar datasets for other train services.

A cluster analysis was undertaken on the average coach-by-coach occupancy data for long trains to identify stations with similar patterns in a more robust manner. In the southbound direction three of the six clusters had some coaches with average occupancy over 70% with other coaches below 55%, while for northbound this was true for two of the five clusters; 15 stations were assigned to these clusters for southbound services and seven stations for northbound services, which suggested that uneven loadings may be a more common issue for southbound trains, compared to northbound trains. For southbound services, Cluster A1 comprised stations where on average the middle of the train was quite busy, while the rear of the train was quiet; the opposite was true for Clusters A2 and A3, with varying levels of average occupancy. Cluster A3 had the greatest range in average occupancy of all the clusters, i.e. it was the 'most uneven'. For northbound services, Cluster B1 comprised stations where on average the middle of the train was quite busy, while other parts of the train were quiet; the opposite was true for Cluster B2. There were often similar stations in the respective southbound and northbound clusters, which suggested that for some but not all stations there were similar patterns in both directions.

A key finding from the cluster analysis was that patterns of inter-coach occupancy were often similar for neighbouring stations; this was also observed from the average occupancy heat maps. The finding that neighbouring stations had similar loading profiles was not a surprise, because in a situation where the net number of passengers alighting is relatively low, if a coach is busy on arrival at a station it will likely also be busy on departure. An important thing to note from this is that if a station has particularly uneven loadings, initiatives at that station in isolation may not be sufficient;

i.e. if the problem is being caused by behaviours at a preceding station, it would be better to tackle the root cause of the problem 'upstream'. In other words, prioritising effort to provide solutions at those stations with uneven loadings and high boarding flows may also alleviate crowding at several subsequent stations 'downstream'.

Uneven inter-coach crowding was identified to be prevalent at some but not all stations. This suggests that efforts to address uneven occupancy should be targeted at particular stations based on insights from existing data, rather than blanket implementation at all stations.

In order to quantify the scale of the problem, it was estimated that across all departures around 2% of passengers were standing while there were sufficient seats for them to sit down. For busy services from one station, this measure was estimated to be 9%, which represents a substantial proportion of passengers at this station who were being unnecessarily inconvenienced.

Regarding the second bullet point on the lack of suitable existing metrics, various measures were calculated to distinguish between even and uneven services in a systematic way. There are existing metrics (§2.2.4.1), but these were not readily applicable to the problem at hand. Two new metrics to describe inter-coach loading diversity were proposed that, unlike existing metrics, contain information relative to the capacity.

Six classes were proposed based on the two busiest and two quietest coaches; around 10% of long train departures were classified as 'very uneven' ('H-L'), i.e. observations where at least two coaches had occupancy less than 75% and at least two coaches had occupancy greater than 100%. This metric was applied individually to each station and direction to identify in a systematic way the stations where the inter-coach loading diversity was high and was successful in overcoming the limitations of the existing metric.

A limitation of this proposed new metric is that it does not contain information on where along the train the crowded coaches are; in order to address this, a second new metric was also proposed. The train was split into 'Rear', 'Middle' and 'Front' and then all the ordered permutations of occupancy were considered. This metric was applied to one station that was identified as having particularly uneven loadings. In the afternoon and evening there were a large proportion of trains that departed classified as 'H-H-L', 'H-M-L' or 'H-L-L', i.e. with coaches over capacity towards the rear of the train, but with

spare capacity towards the front of the train. This metric was successful in quantifying at what point along the train the crowding occurred.

The new metrics are a contribution to knowledge that can be used by other researchers working in the field of loading diversity, as well as by train operators to better understand patterns of crowding across their network.

### **7.1.2 RQ 2 – Link between inter-coach loading diversity and dwell times**

Another gap identified by the literature review was:

- Gap: “There is evidence to suggest that optimising the distribution of passengers on the train may lead to dwell time savings, but there is only limited research on the extent of these savings”.

The occupancy data was combined with door locking data and timetable data. Dwell times were calculated as the time between the doors unlocking and locking again at a station. The time the doors unlocked was also compared with timetable data to identify train services that arrived late. No data was directly available on boarding times, although the dwell times were used as a proxy to estimate the boarding times for train services that arrived late.

These estimated boarding times were correlated against the new inter-coach loading diversity metric and the analysis suggested that there is a link between uneven loadings and dwell times. Trains classified as ‘very uneven’ (‘H-L’) on departure, i.e. with at least two coaches with occupancy less than 75% and at least two coaches with occupancy greater than 100%, typically had dwell times of approximately five to ten seconds greater than services that were classified as being ‘even’, with a similar total number of passengers on board.

These results suggest that if it were possible to optimise the distribution of passengers along the train for ‘uneven’ services, then typically up to ten seconds per stop could be saved for affected services. The size of this potential saving is not unsubstantial, in particular when such delays would be cumulative for each stop on a route. For example supposing a ten-second saving was made at each stop for a service with 20 station calls, this could equate to a time saving of over three minutes. This benefit would primarily be for late-running services in that services that are on-time cannot leave before their scheduled departure time, although it would have the benefit for services

that are running on time that they would be less likely to be delayed due to boarding delay.

As discussed in §2.2.5, the relationship between loading diversity and boarding times is likely two-fold. Firstly, the boarding time for a whole train is equal to the maximum boarding time of each individual door, thus if a disproportionate number of passengers board and alight from one particular door this will result in higher overall boarding time. Secondly, there is research from multiple studies to suggest that once the vestibule reaches a certain threshold of crowding, the flow rate at which passengers can board reduces. Thus it is intuitive that optimising boarding patterns along the platform will result in lower boarding times, although this relationship is likely non-linear if there is an interactive effect of crowding in vestibules.

### **7.1.3 RQ 3 – Quantification of inter-train loading diversity**

The occupancy data was used to investigate patterns of inter-train occupancy and the three gaps/opportunities listed above for RQ 1 were also relevant here for RQ 3.

The services with highest average load factor were identified across all stations, both by different routes and times of day. The variability of load factor according to different variables was investigated, with time of day yielding the greatest range, although train length, day of week, month and late-running also had an effect.

Time series graphs of average load factor with standard deviation were presented for selected stations, illustrating a range of profiles across the network. The TRB ‘peak hour factor’ metric was modified to identify the stations where a small shift of passengers taking earlier or later trains would yield the greatest smoothing of occupancy.

### **7.1.4 RQ 4 – Prediction of inter-train loading diversity**

An opportunity identified by the literature review was:

- Opportunity: “There are many different data mining techniques that may be readily applicable to help realise the potential offered by APC data”.



The Naïve Bayes algorithm was implemented on a training dataset for different groups of predictor variables and the accuracy of this classifier was measured through applying the model to a test dataset.

From looking at thresholds used elsewhere, as well as investigating the relationship between inter-coach and inter-train occupancy, three classes of load factor were justified: '0 to 75%'; '75 to 100%'; and 'Over 100%'. Six variations of the Naïve Bayes model were tested with different selections of predictor variables. The model using train length, departure time (half-hour periods), day of week and month (excluding school holiday weeks) yielded overall accuracy of 73.8%, which was an improvement over the 'historic average' (half-hour periods) baseline model, with overall accuracy of 67.3%.

Adding information on delays only had a marginal improvement; however, this would not be known in advance, suggesting that models without the delay variable may be more appropriate. Ideally information on the train length would be included in the model if this was available, because this had a relatively large impact on the accuracy of the results. The predictions would likely be improved with additional information not currently included in the models; e.g. including information on scheduled events that are expected to have an impact on occupancy.

A decision tree was also implemented for departures from one station and had a substantial improvement over the Naïve Bayes models for this station. This preliminary investigation suggested that decision trees and possibly other techniques may warrant further investigation in a future study.

The review of solutions identified that some train operators already use APC data to provide information on the level of crowding for a particular service through a variety of information channels (§2.3.6). In particular the Swiss rail operator provides information on the predicted level of crowding for a particular service on their ticketing website.

Although an improvement over the baseline model, accuracy of 74% from the Naïve Bayes model is perhaps not sufficiently accurate to be of value to rail users. However, preliminary results for an approach using decision trees were promising, which suggests that with further research such techniques may have potential to power predictive crowding information services, for example on ticketing websites.

### **7.1.5 RQ 5 – Perceived causes and effects of loading diversity**

A focus group was held with a mixture of train staff and station staff (§6.1.1), in which participants discussed particular stations and services that often had uneven loadings and high levels of crowding. These findings were similar to those observed in the occupancy data analysis task and thus went some way to validating these results. Feedback from staff provided many examples of where uneven boarding was exacerbating the problems of crowding and making it harder for them to carry out their work. The consensus from the focus group was that the cause of these situations was a mixture of several factors, including the location of the entrance to the platform and the tendency for commuters to use the coach that was nearest to the exit at their destination station

From the literature review, Kim et al. (2014) found that about three-quarters of subway users in South Korea reported choosing a specific coach intentionally and of these when asked to explain their motivation, 70% said “to minimise the walking distance to exit when they disembarked at a destination station” (§2.2.4.2). These results align partially with the focus group, although the train staff did not perceive this to be the main cause, with a variety of other factors also contributing to uneven loading.

It would be interesting to conduct a survey of passengers to get information on why they choose particular coaches and how this varies for different trip purposes and areas of the network. This would be a useful extension to the study, but has already been covered to a certain extent from existing studies, such as Kim et al. (2014), albeit primarily for commuters on a subway.

An opportunity identified by the literature review was:

- Opportunity: “Crowding on trains in the UK is an important problem from both a passenger and operator perspective. As such, investigating measures to ease crowding would likely yield benefits to both passengers and operators.”

Focus group participants were asked for their opinions on the effects of crowding on trains. Their responses were grouped into three categories: impacts on passengers; impacts on staff; and impacts on revenue. These impacts would be expected to lead to costs to train operators, for example by contributing to increased delay minutes and more indirect costs, such as reduced customer satisfaction and staff morale.

The qualitative evidence obtained on the wide range of impacts of crowding suggests that if solutions were implemented to address localised crowding, there may be benefits to both users and operators. More specifically, the potential benefits of such systems might include:

- Passenger comfort, satisfaction and welfare – Reduced crowding both on-train and on-platform would likely lead to better physical comfort if more passengers are seated rather than standing, reduced perceived value of time and perception of better value for money, all of which could lead to improved passenger satisfaction. Depending on the type of solutions, associated benefits could be improved information on the location of luggage and cycle storage spaces and improved access to on-train services such as the buffet car/trolley.
- Staff welfare – For train staff, reduced crowding on-train could lead to easier movement throughout the train and possibly in turn reduced staff stress. For train staff, better information and more even distribution of passengers along the platform may also lead to reduced workload and stress.
- Performance and revenue – If more even loadings resulted in reduced dwell times for late-running services, this would have a direct financial benefit to train operators. If there was reduced crowding on-train, staff would likely be able to perform their duties more effectively. If there was improved customer satisfaction this would likely lead to more passengers being encouraged to use rail.

It would be interesting to conduct a survey of passengers to investigate passenger perceptions and attitudes to crowding and how important it is to them. This would be a useful extension to the study, but has already been covered to a certain extent by existing studies, such as those summarised in §1.1.

## **7.2 Limitations**

Some limitations of the datasets used are discussed below.

### **7.2.1 Occupancy data**

As discussed in §4.1, the datasets obtained were quite comprehensive in their coverage of a large proportion of the train fleet over an 11-month period. This had

advantages over traditional survey approaches of giving a complete picture of trends rather than a sample for a few stations over a much shorter time period (such as those in §2.2.4.2). Another advantage of this approach was the negligible cost of data collection, in that it was automatically produced as a by-product ‘exhaust’ from the suspension system.

A limitation of the statistics produced was that they were sensitive to the assumption on average passenger weight, which was taken to be 80kg; if this assumption is too low then the proportion of standing passengers would be over-estimated and crowding would not be as high as suggested. Despite this limitation, the primary focus of this research was to investigate localised crowding, both inter-coach and inter-train; as such the relative levels of crowding were of greater interest than the absolute levels of crowding. Comparison against ticket counts suggested sensible results sufficient for relative comparisons, although caution must be taken when making any conclusions on absolute levels of crowding, because no robust validation against manual counts has been undertaken.

It would be useful to collect manual counts to calibrate the method for generating the air suspension passenger counts and to give greater confidence in the findings from the air suspension occupancy data. This could be achieved through surveyors collecting passenger counts while walking along the length of the train, or alternatively through manual cordon counts of the doors and vestibules using CCTV data. The calibration exercise could lead to refinements to the methodology for generating the occupancy counts.

A further limitation of the analysis was that no information was available on boarding and alighting flows. An advantage of door sensors, where available, is that assumptions are not required around the average weight of passengers; additionally, door sensors are able to provide bi-directional boarding and alighting flows, whereas weight sensors cannot. However, door sensors may be susceptible to ‘drift’ in the running total on-board and also to provide accurate passenger counts for each coach, additional door sensors are required between carriages. As discussed in §7.1.1, prioritising effort to provide solutions at those stations with uneven loadings and high boarding flows should be considered.

If information on boarding and alighting flows was available (e.g. for door sensor data, or possibly ticketing data), it may be possible to identify the stations that were the ‘root

cause', i.e. at which stations the effort should be targeted for solutions to smooth the crowding.

### 7.2.2 Dwell time data

The analysis used the times the doors locked and unlocked for late-running services as a proxy for overall boarding time. An extension to the work could be to also investigate the boarding times and passenger boarding and alighting flows at each door; this was not possible from the weight-based APC data, but would be possible if combined with door sensor data if available.

There are several existing research studies that have investigated the factors affecting boarding times through both real-world and laboratory experiments, such as Wiggens (2001), Lee et al. (2007), Fujiyama et al. (2008) (see §2.2.5). The mean alighting and boarding time per passenger has been found to be typically around one second, although this increases or decreases depending on various conditions, such as passenger mobility, platform design, vehicle design and crowding effects.

## 7.3 Effectiveness of solutions to influence loading diversity

Two further gaps/opportunities were identified from the literature review:

- Gap: "Although a wide range of solutions have been proposed that would seek to optimise loading diversity, only limited evaluation evidence is available on the impact of such systems".
- Opportunity: "The use of APC door sensors and weight sensors on trains is becoming increasingly widespread in the UK, with it being specified for installation on all new trains and already over 40% of existing trains equipped".

An additional group of questions was posed at the end of §3 on the effectiveness of solutions to influence loading diversity:

- *How can the solutions be categorised, both inter-coach and inter-train?*
- *How would passengers interact with the solutions, both inter-coach and inter-train?*
- *What would be the impacts of the solutions, both inter-coach and inter-train?*

- *What are the most promising solutions, both inter-coach and inter-train?*

It was unfeasible to conduct full-scale pilots within the scope of the MPhil; nevertheless, the findings from the literature review on solutions (§2.3) have been summarised and various aspects for future work in this area have been considered below.

### **7.3.1 (a) How can the solutions be categorised, both inter-coach and inter-train?**

There are a wide range of measures to increase rail capacity in general, such as longer trains with more coaches, creating more train services by running trains closer together, building new rail lines and so on. This study has not covered these measures, because the overall concept proposed was not to increase total capacity per se, but rather to help make the most of the existing available capacity, for example, so that no-one is standing when seats are available.

The literature review identified a number of potential measures that could be used to help to reduce uneven loadings, ranging from lower-cost measures that could be implemented in the shorter term to also higher-cost measures that would require longer term investment decisions. Passengers' behaviour can be influenced by simple direction signs, notices and announcements, as well as by more complex measures such as real-time information and changes to station layouts. Solutions have been categorised into eight categories of intervention, both by the approach used and their desired effect.

The approach has been split into three types:

- Active passenger management – solutions with a dynamic or real-time element;
- Passive behavioural modifiers – static solutions that seek to subconsciously modify user behaviour;
- Static crowding information – static information on historic levels of occupancy.

The desired effect, i.e. the type of loading diversity that the solutions are trying to address, has been split into four types:

- Within coach – spreading loadings out within a particular coach;

- Inter-coach – spreading loadings out between coaches on a particular train;
- Inter-train – spreading loadings between multiple trains;
- Inter-route – spreading loadings between multiple routes.

The eight categories are defined in Table 13, with links to the section of the literature review in which examples were discussed in greater detail. The number of examples in this ‘toolkit’ of solutions is expected to grow in time as such interventions are increasingly piloted and implemented.

**Table 13 – Proposed categorisation of ‘solutions toolkit’ to optimise loading diversity**

Approach	Desired effect	Solution category	Examples found
Active passenger management	Inter-coach	<b>1. Active passenger management: inter-coach real-time information</b> e.g. real-time platform displays or apps with information on where to stand on the platform in order to be more likely to get a seat	See §2.3.1
	Inter-coach	<b>2. Active passenger management: inter-coach near real-time seat reservations</b> Modifications to algorithms for seat reservations to assign certain origin-destination trips for advanced tickets to particular coaches	See §2.3.2
	Inter-train	<b>3. Active passenger management: inter-train real-time information</b> e.g. a personalised smartphone app with real-time information / predictions on the loading of subsequent trains with suggestions like, ‘You are more likely to get a seat if you take a later train’	See §2.3.5
	Inter-route	<b>4. Active passenger management: inter-route real-time information</b> e.g. real-time crowding information for alternative routes, possibly inter-modal as well	Not covered in thesis, but perhaps relevant for urban light rail
Passive behavioural modifiers	Inter-coach	<b>5. Passive behavioural modifiers: inter-coach platform design</b> e.g. modifications to platform designs to subconsciously modify behaviour, such as moving information screens or benches	See §2.3.4

	Within-coach	<b>6. Passive behavioural modifiers: within-coach train design</b> Increasing the capacity through design of train layout, such as fewer seats, removal of toilets, larger vestibules or wider doors	Not main focus of thesis, but partially covered in §2.2.5
Static crowding information	Inter-coach	<b>7. Static crowding information: inter-coach static information</b> Fixed signage, e.g. "Walk to end of platform ->", based on historic trends in the data and repeated patterns	See §2.3.3
	Inter-train	<b>8. Static crowding information: inter-train static information</b> Information on posters or ticketing websites on the typical patterns of loading	See §2.3.6



### 7.3.2 (b) How would passengers interact with the solutions, both inter-coach and inter-train?

#### Human factors

Human factors considerations are discussed in Table 14 for each category in the toolkit of solutions.

**Table 14 – Human factors considerations for each category in the toolkit**

Category	Human factors considerations
1. Active passenger management: inter-coach real-time information	<p>In order to be able to achieve the desired response real-time information must be made available and in a form that is identifiable, comprehensible and trustworthy:</p> <ul style="list-style-type: none"> <li>Information must be presented <b>clearly and unambiguously</b>. People standing at the platform must be able to see (or hear) the information, and be able to do this at useful locations. Ideally the information would be identifiable along the length of the platform, but as a minimum it should be available at the point where passengers first enter the platform so they can make an informed decision where to stand. Visually presented information must be clearly visible and (if relevant) readable. If information is presented audibly this must be clear and account for background noise.</li> <li>The information must be <b>simple enough to avoid potential confusion</b> and intuitive enough to cater for people of differing language, culture or experience of the system. Symbols and colour-coding would probably be most easily comprehended by the majority of people, although 'traffic-light' colour coding may be difficult for red-green colour-blind individuals.</li> <li>Real-time information must be perceived to be <b>accurate and up-to-date</b>. This means the information must accurately represent the actual situation on-board the train, and must be presented in enough time to allow people to act on the information. The information should also be able to cater for last-minute changes in the advice as if this is not conveyed it could be perceived as incorrect advice. If people grow to mistrust the system it will lose its effectiveness.</li> </ul>
2. Active passenger management: inter-coach near real-time seat reservations	There would be no human factors considerations, because no decision is required by the passenger.

Category	Human factors considerations
3 and 4. Active passenger management: inter-train or inter-route real-time information	The critical factor as to the success of this approach is likely to be in ensuring that the information is made available at a point in time where the user is reasonably able to alter their travel plans. For example, if on arrival at a station a traveller is informed that the next train is full and that the following train is less full, the individual's decision on whether to act on that information is likely to be influenced significantly by the arrival time of the following train. How long a traveller is prepared to wait is itself likely to be influenced by a host of other factors, such as: how long their overall journey is expected to take; the purpose of their travel (e.g. potential scheduling commitments); trust in the information provided (both as to the predicted waiting time and the certainty that they will get a seat); facilities at the station (leisure, food and drink, toilets, seating, waiting rooms); weather conditions (linked to waiting facilities); whether or not they are booked on a particular service. The effectiveness of any measures may therefore vary considerably between stations, between individuals, and at different times of the day/year. Measures seeking to encourage passengers to take alternative trains would therefore likely be more effective if the information can be provided in advance, perhaps when buying a ticket online or through the use of a smartphone app. However, an app could be expected to have limited market penetration and would be limited by how far in advance information is available to pass on.
5 and 6. Passive behavioural modifiers	Such measures have the advantage that information does not need to be identified and comprehended by individuals, instead relying on people to simply take the path of least resistance. Before employing such nudge tactics it is necessary to consider how different people may behave under different circumstances, and to be aware of the potential for unintended consequences of changing the flow of crowds.
7. Static crowding information: inter-coach	The purpose of static inter-coach crowding information is to persuade passengers to change their choice of coach. Providing information to passengers about such patterns of crowding may be less susceptible to issues of trust as with real-time information, as passengers may accept that the guidance will not be right all the time.
8. Static crowding information: inter-train	The purpose of static inter-train crowding information is to persuade passengers to change their time of travel. Posters may be a useful way of targeting regular commuters; however, providing information at the point of sale may allow more opportunity to influence the travel choices of individuals making one-off journeys.

### ***The optimisation challenge for inter-coach solutions***

Regarding all inter-coach solutions, it should be noted that there are essentially two optimisation problems that in some cases may not be complementary:

- Spreading occupancy evenly across the train (to reduce crowding and improve passenger comfort, satisfaction and welfare, as well as staff welfare)
- Spreading the number of boarders and alighters evenly across the platform (to reduce dwell times)

As an example, there is some evidence to suggest that some commuters tend to use coaches closest to their exit, meaning it is likely more people alight at these doors compared to other doors along the train. As such, if the primary goal is to optimise dwell times, it would likely be best to instruct boarding passengers to use other parts of the platform. However, if the primary goal is to optimise levels of crowding, it may be best to instruct the boarding passengers to use the doors where most people have just alighted, because there may now be more capacity on that coach.

Furthermore, there is a possibility that such interventions might make the situation worse; for example, the type of information (message channel, message location/size, message content etc) will have an effect on how many passengers comply with the instruction and there is the possibility that 'too many' will comply and more people than intended move to other parts of the platform. Therefore if piloting such real-time systems, it may be sensible to initially trial at a station with a relatively low number of passengers and different types of messages, starting with the least intrusive first before piloting at a busier station.

It seems that most real-time inter-coach solutions identified in §2.3.1 focus purely on how busy each coach is on the approaching train before anyone has alighted, i.e. there is no predictive element on the expected boarding and alighting flows.

### **7.3.3 (c) What would be the impacts of the solutions, both inter-coach and inter-train?**

While the review identified a number of approaches that could in principle be adopted, very little published evaluation evidence was found of the impact of such measures, namely: i) impact on how even occupancy is on the train; ii) impact on boarding times; iii) whether such interventions are safe.

As discussed above in §7.1.2, one aspect of this study was to identify the potential dwell time savings if such systems were able to deliver more even occupancy, although this has only partially addressed this gap. There remain unanswered questions on whether the solutions will work and in which situations they will work and in turn deliver

the potential benefits identified (see §7.1.5); this lack of evaluation evidence needs to be addressed before such systems are implemented.

A suggested approach would be to conduct pilots and impact evaluations of various solutions from the toolkit. This might involve pilot trials of interventions supported by 'Before' and 'After' monitoring to assess the impact of these measures in particular on occupancy, dwell times and user acceptance. Furthermore, after the evaluation has finished interventions would need to be supported by ongoing monitoring to understand how effects change over time. Such work should be undertaken in the context of the RSSB Platform Train Interface Strategy (RSSB 2015).

Considerations for future impact evaluations are discussed in Table 15 for each category in the toolkit of solutions.

**Table 15 – Considerations for pilots and impact evaluations for each category in the toolkit**

Category	Considerations for pilots and impact evaluation
1, 3 and 4. Active passenger managemen t: real-time information (both inter- coach or inter-train)	<p><b>Pilot trials of real-time passenger information on loadings, using existing air suspension data</b> – The review of existing solutions identified pilots of real-time passenger information systems of coach-by-coach occupancy of the incoming train, the most advanced of which was the Dutch pilot (§2.3.1.1). It may be possible to replicate a more basic version, but using existing systems. This would involve downloading air suspension data off the train in real-time, processing it and sending to the next station before the train arrives, then displaying a message on existing CIS. A pilot of such a system at a selected station would require effort at the system design stage and monitoring through video surveys to complement existing data sources. There would also be a human factors element, in particular on the most appropriate way to display the message.</p> <p><b>Pilot trials of real-time passenger information on loadings, using new door counters for passenger counts</b> – A more advanced version of a real-time system would be to use automatic door counters to generate the passenger counts. The main benefit of this approach would be that it would contain information on the boarding and alighting flows at each door, so could seek to optimise the passenger flows as well as the passenger occupancy. A potential pitfall with such systems is that they may have an adverse impact in that 'too many people comply with the announcements' and they could actually increase boarding times. Another key thing to be investigated is that optimising boarding times would not necessarily result in more even crowding. Such systems may therefore need to include predictions for the number of alighting passengers from each door as well as predictions on which doors the boarding passengers should use. This may be necessary to ensure no adverse impacts for dwell times or crowding.</p>

Category	Considerations for pilots and impact evaluation
2. Active passenger management: inter-coach near real-time seat reservations	<b>Modifying the seat reservation system to direct advanced bookings to the least crowded coaches</b> – For train services with seat reservations, it is possible to develop an algorithm to allocate a reserved seat to passengers in particular coaches based on their origin station. This might involve: analysing coach-by-coach seat reservation data; estimating the proportion of trips that the modified algorithm would affect through looking at ticket data; developing the algorithm; collecting data for a trial period to see the effect of the new algorithm on the reservations and also the overall occupancy and dwell times.
5 and 6. Passive behavioural modifiers	<b>Modifying platform layouts and train layouts to sub-consciously modify passenger behaviour</b> – Rather than providing information, measures could be implemented that seek to modify passenger behaviour through the use of ‘nudge’ tactics (§2.3.4). This might involve: conducting desktop analysis for selected stations of fixed features and train stopping positions; creating a shortlist of stations for site visits; collecting observations of passenger behaviour at the selected stations; creating a list of measures for particular stations that may be appropriate; assessing the impact on occupancy and dwell times after a trial period. In particular this could be aligned with any station redevelopments that this information could inform/influence.
7. Static crowding information: inter-coach	<b>Installation of fixed signs to direct passengers to the least crowded ends of the train on selected station platforms</b> – Some London Underground stations have signs pointing to less busy ends of the platform (§2.3.3); a similar implementation could be trialled on selected station platforms. This might involve: using the existing analysis along with ticket data to identify a shortlist of stations for the fixed signs; considering human factors options for installing the type of sign (location; size; message content); assessing the impact on occupancy and dwell times after a trial period.
8. Static crowding information: inter-train	<b>Providing passenger information on the ‘expected’ level of crowding on specific trains</b> – Rather than spreading passengers out to other coaches, this measure would seek to encourage passengers to take earlier or later trains. This could be achieved through information on a ticketing website (e.g. §2.3.6.1). Such information systems could be powered by the findings from Task 2 to indicate the most appropriate techniques. An alternative approach could be to use posters (§2.3.6.2), which may have higher exposure to passengers, although with a lower level of detail on particular services. A ticket website approach would have the benefit of being more likely to influence behaviour, because would be at the point of sale. The impacts of the intervention could be assessed through an assessment of changes in occupancy, supported by passenger surveys.

For such pilots, typically a three-stage evaluation would be appropriate:

- Stage 1 – One of the key insights from the findings in §5 was that the patterns of inter-coach and inter-train occupancy vary substantially in different circumstances. For any future pilots a proposed necessary first step therefore is to conduct, similar to as done here, a spatio-temporal analysis of uneven occupancy using data from automatic passenger counters across all stations and train services, applying the new metrics identified. This should be combined with passenger flow data if available to identify the 'root cause' stations. This review of existing data could be complemented with the collection of new data in a human factors study in order to take into account the trip purpose and type of passengers to understand the behavioural drivers. The outcomes of this first stage would be as follows:
  - A better understanding of inter-coach crowding, i.e. for which stations and services uneven loadings is most prevalent and the potential for improvement;
  - A better understanding of the causes and effects of inter-coach crowding;
  - A better understanding of inter-train occupancy, i.e. how occupancy varies by time of day and the potential for 'spreading the peak';
  - Identification of which stations would form suitable locations for pilots of solutions from the toolkit.
- Stage 2 – Based on the findings of Stage 1, appropriate solutions should be piloted at carefully selected stations and the effects of the interventions on passenger behaviour should be assessed. This should comprise a 'Before' and 'After' impact evaluation using quantitative analysis from video surveys, supported by qualitative analysis from passenger questionnaires and focus groups. Such studies could focus on the following measures:
  - Impact on passenger comfort, satisfaction and welfare, including the average coach-by-coach occupancy and other loading diversity metrics;
  - Impact on performance, including dwell times for late-running services;

- Impact on staff welfare, including ability to carry out duties effectively and stress.

The outcome of this stage would be to collect measurable evaluation evidence on the effectiveness of the solutions to inform whether implementation at other locations may be appropriate.

- Stage 3 – Following the findings from the pilots in Stage 2, it may be appropriate to implement solutions at other stations. This would need to be accompanied by ongoing monitoring to assess any changes in behaviour as passengers adapt over time.

#### **7.3.4 (d) What are the most appropriate solutions, both inter-coach and inter-train?**

Robust conclusions on which solutions are most appropriate cannot be made without evidence from future impact evaluations on their impact as discussed in §7.3.3. Nevertheless, it can be concluded that the selection of the most appropriate solution would likely depend on the nature of each station and other factors:

- Variability of occupancy patterns – As discussed in §7.1.1, static solutions may be most appropriate where the distribution of passengers is often a repeated trend; for example at a terminus where the end of the train closest to the station is often more heavily crowded than the other end of the train. However, for situations where the distribution of passengers is more variable, real-time measures may be more appropriate.
- Platform layouts – Observational studies or interviews on passenger behaviour on platforms could reveal the underlying causes for uneven loading patterns at particular stations. Where such causes are identified, solutions that seek to passively modify behaviour may be an appropriate first step before attempting more expensive 'Active passenger management' solutions.
- Availability of technology – The selection of intervention will likely depend on the availability of existing technology, namely the coverage of APC sensors on the train (either weight-based and/or door sensors), but also the nature of CIS on the platform. The proportion of the UK fleet that is equipped with some type of APC system was around 40% as of 2010 and it is understood that this proportion has since increased.

- Types of passengers – It is likely that the type of passengers on a service would have an influence on the effectiveness of any solutions, in particular the journey purpose and whether they are regular or occasional rail users.
- Boarding flows – Stations with relatively lower boarding flows may be a lower priority compared to other stations, because there may be less potential for making a substantial difference.
- Funding available – The choice of solution will naturally be affected by the availability of funding. However, as discussed in the points above the most expensive solution may not necessarily be the best choice and it may be possible to achieve the same or greater benefits through less expensive alternatives.



## **8 Conclusions, recommendations, contribution and further work**

### **8.1 Conclusions**

This thesis aimed to investigate the hypothesis, “Automatic passenger counting (APC) data has the potential to deliver more even loadings on trains through the provision of new real-time and static solutions; furthermore, such solutions have benefits of reduced dwell times and reduced crowding and are of value to rail passengers and train operators”.

An 11-month sample of data was obtained from a UK train operating company, which included air suspension pressure, the times the doors lock/unlock and timetable information. The research aims were to address the gaps identified in the literature review, while exploiting opportunities offered by data mining techniques and the manipulation of the sample of on-train data.

The concept of 'loading diversity' has been defined as having four components: 'Within coach' (near doors within a coach); 'Inter-coach' (from coach to coach); 'Inter-train' (between trains across the peak period); 'Inter-route' (between different routes). Based on the type of data that was available, this study investigated 'inter-coach' and 'inter-train' occupancy.

Based upon the findings in §5 and §6 and the discussion in §7, it can be concluded that:

- It was possible to use the air suspension data to generate estimated coach-by-coach occupancy on departure. The primary focus of this research was to investigate localised crowding, both inter-coach and inter-train. As such the relative levels of crowding were of greater interest than the absolute levels of crowding. Comparison against ticket counts suggested sensible results sufficient for relative comparisons, although caution must be taken when making any conclusions on absolute levels of crowding, because no robust validation against manual counts has been undertaken.
- There was uneven inter-train occupancy on many train services and these trends were station-specific, i.e. for some stations it was usually one end of the

train that was busiest, whereas for other stations it was the middle of the train where this was the case.

- A cluster analysis on average coach-by-coach occupancy identified stations with similar patterns of inter-coach occupancy, in particular:
  - For southbound services, Cluster A1 comprised stations where on average the middle of the train was quite busy, while the rear of the train was quiet; the opposite was true for Clusters A2 and A3, with varying levels of average occupancy. Cluster A3 had the greatest range in average occupancy of all the clusters, i.e. it was the 'most uneven'. For all other stations (Clusters A4, A5, A6) average occupancy was relatively lower.
  - For northbound services, Cluster B1 comprised stations where on average the middle of the train was quite busy, while other parts of the train were quiet; the opposite was true for Cluster B2. For all other stations (Clusters B3, B4, B5) average occupancy was relatively lower.
- A key finding was that patterns of inter-coach occupancy were often similar for neighbouring stations. This was not a surprise, because in a situation where the net number of passengers alighting is relatively low, if a coach is busy on arrival at a station it will likely also be busy on departure. An important thing to note from this is that if a station has particularly uneven loadings, initiatives at that station in isolation may not be sufficient; i.e. if the problem is being caused by behaviours at a preceding station, it would be better to tackle the root cause of the problem 'upstream'. In other words, prioritising effort to provide solutions at those stations with uneven loadings and high boarding flows may also alleviate crowding at several subsequent stations 'downstream'.
- It was found that uneven loadings often resulted in passengers having to stand while seats were available elsewhere on the train. It was estimated that across all departures, around 2% of passengers were standing while there were sufficient seats for them to sit down. For busy services from one station, this measure was estimated to be 9%, which represents a substantial proportion of passengers at this station who were being unnecessarily inconvenienced.
- Existing inter-coach 'loading diversity' metrics were found to be inadequate and two new measures have been defined and applied to the dataset to

successfully quantify uneven occupancy. The new metrics are a contribution to knowledge that can be used by other researchers working in the field of loading diversity, as well as by train operators to better understand patterns of crowding across their network.

- Estimated boarding times were correlated against the new inter-coach loading diversity metric and the analysis suggested that there is a link between uneven loadings and dwell times. Trains classified as 'very uneven' ('H-L') on departure, i.e. with at least two coaches with occupancy less than 75% and at least two coaches with occupancy greater than 100%, typically had dwell times of approximately five to ten seconds greater than services that were classified as being 'even', with a similar total number of passengers on board. This suggests that a key benefit to industry of inter-coach solutions may be in the potential to reduce boarding times and in turn reduce the number of late-departing services.
- Different predictive models were applied to predict how busy a particular train will be in the future. The Naïve Bayes model using predictor variables of train length, departure time (half-hour periods), day of week and month (excluding school holiday weeks) yielded overall accuracy of 73.8%, which although an improvement over the baseline model (67.3%) is perhaps not sufficiently accurate to be of value to rail users. However, preliminary results for an approach using decision trees were more promising.
- Feedback from staff through a focus group was consistent with the patterns identified from the air suspension data. There was indicative evidence to suggest that main factors to influence 'inter-coach loading diversity' could be: for boarding passengers, the station layout (e.g. location of stairs and platform entrances/exits); for alighting passengers, the type of travellers (e.g. commuters more likely to choose carriage close to exit on arrival). A range of other factors that may influence passenger behaviour were also discussed.
- Feedback from staff on the effects of crowding on trains was grouped into three categories: impacts on passengers; impacts on staff; and impacts on revenue. These impacts would be expected to lead to costs to train operators, for example by contributing to increased delay minutes and more indirect costs, such as reduced customer satisfaction and staff morale.

- A review of solutions identified a number of interventions for reducing uneven loadings on trains, in both metro operations and longer distance inter-city trains. The purpose of this review was to understand what concepts have been proposed or piloted and to explore the extent to which there exists evaluation evidence on their impact. The literature review identified a number of potential measures that could be used to help to reduce uneven loadings, ranging from lower-cost measures that could be implemented in the shorter term to also higher-cost measures that would require longer term investment decisions. Solutions have been categorised into eight categories of intervention, both by the approach used and their desired effect.
- While the review identified a number of approaches that could in principle be adopted, very little published evaluation evidence was found of the impact of such measures, or their costs. A simulation study by the Korean Transport Institute concluded that a 15% reduction in dwell times could be achieved by giving passengers real-time information on where seats are most likely to be available, however no equivalent figure has been identified from an in-service trial.

## 8.2 Recommendations

It can be concluded that APC data can be used to understand patterns in uneven occupancy and loading diversity, which is an important first step before attempting to address the issue with any interventions. Furthermore, the analysis of the three on-train datasets found that uneven inter-coach loadings have a measurable impact on dwell times, with the dwell time cost of uneven occupancy being estimated at between five to ten seconds per station. Based on these findings it is recommended that further actions should be taken to manage passenger boarding and promote more even loadings.

Uneven inter-coach crowding was identified to be prevalent at some but not all stations. It is therefore recommended that efforts to address uneven occupancy should be targeted at particular stations based on insights from existing data, rather than blanket implementation at all stations.

For the stations analysed in this thesis, the various indicators point to one station in particular as being a candidate for some type of intervention. There was evidence to suggest crowding at the rear of the train, with spare capacity towards the front of the train and furthermore that this has a knock-on impact for downstream stations. There

were very clear repeated trends, suggesting that an appropriate solution might be either a static sign (e.g. as in Figure 7) or a variable message sign (activated only at busy times in the afternoon) saying, for example, 'Please walk down platform to find a seat', see §7.1.1.

For stations on a different route, the opposite trend was observed, where often the middle of the train was the most heavily utilised section of the train. It is recommended to pilot similar interventions at this station, although with the messages reversed to encourage people to use the end of the train.

### **8.3 Contribution to scholarship**

This area of study is an emerging field and my thesis contributes to the Literature on this subject in several ways:

- I propose two new metrics to describe inter-coach loading diversity that, unlike existing metrics, contain information relative to the capacity – namely the unordered metric of 'two busiest coaches and two quietest coaches' and the ordered metric of 'Front-Middle-Rear' occupancy.
- I have revealed a link between the inter-coach loading diversity metrics and estimated boarding times, with trains classified as 'very uneven' on departure typically having dwell times of approximately five to ten seconds greater than services that were classified as being 'even' with a similar total number of passengers on board. The methodology used is repeatable for other researchers working with similar datasets on other train services. In particular, it may be possible to obtain more robust estimates where door sensor data is available, rather than just weight-based sensors and door locking data, as in this study.
- I have conducted preliminary investigations into applying classification supervised learning techniques to predict the load factor for a given service and these predictors were an improvement over taking the historical average. With further research such techniques may have potential to power predictive crowding information services, for example on ticketing websites.

## 8.4 Further work

The second part of the hypothesis was that “such solutions have benefits of reduced dwell times and reduced crowding and are of value to rail passengers and train operators”. This thesis went some way towards this by identifying a potential dwell time saving of between five and ten seconds per station on ‘uneven services’ if it was possible to optimise boarding flows; however further work is required in order to fully answer this part of the hypothesis.

A variety of interventions that use APC data in some way have been identified in the literature review; these solutions have been categorised and discussed. However, there was almost no published evaluation evidence on the impact of the various solutions and so there remain many unanswered questions on their impact and effectiveness: whether they would be safe; whether they would work; and if so in which situations they would work; and in turn what benefits they would provide. This lack of evaluation evidence needs to be addressed before such systems are implemented. An outline has been given in §7.3 for a proposed impact evaluation methodology for pilots of the eight categories of solutions. It is recommended any evaluations comprise three stages:

- Stage 1 – For any future pilots a proposed necessary first step is to conduct, similar to as done here, a spatio-temporal analysis of uneven occupancy using data from automatic passenger counters across all stations and train services, applying the new metrics identified. This should be combined with passenger flow data if available to identify the 'root cause' stations. This review of existing data could be complemented with collection of new data in a human factors study in order to take into account the trip purpose and type of passengers to understand the behavioural drivers.
- Stage 2 – Based on the findings of Stage 1, appropriate solutions should be piloted at carefully selected stations and the effects of the interventions on passenger behaviour should be assessed with a 'Before' and 'After' impact evaluation.
- Stage 3 – Following the findings from the pilots in Stage 2, it may be appropriate to implement solutions at other stations. This would need to be accompanied by ongoing monitoring to assess any changes in behaviour as passengers adapt over time.

Regarding the selection of which intervention may be most appropriate and at which station, the recommended overall approach is a 'problem looking for a solution' rather than a 'solution looking for a problem'. Although it may be tempting to opt for the most expensive real-time system, it may be the case that simpler solutions are more appropriate. It is therefore recommended that lower-cost initiatives in the toolkit are at least considered first before opting for higher-cost solutions. The selection of the most appropriate solution would likely depend on the nature of each station and so before selecting the location of pilot and type of solution several factors need to be taken in to account, such as: variability of occupancy patterns; platform layouts; availability of technology; types of passengers; boarding flows; funding available; other factors.

The following specific areas for further work would also be useful extensions to the study.

- The existing analysis should be combined with boarding and alighting flow data, ideally from APC door sensor data if this was available or alternatively from ticket flow data. In particular it would be of value to investigate which of the stations identified by the cluster analysis were the 'root cause', i.e. at which stations the effort should be targeted for solutions to smooth the crowding, see §7.1.1.
- Regarding inter-train predictive crowding information, further work is required on techniques to predict levels of occupancy for a given service, see §7.1.4.
- Passenger surveys could be conducted to investigate in more depth the causes and impacts of loading diversity, see §7.1.5.
- An extension to the study could be to collect manual counts to calibrate the method for generating the air suspension passenger counts, see §7.2.1.



## 9 References

- Adhikari, R. (2011). Big Data, Big Open Source Tools. Retrieved from <http://www.linuxinsider.com/story/71945.html>
- Ball, S. (2011). [Intelligent Transport Systems \(ITS\): latest developments and the use of micro-simulation assessment - \(PPR 606\)](#). Crowthorne: TRL
- Ball, S. D. (2014). [Data differences - Data science in transport: where do we start?](#) TEC Magazine, Vol.55, No.4, December 2014, p159.
- Ball, S. D. (TRL), Tindall, D. (TRL), McMurran, R. (WMG), & Chan, E. (Ricardo). (2009). [Framework Architecture and Classification for Intelligent Transportation Systems \(FACITS\), Task 5 - Issues Report. - \(PPR 566\)](#). Crowthorne: TRL.
- Barber, G. (2012). [Green's Victoria utilise smartphone application to map public transport overcrowding in real time](#).
- Bell, M., Galatioto, F., Hill, G., & Namdeo, A. (2009). Modelling Environmental Impacts Of Traffic Using A New Generation Of Pervasive Sensors. 16th ITS World Congress, Stockholm 2009.
- Bijlsma, M. (2014). [Sustainable Social Network Services for Transport \(SUNSET\) - Project Final Project](#).
- Bike Sharing Map. (n.d.). The Bike-sharing World Map. Retrieved from [www.bikesharingmap.com](http://www.bikesharingmap.com)
- Birmingham City Council. (2012). [Birmingham's Smart City Vision, Establishing Birmingham's Roadmap to a Future City](#).
- Bolbol, A., Cheng, T., Tsapakis, I., & Haworth, J. (2012). [Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification](#). Computers, Environment and Urban Systems, 36(6), 526–537.
- Borg, J. (2014). Public Transport Interest Group Update on smart & integrated ticketing. ITS UK PTIG, December 2014.
- Bramer, M. (2007). Principles of Data Mining. Springer-Verlag London Limited 2007.
- Brownlee, J. (2013). A Tour of Machine Learning Algorithms. Retrieved from <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms>
- C2C. (n.d.). c2c Live app: keeping you on track. Retrieved from <http://www.c2c-online.co.uk/travel-information/c2c-live-app>
- Calabrese, F., Ferrari, L., & Blondel, V. (2014). [Urban sensing using mobile phone network data: A survey of research](#). ACM Comput. Surv., 47, 2, Article 25 (November 2014), 20 pages.
- Capita Symonds. (2012). [DfT announces £1.9m passenger-counting contract](#).
- CarbonDiem. (n.d.). How CarbonDiem works. Retrieved from <http://www.carbondiem.com/how-carbondiem-works.xhtml>
- Ceapa, I., Smith, C., & Capra, L. (2012). [Avoiding the crowds: understanding tube station congestion patterns from trip data](#). UrbComp '12, August 12, 2012. Beijing, China., 134–141.
- Charles, P., & Hale, C. (2009). [Managing peak demand for passenger rail: A literature review](#). In ATRF09.
- Collins, C., Hasan, S., & Ukkusuri, S. V. (2013). [A Novel Transit Rider Satisfaction Metric : Rider Sentiments Measured from Online Social Media Data](#). Journal of Public Transportation, 16(2), 21–45.
- Collins, N. (2014). [Automatic refunds for delayed commuters](#). Telegraph, 27 Jun 2014.
- CommuteLondon. (n.d.). CommuteLondon: a Live Overview of London's Rail Services Based on Customers' Feedback. Retrieved from <http://commutelondon.com/CommuteLondon>
- Corsar, D., Edwards, P., Baillie, C., Markovic, M., Papangelis, K., & Nelson, J. (2013). [GetThere: A Rural Passenger Information System Utilising Linked Data & Citizen Sensing](#). In Proceedings of the ISWC 2013, Vol. 1035, pp 85-88, CEUR Workshop.
- COSMOS Project. (n.d.). [Collaborative Online Social Media Observatory \(COSMOS\): Social Media and Data Mining](#).
- Cox, T., Houdmont, J., & Griffiths, A. (2006). [Rail passenger crowding, stress, health and safety in Britain](#). Transportation Research Part A: Policy and Practice, 40(3), 244–258.
- Crawford, D. (2013). [Apps help passengers avoided overcrowded public transport](#). ITS International, May-June 2013.
- CycleStreets. (n.d.). About CycleStreets. Retrieved from <http://www.cyclestreets.net/about>



- Data Science London. (2014). About @DS\_LDN. Retrieved from <http://datasciencelondon.org/data-science-london>
- De Vos, M. (2013). [Working with NS and ProRail to improve train platform information](#).
- DfT. (2005). [ITS Policy Framework for the Roads Sector](#).
- DfT. (2013a). [Rail passenger numbers and crowding statistics, notes and definitions](#).
- DfT. (2013b). [Top 10 overcrowded train services: England and Wales, spring and autumn 2012](#).
- DfT. (2013c). [Norfolk bus trial to inform government's smart ticketing strategy](#).
- DfT. (2013d). [Road Transport Forecasts 2013, Results from the Department for Transport's National Transport Model](#).
- DfT. (2014). [Rail passenger numbers and crowding on weekdays in major cities in England and Wales: 2013](#).
- DfT. (2015). [The Pathway to Driverless Cars, Summary report and action plan](#).
- Dilax (n.d.) [Seat management](#)
- Dobson, E. (2014). [Real-Time Journey Prediction From Car-Sensor Data](#). Advanced Data Visualization Group, Nov 2014.
- EC. (2010). [Directive 2010/40/EU of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport](#).
- EC. (2015). [Public consultation, Provision of EU-wide multimodal travel information services under the ITS Directive 2010/40/EU](#).
- Enninga, T., Manschot, M., van Gessel, C., Gijbels, J., van der Lugt, R., Visser, F. S., ... Godfroij, B. (2013). [Service design: insights from nine case studies](#).
- Ferrari, L., Berlingerio, M., Calabrese, F., & Curtis-Davidson, B. (2013). [Measuring public-transport accessibility using pervasive mobility data](#). IEEE Pervasive Computing, 12(1), 26–33.
- Fiat. (2010). [Eco-driving uncovered, full report 2010](#).
- FixMyTransport. (n.d.). FixMyTransport. Retrieved from <http://www.fixmytransport.com>
- Froehlich, J., & Krumm, J. (2008). [Route Prediction from Trip Observations](#). Proceedings of SAE World Congress, 2193(2008-01-0201), 53.
- Fujiyama, T. (2014). Private correspondence.
- Fujiyama, T., Nowers, J., & Tyler, N. (2008). [Investigation into train dwell time, research report submitted to UK Department for Transport – Executive Summary](#).
- Gault, P., Corsar, D., Edwards, P., Nelson, J., & Cottrill, C. (2014). [You'll Never Ride Alone: The Role of Social Media in Supporting the Bus Passenger Experience](#). In Proceedings of Ethnographic Praxis in Industry Conference Volume (EPIC 2014), Issue 1.
- Gillan, B. (2009). The price of progress. Public Service Review: Transport. Issue 23, pp64.
- Graham, A., & Petrie, A. (2012). [The Origin of Species](#). ITS UK Newsletter, Sep 2012, pp8–11.
- Grant-Muller, S. (2015). Social media in sustainable transport management, TSUG seminar on big data, 21 January 2015.
- Grant-Muller, S. M., Gal-Tzur, A., Minkov, E., Nocera, S., Kuflik, T., & Shoor, I. (2014). [The Efficacy Of Mining Social Media Data For Transport Policy And Practice](#). TRB 93rd Annual Meeting, 1–19.
- Grant-Muller, S. M., Gal-Tzur, A., Minkov, E., Nocera, S., Kuflik, T., & Shoor, I. (2015). [Enhancing transport data collection through social media sources: methods, challenges and opportunities for textual data](#). Intelligent Transport Systems, IET, 9(4), 407–417.
- Harding, M., Finney, J., Davies, N., Rouncefield, M., & Hannon, J. (2013). [Experiences with a social travel information system](#). In Proceedings of UbiComp '13 (p. 173). ACM Press.
- Harford, T. (2014). [Big data: are we making a big mistake?](#) Financial Times.
- Hemminki, S., Nurmi, P., & Tarkoma, S. (2013). [Accelerometer-based transportation mode detection on smartphones](#). In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys '13 (pp. 1–14). ACM Press.
- Hey, T., Tansley, S., & Tolle, K. (2009). [The Fourth Paradigm: Data-Intensive Scientific Discovery](#). Microsoft Research.
- HM Government. (2012). [Open Data White Paper: Unleashing the Potential](#).
- Holland, M. (2015). [Swedish big data project predicts train delays before they happen](#).

- Holthuis, J. (2013). [Rotterdam Design Public Prize 2013 - Edenspiekermann, ProRail, STBY en NS Reizigers - Dynamic travel information on platforms](#).
- Hopkin, J., Ball, S. D., Hannay, A., Hutchins, R., Palmer, D., Rahman, S., ... Longhi, D. (2014). [Pilot trial of a transport crowdsourcing smartphone app: final report – \(PPR 719\)](#). Crowthorne: TRL
- Hummels, C. (2013). [Edenspiekermann, STBY, ProRail and NS Reizigers for Dynamic boarding information above train platforms](#).
- IBM. (2013). [Infographic: The Life of the Connected Commuter](#).
- IBM. (n.d.). Smarter Cities. Retrieved from [http://www.ibm.com/smarterplanet/us/en/smarter\\_cities/overview](http://www.ibm.com/smarterplanet/us/en/smarter_cities/overview)
- IFSTTAR. (2012). [International Workshop December 5, 2012, Treatment of spatio-temporal data for a better understanding of urban mobility](#).
- Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., & Kitsuregawa, M. (n.d.). [Visualization of Passenger Flows on Metro](#). The University of Tokyo.
- JustPark. (n.d.). About JustPark. Retrieved from <https://www.justpark.com/about>
- Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). [Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system](#). Pervasive and Mobile Computing, 6(4), 455–466.
- KDD. (n.d.). [KDD Lab Publications](#), Filters: Mobility-profiling.
- Kelly, J., & Hamm, S. (2013). [Smart Machines IBM'S Watson and the era of cognitive computing](#). Columbia University Press.
- Kiefer, C., & Behrendt, F. (2015). [Smart e-bike monitoring system: real-time open source and open hardware GPS assistance and sensor data for electrically-assisted bicycles](#). IET Intell. Transp. Syst., pp. 1–10.
- Kim, H., Kwon, S., Wu, S. K., & Sohn, K. (2014). [Why do passengers choose a specific car of a metro train during the morning peak hours?](#) Transportation Research Part A: Policy and Practice, 61, 249–258.
- Kincaid, C. (2014). [How to be a Data Scientist Using SAS](#). Kalamazoo, MI.
- Knowledge Transfer Network & Smith Institute. (2014). [Data Science: Exploring the Mathematical Foundations](#).
- Kotsiantis, S. B. (2007). [Supervised Machine Learning : A Review of Classification Techniques](#). Informatica, 31, 249–268.
- Kumar, N. (2010). [Smart Display Shows Passenger Density In Trains](#).
- Lathia, N., & Capra, L. (2011). [Mining Mobility Data to Minimise Travellers' Spending on Public Transport](#). In KDD'11, August 21–24.
- Lathia, N., Ahmed, S., & Capra, L. (2012). [Measuring the impact of opening the London shared bicycle scheme to casual users](#). Transportation Research Part C: Emerging Technologies, 22, 88–102.
- Lathia, N., Froehlich, J., & Capra, L. (2010). [Mining public transport usage for personalised intelligent transport systems](#). In Proceedings - IEEE International Conference on Data Mining, ICDM (pp. 887–892).
- Lathia, N., Smith, C., Froehlich, J., & Capra, L. (2013). [Individuals among commuters: Building personalised transport information services from fare collection systems](#). Pervasive and Mobile Computing, 9(5), 643–664.
- Lee, J. S., Cho, H. S., Bae, C. B., & Lee, J. J. (2012). [Improving the Urban Railway Signal System and Operating Efficiency](#). KOTI World Brief, 4(39), 10–11.
- Lee, Y. C., Daamen, W., & Wiggeraad, P. (2007). Boarding and alighting behavior of public transport passengers. In TRB 2007 Annual Meeting. Washington.
- Legara, E. F., Monterola, C., Lee, K. K., & Hung, G. G. (2014). [Critical capacity, travel time delays and travel time distribution of rapid mass transit systems](#). Physica A, 406, 100–106.
- Li, M., Dai, J., Sahu, S., & Naphade, M. (2011). Trip analyzer through smartphone apps. In Proceedings of 19th ACM SIGSPATIAL (GIS '11).
- Libby, R. (2008). A simple method for reliable footstep detection on embedded sensor platforms. Sensors Peterborough NH, (June 25), 1–16.
- London Assembly. (2015). [The London Assembly Transport Committee's response to four Network Rail Route Study drafts](#), 23 January 2015.
- Lyons, G. (2012). Social Media in Transport Evolution or Revolution. ITS UK. Social Networks in Transport, 15 November 2012, London.

- Macal, C. M., & North, M. J. (2010). [Tutorial on agent-based modelling and simulation](#). Journal of Simulation.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). [Big data: The next frontier for innovation, competition, and productivity](#).
- Manzoni, V., Maniloff, D., Kloeckl, K., & Ratti, C. (2011). [Transportation mode identification and real-time CO2 emission estimation using smartphones](#).
- Mims, C. (2013). [Most data isn't "big" and businesses are wasting money pretending it is](#).
- Mitchell, T. M. (2006). [The Discipline of Machine Learning](#).
- Network Rail. (2011). [Network RUS Stations](#).
- Network Rail. (n.d.) Performance. Retrieved from <http://www.networkrail.co.uk/about/performance>
- Ng, A. (2013). [Machine Learning online course, Lecture 1](#). Stanford University, Coursera.
- Nielsen, J. (2006). [The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities](#). Nielsen Norman Group.
- NS. (2013). NS iNStApp - Real-time occupancy information. Retrieved from <https://www.youtube.com/watch?v=gVKM158XSI4>
- ONS. (2010). ['Average' Briton highlighted on UN World Statistics Day](#).
- Opencapacity. (n.d.) Design for all. Retrieved from <http://opencapacity.co.uk>
- ORR. (2012). [The impact of publishing more information on seat availability: South West Trains case study](#).
- ORR. (2014). Crowding on trains. Retrieved from <http://orr.gov.uk/what-and-how-we-regulate/health-and-safety/guidance-and-research/passenger-safety/crowding-on-trains>
- PAC. (2010). [Increasing Passenger Rail Capacity](#).
- Path Intelligence. (2015). [Path Intelligence Technology](#)
- Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). [Smart card data use in public transit: A literature review](#). Transportation Research Part C: Emerging Technologies.
- Pervasive and Mobile Computing. (n.d.). [Pervasive and Mobile Computing: aims and scope](#).
- Pickzenik, I. (2013). [Dealing with Crowding in the UK Rail Industry](#). In TSUG congestion seminar May13.
- Rajbhandari, R., Chien, S., & Daniel, J. (2003). [Estimation of Bus Dwell Times with Automatic Passenger Counter Information](#). Transportation Research Record.
- Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2010). [Using mobile phones to determine transportation modes](#). ACM Transactions on Sensor Networks, 6(2), 1–27.
- Renso, C., Spaccapietra, S., & Zima, E. (2013). [Mobility Data: Modeling, Management, and Understanding](#).
- Ross, T. (2012). [Social Media in Transport – The Role of Crowd-Sourced Data](#). 18 Jun 2012, IET.
- RSSB. (2003). [Managing large events and perturbations at stations - initial research report](#). T161. October 2003. RSSB: London.
- RSSB. (2005). [Health and safety effects of crowding](#). T307. RAS/04/DRAF. London: RSSB.
- RSSB. (2009a). [Crowd management on trains a good practice guide](#). T605. London: RSSB.
- RSSB. (2009b). [Management of on-train crowding](#). T605. London: RSSB.
- RSSB. (2012). [Research into station design and crowd management](#). T916. April 2012. RSSB: London.
- RSSB. (2015). [Platform train interface strategy](#). January 2015. RSSB: London.
- Ryan, C. (2012). [New app keeps Londoners in the Loop about the Tube](#). UCL News, July 3 2012. London.
- SBB. (n.d.). SBB Ticket Shop. Retrieved from <https://www.sbb.ch/ticketshop>
- Schrammel, J., Prost, S., & Röderer, K. (2015). [Persuasive Advisor for CO2-reducing cross-modal trip planning \(PEACOX\), D5.2 Guidelines for designing persuasive strategies within the mobility context](#).
- Senin, P. (2008). [Dynamic time warping algorithm review](#). Information and Computer Science Department, University of Hawaii, Dec 2008.
- Siemens. (2009). [Desiro City - Evolution in motion](#).
- Smart Streets Hub. (n.d.). Smart Streets. Retrieved from [http://www.smartstreetshub.com/smart\\_streets](http://www.smartstreetshub.com/smart_streets)

- Smith, C. (2011). Station capacity and safety policy. Retrieved from [https://www.whatdotheyknow.com/request/station\\_capacity\\_and\\_safety\\_poli](https://www.whatdotheyknow.com/request/station_capacity_and_safety_poli)
- Sohn, K. (2013). [Optimizing Train-Stop Positions along a Platform to Distribute the Passenger Load More Evenly across Individual Cars](#). IEEE. T. Intell. Transp., 14(2), 994–1002
- Steinfeld, A., Zimmerman, J., Tomasic, A., Yoo, D., & Aziz, R. D. (2011). [Mobile Transit Information from Universal Design and Crowdsourcing](#). Transportation Research Record: Journal of the Transportation Research Board.
- Stenneth, L., Wolfson, O., Yu, P. S., & Xu, B. (2011). [Transportation mode detection using mobile phones and GIS information](#). In Proceedings of Proceedings of 19th ACM SIGSPATIAL (GIS '11).
- Stockholm Environment Institute. (n.d.). [Cargo vs. Passengers](#).
- Syncromatics. (n.d.). Key West Transit. Retrieved from <http://www.kwtransit.com/map>
- Taylor, D. (2011). [Next-gen London underground display](#).
- TfL. (2015). Crowding Indicators Research Complete. Engineering Community Newsletter, Issue 42, July 2015, p14.
- Thiagarajan, A., Biagioni, J., Gerlich, T., & Eriksson, J. (2010). [Cooperative transit tracking using smart-phones](#). Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems - SenSys '10, 85.
- Transport Focus. (2011). [Putting rail information in the public domain](#).
- Transport Focus. (2012). [Short and Tweet How passengers want social media during disruption](#). June 2012.
- Transport Focus. (2014). [NRPS Spring 2014 main report](#).
- TRB. (2003). [Transit Co-operative Research Program \(TCRP\) Report 100, 2003](#). Transit Capacity and Quality of Service Manual, 2nd Edition, Part 5 Rail Transit Capacity. TRB.
- TRB. (2008). [Transit Co-operative Research Program \(TCRP\) Synthesis 77, 2008. Passenger Counting Systems](#). Transportation Research Board/National Academy Press.
- TSC. (2014). CommonRoute. Retrieved from <https://commonroute.commonplace.is/>
- Upton, L. (2014). [SmartRail Speaks: Dr. Ramon Lentink, Senior Research Leader, NS \(Netherlands Railways\)](#).
- Van Dijk, E. (2014). [Dynamic boarding information on train platforms](#).
- VTPI. (2014). [Online TDM Encyclopaedia](#).
- Wang, W., Attanucci, J. P., & Wilson, N. H. M. (2011). [Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems](#). Journal of Public Transportation, 14(4), 131–150.
- Wardman, M., & Whelan, G. (2011). [Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience](#). Transport Reviews.
- Watt, N. (2013). [Rail firms asked to come clean on overcrowding](#). The Guardian, 26-08-2013.
- Watt, T. (2013). [TransportBuzz - social sentiment for the transport industry](#). Innovate UK, Intelligent Mobility, 19/11/13.
- Wiggenraad, P. (2001). [Alighting and boarding times of passengers at Dutch railway stations: Analysis of data collected at 7 railway stations in October 2000](#). TRAIL Research School: Delft University of Technology.
- Wood, J., Slingsby, A., & Dykes, J. (2011). [Visualizing the Dynamics of London's Bicycle-Hire Scheme](#). Cartographica, 46(4), 239–251.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). [Top 10 algorithms in data mining](#). Knowl Inf Syst (2008) 14:1–37.
- Zaki, M. J., & Meira, W. (2014). [Data mining and analysis: fundamental concepts and algorithms](#).
- Zheng, Y., Chen, Y., Li, Q., Xie, X., & Ma, W. (2010). [Understanding transportation modes based on GPS data for web applications](#). ACM Transactions on the Web, 4(1), 1–36.