



The
University
Of
Sheffield.

A methodology for passenger-centred rail network optimisation

By:

Bob Hickish

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Engineering
Department of Mechanical Engineering

September 2019

Abstract

Optimising the allocation of limited resources, be they existing assets or investment, is an ongoing challenge for rail network managers. Recently, methodologies have been developed for optimising the timetable from the passenger perspective. However, there is a gap for a decision support tool which optimises rail networks for maximum passenger satisfaction, captures the experience of individual passengers and can be adapted to different networks and challenges. Towards building such a tool, this thesis develops a novel methodology referred to as the Sheffield University Passenger Rail Experience Maximiser (SUPREME) framework. First, a network assessment metric is developed which captures the multi-stage nature of individual passenger journeys as well as the effect of crowding upon passenger satisfaction. Second, an agent-based simulation is developed to capture individual passenger journeys in enough detail for the network assessment metric to be calculated. Third, for the optimisation algorithm within SUPREME, the Bayesian Optimisation method is selected following an experimental investigation which indicates that it is well suited for 'expensive-to-compute' objective functions, such as the one found in SUPREME. Finally, in case studies that include optimising the value engineering strategy of the proposed UK High Speed Two network when saving £5 billion initial investment costs, the SUPREME framework is found to improve network performance by the order of 10%. This thesis shows that the SUPREME framework can find 'good' resource allocations for a 'reasonable' computational cost, and is sufficiently adaptable for application to many rail network challenges. This indicates that a decision support tool developed on the SUPREME framework could be widely applied by network managers to improve passenger experience and increase ticket revenue. Novel contributions made by this thesis are: the SUPREME methodology, an international comparison between the Journey Time Metric and Disutility Metric, and the application of the Bayesian Optimisation method for maximising the performance of a rail network.

Acknowledgements

I am indebted to David Fletcher and Robert Harrison for their supervision, guidance and patience. I am thankful to other staff and students at the University of Sheffield who have also provided help and stimulating discussion: Ozgun Sunar, Jon Paragreen, Rob Chisolm and Paul Richmond. My thesis mentor, Nick Weston, has encouraged me and given me many useful techniques to aid me with writing. The project has received support from various people within industry, for which I am grateful to Meena Dasigi and Nadia Hoodbhoy at Network Rail, David Gill at Siemens and Nigel Kelt, David Winslett and Ken Lamacraft at London Underground Limited.

For supporting me on the way I am incredibly grateful to Linda and Tam, Heliya, Theo, Steph, Chris, Malcolm and Howard.

This project has been funded by Network Rail and EPSRC grant number EP/M508135/1.

Table of Contents

Abstract	i
Acknowledgements.....	ii
Terminology.....	viii
Abbreviations	ix
Symbols	xi
Chapter 1 Introduction	1
1.1 Research question and scope	2
1.2 Method, research objectives and chapter outline.....	3
1.3 Contribution.....	7
Chapter 2 Literature review.....	8
2.1 Rail network management challenges.....	8
2.2 Optimisation methods.....	10
2.3 Metrics for assessing rail networks.....	13
2.3.1 Established journey metrics and network metrics	15
2.4 Network models.....	17
2.5 Heuristic optimisation algorithms.....	19
2.6 Summary and research direction	23
Chapter 3 A metric to assess passenger rail networks	25
3.1 Network assessment metrics that capture the passenger perspective.....	25
3.1.1 Calculating an individual journey score	26
3.1.2 Calculating a network score from journey scores	30
3.2 Validation and comparison.....	31

3.2.1	Data sources.....	32
3.2.2	Input data.....	32
3.2.3	Experimental method.....	33
3.2.4	Results.....	35
3.3	Discussion.....	39
3.4	Summary.....	42
Chapter 4 A passenger rail network model.....		44
4.1	The Passenger Rail Model.....	44
4.2	Infrastructure environment.....	46
4.3	Train agents	47
4.3.1	The movement behaviour of train agents	47
4.3.2	The Dispatcher Model.....	51
4.4	Passenger agents.....	52
4.5	Validation of train movement	55
4.5.1	Sensitivity analysis of Train Dynamics Model parameters.....	60
4.5.2	Sensitivity analysis of time step used.....	62
4.5.3	Discussion of results.....	63
4.6	Validation of passenger agent behaviour.....	64
4.7	Testing the performance of the Passenger Rail Model.....	69
4.8	Summary.....	72
Chapter 5 Selecting an optimisation algorithm.....		75
5.1	Genetic Algorithms and Bayesian Optimisation	76
5.1.1	Genetic Algorithms.....	76
5.1.2	Bayesian Optimisation.....	77

5.1.3	Comparing the computational cost of Genetic Algorithms and Bayesian Optimisation.....	80
5.2	Experimental comparison of Genetic Algorithms and Bayesian Optimisation.....	82
5.2.1	The test-tasks.....	82
5.2.2	Formal definition of the test-tasks.....	84
5.2.3	Calculating the value of the objective function	86
5.2.4	Features of the search space	86
5.2.5	The Genetic Algorithm and Bayesian Optimisation implementations	87
5.2.6	Experimental method	89
5.3	Results.....	90
5.3.1	The number of objective function evaluations and target performance	92
5.3.2	Algorithm computation time	93
5.3.3	Computation time of the whole procedure.....	94
5.4	Discussion	95
5.5	Summary.....	96
	Chapter 6 Applying SUPREME.....	98
6.1	Case study one – optimising the timetable of the regional train network around Derby	98
6.1.1	Formal definition.....	100
6.1.2	Constant parameters describing the network.....	101
6.1.3	Capturing the passenger load	102
6.1.4	Experimental method	104
6.1.5	Results	105

6.2	Case study two – optimising the allocation of attributes to trains in the regional network around Derby	115
6.2.1	Formal definition	117
6.2.2	Data input and experimental method	117
6.2.3	Results.....	118
6.3	Case study three – optimising a value engineering strategy for the High Speed Two network	124
6.3.1	Formal definition	126
6.3.2	Parameters capturing the fixed network components.....	127
6.3.3	Parameters capturing the variable network components	128
6.3.4	Parameters capturing the passenger load	130
6.3.5	The savings functions	131
6.3.6	Experimental method.....	133
6.3.7	Results.....	133
6.4	Summary.....	138
	Chapter 7 Conclusions.....	140
7.1	The network assessment metric.....	144
7.2	The network model.....	146
7.3	The optimisation algorithm	149
	References.....	151
	List of personal communications.....	169
	Appendix I.....	170
	Appendix II	171
	Demonstration of expected power fit in Section 4.5.1	171
	Parameter values for different trains referred to in Section 4.5.3.....	172

Discussion of signalling systems referred to in Section 4.2.....	172
Appendix III.....	174
Determining Davis formula coefficients used in Chapter 6.....	174
High Speed Two savings functions used in Section 6.3	176

Terminology

Agent	A software entity representing a real-world passenger or train
Candidate	A network evaluated by an optimisation algorithm
Challenge	A real-world situation which requires action from rail network managers
Conditions	Factors experienced by a passenger during a journey stage
Crowding penalty	Addition to the Value of Time weighting owing to crowding
Job	A specific combination of a task and minimum acceptable solution quality
Journey metric	A metric to assess the journey of an individual passenger
Journey score	The output of a journey metric, quantifying the experience of a single passenger with a single parameter
Journey stage	A passenger activity within the model of a journey
Network metric	A metric to assess a whole network considering all passengers
Network score	The output of network metric quantifying the performance of the network, from the perspective of all passengers using it, with a single parameter
Passenger state	A specific combination of journey stage and conditions
Solution	The best network returned by an optimisation algorithm
Task	A model optimisation problem

Abbreviations

ABM	Agent-Based Model
AEI	Access, Egress, Interchange
BO	Bayesian Optimisation
BR	British Rail
BRaVE	Birmingham Rail Virtual Environment
CBA	Cost-Benefit Analysis
CSS	Customer Satisfaction Survey
DM	Disutility Metric
DM(UN)	Disutility Metric (Un-normalised)
DMU	Diesel Multiple Unit
EMT	East Midlands Trains
EMU	Electric Multiple Unit
ERTMS	European Rail Traffic Management System
ESEE	Estimated Station Entry Exit
GA	Genetic Algorithm
GB	Great Britain
HS2	High Speed Two
ILP	Integer Linear Program
JTM	Journey Time Metric
JTM(UN)	Journey Time Metric (Un-normalised)
LCB	Lower Confidence Bound
LMA	Limit of Movement Authority
LP	Linear Program
LUL	London Underground Limited

MILP	Mixed Integer Linear Program
ODM	Origin Destination Matrix
ORR	Office of Rail and Road
PB	Passenger Behaviour
PDA	Performance Data Almanac
PRaM	Passenger Rail Model
PS	Particle Swarm
RODS	Rolling Origin Destination Survey
S2AM	Siemens-Sheffield Advance Multimodal Simulator
SA	Simulated Annealing
SM	Signalling Model
SUPREME	Sheffield University Passenger Rail Experience Maximiser
TCM	Train Control Model
TDM	Train Dynamics Model
TS	Tabu Search
UCB	Upper Confidence Bound
UK	United Kingdom
WTT	Working Timetable

Symbols

$\mathcal{A}, \mathcal{B}, \mathcal{C}$	Davis formula coefficients
c_1 to c_3	JTM constants
C_0	Door opening and closing time
d	Distance travelled
$f(\mathbf{x})$	General objective function
F_1^*	Minimum acceptable solution quality
$F_1(\mathbf{x}; \lambda, \theta)$	Objective function found in Chapter 5
$F_2(\mathbf{x}; \lambda, \theta)$	Objective function found in Chapter 6
$F_{available}$	Tractive force available
F_{max}	Tractive force available at rest
$F_{resistance}$	Resistance force
$g(\mathbf{x}), h(\mathbf{x})$	General constraint functions
$G(\mathbf{x})$	Total savings function
$G_C(\mathbf{x})$	Train comfort savings function
$G_L(\mathbf{x})$	Permissible line speed savings function
$G_T(\mathbf{x})$	Rolling stock savings function
$G_V(\mathbf{x})$	Station comfort savings function
H	Crowding factor
i	Optimisation variable index
j, k	Constraint function indexes
J, K	Number of constraint functions
\mathcal{L}	Number of lines
m, p	Observation index
m_C, m_L, m_V	Fitting parameters

M_E	Effective mass
M_T	Tare mass
M_P	Mass of passengers
n	Number of optimisation variables within x
P	Population size of a GA
\wp	Number of lines speed choices
q_1 to q_7	DM constants
r	Passenger index
r	Rotary allowance
R	Number of passengers
s	Passenger state index
S	Number of passenger states
t_s	Time spent in s^{th} state
T	Number of trains
v_0	Boundary speed
x	Optimisation variable
x	Vector containing optimisation variables
x^*	Optimum vector
X	Search space
α_s	Journey stage of the s^{th} state. Can be superscripted with JTM or DM
β_s	Conditions of the s^{th} state. Can be superscripted with JTM or DM
γ	Cost of a single objective function evaluation
Γ	Total cost of all objective function evaluations. Can be superscripted with BO or GA

δ	Number of passengers
δ_{\max}	Maximum passenger capacity of a train
ζ	Value of Time unit
η	Number of objective function evaluations. Can be superscripted with BO or GA
η_{\max}	Objective function evaluation budget
θ	Passenger load
Θ	Total number of algorithm iterations. Can be superscripted with BO or GA
κ_D	Experimental parameter multiplying C_0
κ_S	Minimum total savings
κ_V	Experimental parameter multiplying the VoT weighting for the 'Spare Time' stage
κ_Ω	Experimental parameter multiplying the Value of Time weighting for the 'On Train' stage
λ	Fixed network parameters
μ	Number of seats
π	Cost of single iteration of algorithm excluding objective function evaluation. Can be superscripted with BO or GA
Π	Total cost of all algorithm computation excluding objective function evaluations. Can be superscripted with BO or GA
ϖ	Crush capacity
σ_f	Signal standard deviation parameter
σ_i	Kernel length scale
ρ	Spearman's rho
ϱ	Algorithm iteration counter

τ_B	Kendall's rank correlation coefficient B. Can be superscripted with JTM, DM, CSS, JTM(UN) or DM(UN)
ϕ	Network score. Can be superscripted with JTM, DM, CSS, JTM(UN) or DM(UN)
ψ	Individual passenger journey score. Can be superscripted with JTM or DM
Ψ	Distribution of journey scores
ω	Crowding penalty function. Can be superscripted with JTM or DM
Ω	VoT weighting function. Can be superscripted with JTM or DM
\aleph	Number of carriages
$\beta_1, \beta_2, \beta_3$	Model parameters for time passengers spend buying ticket, moving through station and exiting station

Chapter 1

Introduction

In 2017, 8% of Great Britain's (GB) passenger journey miles were undertaken by rail (Davis, 2018). Between 2002 and 2017, the number of passenger road miles fell, whereas the rail miles increased 28% (Davis, 2018), and the number of rail journeys is forecast to increase 40% further by 2040 (Carne, 2018). These figures demonstrate that rail travel has an increasingly important role to play in meeting the country's passenger journey needs. To fulfil this role, the rail industry Technical Leadership Group (2017) set targets for the GB network that included improving customer experience (passengers and freight – this thesis considers passengers only). However, traditionally, rail networks have been assessed with train focussed metrics. For example, the GB industry standard Public Performance Measure describes the percentage of services that arrive at their final destination within five (ten for long distance trains) minutes of the timetabled time, but this metric does not capture the effect on passengers if the train arrives late to intermediate stations, or if the train is crowded. Therefore, to accurately monitor changes to passenger experience requires a different approach using a metric that directly captures passenger experience.

Operational resources of a rail network include timetable slots on track sections, junctions and platforms as well as rolling stock and crew. Rail network managers, such as infrastructure managers and train operating companies, may wish to alter the allocation of these resources to improve passenger experience. Two examples of this activity are: re-timetabling trains for better passenger connections and re-allocating rolling stock amongst trains to reduce on train crowding. In some cases no further benefit can be obtained through re-allocation and investment is required to either increase or upgrade the resources available. To illustrate the capital required for investment: upgrading 110 miles of the GB West Coast Mainline to increase train throughput has cost £2.8 billion (Committee of Public Accounts, 2017) and purchasing new rolling stock often costs millions of pounds per train (Rail Delivery Group et al., 2018).

Because increasing the operational resources is often expensive, it is usually desirable for network managers to ensure that resources are well allocated and that capital is allocated to the best investment scenario. A decision support tool that optimises resource allocation considering the passenger perspective would allow network managers to select 'good' operational and investment strategies to improve passenger experience cost effectively. From here on in, the word 'methodology'¹ is used to describe the output of this thesis to emphasise that it is a collection of ideas rather than a single, commercially viable, 'product'. Although this thesis incorporates international findings relating to the topic, e.g. a passenger journey metric developed in Japan is investigated, it concentrates on the GB network. Consequently, the institutions referred to are UK institutions unless otherwise stated. Nonetheless, the discussions herein also have relevance to rail networks in other countries where passenger experience is considered by network managers, e.g. across Europe (TNS Political & Social, 2013) and Japan (Kunimatsu et al., 2012). Because the management structure of rail networks varies between different networks, this thesis generalises all network managers as an entity controlling all the components of a rail network, e.g. in the UK no distinction is made between infrastructure managers and train operating companies, implicitly assuming that cooperation between different managing entities can occur.

1.1 Research question and scope

The research question considered by this thesis is: can a methodology be developed for optimising resource and investment allocation, explicitly for maximum improvement to passenger experience? The aim of this thesis is to develop such a methodology and investigate its value, applicability and limitations. There are three points defining the scope of this thesis:

- 1) **An adaptable methodology:** The GB rail network is large and can be abstracted to smaller sub-networks. Furthermore, on a global

¹ Defined by the Oxford University Press (c2019) to mean 'a system of methods used in a particular area of study or activity'.

perspective, there are many other international rail networks and associated sub-networks. Each of these networks may have their own resource and investment allocation challenges, therefore there are many associated optimisation tasks. This thesis does not aim to solve a specific task on a specific network, it considers developing a general methodology that can be adapted and applied to many different specific challenges.

- 2) **Optimising resource allocation:** This thesis considers optimising resource allocation between the components of a network. Optimising the design of a single component in isolation, e.g. the suspension of trains for better passenger comfort, will not be considered.
- 3) **Individual rail passengers:** The primary objective is to improve passenger experience. Other customers to the network (freight) are not considered and nor are other targets such as those for network capacity (Technical Leadership Group, 2017). Passenger journeys are modelled individually, but simultaneously, so that factors such as waiting times, journey times and on-train crowding can be captured. Furthermore, Goodman and Takagi (2004) describe how information from individual passenger journeys could be used to personalise the service provided to them. The recent widespread availability of increased computational power makes it timely to consider individual passenger journeys, despite the increase in computational cost. Factors such as safety and passenger information are not considered.

1.2 Method, research objectives and chapter outline

In this thesis, a rail network is defined to be a specific combination and configuration of components including, but not limited to, the network topography, infrastructure properties, rolling stock, the timetable and sub-components of these. Passengers use the network to make their journey, but are not part of it. In a resource allocation task, these components can either be the resources-to-be-allocated, e.g. the rolling stock allocated to trains, or are sensitive to the allocation of investment, e.g. the comfort of rolling stock.

Consequently candidate ‘allocations’ each define a specific ‘network’ and, in this thesis, ‘optimising the network’ is synonymous with ‘optimising the resource allocation’. To address the research question this thesis develops an optimisation methodology, referred to as the Sheffield University Passenger Experience Maximiser (SUPREME) framework. The procedure of SUPREME is shown by Figure 1 where, after initialising a candidate network, there is an iterative cycle of modelling the candidate network, using the model to assess network performance and selecting a new candidate network. During each iteration a decision is made whether to continue the cycle or terminate and output the ‘best’ network encountered so far, i.e. the *solution*.

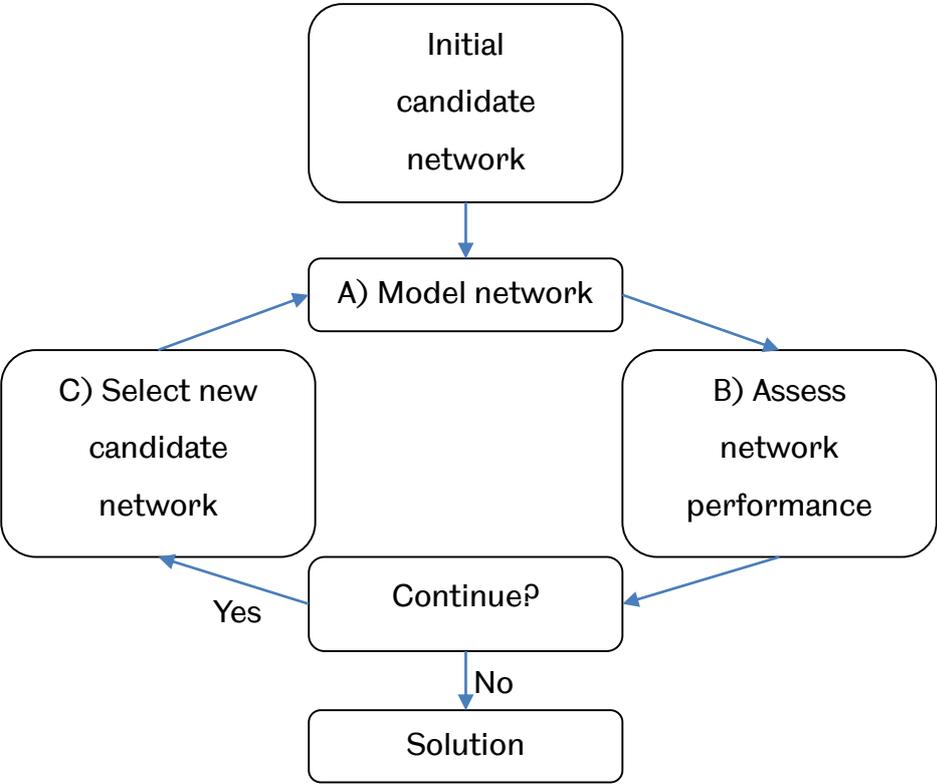


Figure 1 – The flow of processes within SUPREME.

Development of the modules to perform A, B and C in Figure 1 form the majority of this thesis:

- A. **Model network:** It is often time consuming and expensive to alter rail networks, therefore computer modelling is used within SUPREME so that many candidate networks can be assessed. The network model within SUPREME must adequately capture the passenger experience

and be adaptable for different challenges. To fulfil these requirements an agent-based simulation model of train and passenger movements is developed.

- B. **Assess network performance:** A quantitative assessment of network performance is required to differentiate between candidates. Two metrics are identified (one from the UK and one from Japan) that quantify the experience of an individual passenger journey by capturing the multi-stage nature of passenger journeys and the effect of crowding. From these individual journey metrics a whole network metric is developed and validated against data from the Victoria Line of the London Underground Limited network. In this thesis, the network assessment metric is developed first because it informs the design of a network model.
- C. **Select new candidate network:** An optimisation algorithm is required to select new candidate networks. The choice of optimisation algorithm is affected by the computation-cost and formulation of the model. This thesis focuses on investigating and evaluating potential algorithms with an in-depth comparison of the Genetic Algorithm and Bayesian Optimisation methods. Bayesian Optimisation is identified as a suitable method and, to the best knowledge of the author, is previously unused in the context of maximising the performance of a rail network model.

The research aim of this thesis can therefore be broken into four objectives which are the focus of the following chapters:

- In Chapter 3, develop a method to quantitatively assess network performance considering passenger experience.
- In Chapter 4, develop a model of the passenger experience supplied by a network.
- In Chapter 5, identify a suitable optimisation algorithm for selecting new candidate networks and conduct an in-depth study of algorithm effectiveness.

- In Chapter 6, apply SUPREME to case studies and investigate its suitability for the challenges identified. The case studies involve a regional GB sub-network and the planned High Speed Two network. The resources allocated vary between the case studies.

Across the chapters a range of network examples and case studies are used to develop and investigate the SUPREME methodology. This strategy, rather than focussing on just one, is chosen to demonstrate that the SUPREME methodology is adaptable and not specific to one network or challenge type. This is consistent with the project scope described in Section 1.1. Consequently, metro, regional, and inter-city networks are investigated, with challenges with long, medium, and short time windows available for decision making.

The specific networks used for the validation experiments presented in Chapter 3 and Chapter 4 are chosen because they capture all the features of the phenomena being modelled, and suitable data sets were available. The models are successfully validated with input parameters gathered from real-world measurements, without the need for 'parameter tuning'. This indicates that the form of the models is suitable for the validation networks. Since the behaviours of the phenomena being modelled remain similar between networks, e.g. passenger experiences accumulate and train motion obeys Newton's laws, this suggests the form of the models can likely be transferred to other networks if appropriate parameters values are used. Input parameter values might need to be re-evaluated for networks in different locations, and the same network at different times. For example, the maximum line speed of some track sections may need altering to reflect line speed upgrades or temporary restrictions.

Whilst Chapter 6 presents findings that can be used to inform addressing the identified challenges, it is stressed that this is not the purpose of including these case studies in this thesis. Rather, the case studies are used as a basis to demonstrate the application of the SUPREME methodology and investigate its

strengths and weaknesses. Since there are similarities in some behaviours of some rail networks with relatable features, it might be attractive to consider transferring the findings of these case studies to other suitable networks. In some cases expert opinion might be sufficient for determining whether two networks are similar enough for the findings to be translated, in other cases modelling might be required to validate this. If there is need and sufficient data for the latter activity, it would be more beneficial to investigate the specific case directly – thus utilising the adaptability of the SUPREME methodology.

1.3 Contribution

The primary contribution of this thesis is a methodology that is adaptable to different rail network challenges and can be used for maximising passenger satisfaction through optimum resource allocation. Furthermore, to the best of the author's knowledge, this thesis also describes the first:

- International comparison between two passenger journey metrics; the Journey Time Metric and Disutility Metric.
- Application of the Bayesian Optimisation method for maximising the performance of a rail network.

The work in Chapter 3 has been described in the paper 'A rail network performance metric to capture passenger experience' published in the Journal of Transport Planning and Management (Hickish et al., 2019). The work in Chapter 5 has been described in the paper 'Investigating Bayesian Optimization for rail network optimization' and accepted for publication in the International Journal of Rail Transportation. The network model and assessment method have been described in the paper 'Maximising passenger satisfaction through optimised train movements', presented at the 2017 Stephenson Conference (Hickish et al., 2017). The author's contribution to these papers is described in more detail in Appendix I. Further to these papers, part of the network model developed in this thesis has been used in an industry-led project to develop the Siemens-Sheffield Advance Multimodal Simulator (S2AM) for simulating passenger journeys through a multi-modal network (University of Sheffield, 2017).

Chapter 2

Literature review

The large number of components to a rail network often mean that there are many possibilities for network alterations, through investment or operational changes, to address challenges it may face. Furthermore, because the components are highly interconnected, often predicting the effect of alterations can be difficult. To assist network managers, prior to this thesis, decision support tools have been developed which capture the real-world challenge and the effect of alterations. In this thesis, the word 'tool' is used regardless of a methodology's development stage, i.e. any of the Rail Industry Readiness Levels described by the UK's Rail Safety and Standards Board (c2019), so does not imply that it is ready to be applied commercially. Whilst improving components in isolation might improve the experience of passengers, given the scope of this thesis described in Chapter 1, this literature review focusses on tools which capture the interconnected nature of networks as well as resource allocation. The remainder of this literature review is organised as follows: the next section gives an overview of some of the challenges identified in rail network optimisation, Section 2.2 discusses optimisation methods and, because the heuristic method is selected, defines the modules of the methodology developed in this thesis. Sections 2.3, 2.4 and 2.5 review the literature surrounding each one of these modules.

2.1 Rail network management challenges

Marinov et al. (2013) describe that it is common to consider the management of rail networks on three levels: strategic, tactical and operational. The strategic level is the long term planning for the network, generally considering how capital should be invested to acquire resources for the network, e.g. extending or improving network infrastructure. The tactical level is the medium term planning and involves allocating existing resources, for example, deciding the station stopping pattern of trains, i.e. line planning. The operational level is short term and deals with the implementation of plans, timetables and

schedules. Table 1 gives some examples of challenges existing at each management level as well as references to tools which address them. The list is not exhaustive but demonstrates the range of challenges and tools encountered in this review. Some challenges can span between two management levels. For example, Goodwin et al. (2016) consider train control at the tactical level, but Howlett and Pudney (1995) consider this at the operational level.

Management Level	Challenge
Strategic	Expanding stations with extra platforms or sidings (Qi et al., 2016)
Tactical	Timetabling trains (Gupta et al., 2016)
	Scheduling train crew (Lin and Tsai, 2019)
	Assigning rolling stock to trains (Cacchiani et al., 2010)
Operational	Re-timetabling trains after disruption (Meng and Zhou, 2014)
	Train control (Goodwin et al., 2016)
	Train priority at junctions (Fan et al., 2012)

Table 1 – Challenges for rail networks identified in this review and references to tools addressing them. The management level that the challenge exists in is shown by the left column.

By considering multiple challenges simultaneously a better network management strategy can be identified than if they are considered separately. This is because the optimum operational plan for a challenge, which does not consider all aspects of rail network operation, might force subsequent optimisations to only consider extremely sub-optimal candidates. For example, Burggraeve et al. (2017) consider line planning and timetabling simultaneously because an optimal line plan does not guarantee a operationally-feasible timetable. Optimising the line plan and timetable sequentially in isolation might cause a line plan and timetable to be created that is sub-optimal, since selecting a line plan without considering the timetables that can accommodate it might result in a choice of line plan that forces a timetable which is sub-optimal. The example of Burggraeve et al. also demonstrates the advantage of tools that have the adaptability to be extended to include further challenges. Few of the tools found in this review have the adaptability to address the range

of challenges and management levels shown in Table 1, thereby confirming one of the aims of this thesis as a gap in the literature.

2.2 Optimisation methods

Owing to the difficulties associated with altering networks, e.g. cost and disruption, it is common for decision support tools to represent the real-world challenge with a simplified model optimisation problem, i.e. task, which is then solved either exactly or approximately. Here, using the formal notation commonly applied by authors such as Chong (2013), the tasks are generalised to consider selecting the optimum value of a vector, \mathbf{x} , which maximises a non-negative *objective function*, $f(\mathbf{x})$. In the case of a rail network, the objective function quantifies the performance of the network and \mathbf{x} represents the components to be optimised. The vector, \mathbf{x} , comprises n elements and exists in the n -dimensional *search space*, X , bounded by the upper and lower bounds (constraints) of each optimisation variable, x_i , $i = 1, 2, \dots, n$, leading to the optimisation task given by (2.1) subject to constraints (2.2) and (2.3):

$$\mathbf{x}^* = \underset{\mathbf{x} \in X}{\operatorname{arg\,max}} f(\mathbf{x}) \tag{2.1}$$

$$g_j(\mathbf{x}) = 0 \text{ for } j = 1, 2, \dots, J \tag{2.2}$$

$$h_k(\mathbf{x}) \leq 0 \text{ for } k = 1, 2, \dots, K \tag{2.3}$$

where \mathbf{x}^* denotes the global optimum. The set of equality constraints, $g_j(\mathbf{x})$, and inequality constraints, $h_k(\mathbf{x})$, are indexed by j and k respectively up to J and K . For a case requiring minimisation of $f(\mathbf{x})$, maximisation of $-f(\mathbf{x})$ would be used (Press et al., 1992). A task is defined by $f(\mathbf{x})$, $g_j(\mathbf{x})$ and $h_k(\mathbf{x})$. In this thesis, a vector evaluated in the optimisation process is referred to as a candidate.

Some optimisation tasks can be classified by their properties. For example, as well as having a linear objective and constraint function, Linear Programs (LP) have continuous variables whereas Integer Linear Programs (ILP) have exclusively integer optimisation variables and Mixed Integer Linear Programs (MILP) have a mixture. Methods to find or approximate \mathbf{x}^* for the above types

of formulation are well-established and the reader is referred to Chong (2013) and Matoušek and Gärtner (2007) for a discussion of these. These task formulations have been popular in tools for rail network optimisation challenges. For example, LP and MILP formulations are used in tools included in reviews of: timetable planning (Cacchiani and Toth, 2012), timetable recovery after perturbation (Visentini et al., 2014), train control optimisation (Yin et al., 2017) and rolling stock allocation (Piu and Speranza, 2014). However whilst some challenges might naturally relate to a LP, ILP or MILP formulation, others do not. For example, Kanai et al. (2011) state that it is difficult to adequately capture the detail of individual passenger journeys with one of these formulations and Yang (2008) states that most real-world problems are non-linear. Caimi et al. (2017) suggest that simplifying assumptions can restrict the applicability of decision support tools. This is because x^* for a task which does not adequately capture the challenge, e.g. a task with an ill-fitting formulation, might not perform well in the real-world, thus diminishing the usefulness of the tool to network managers. Furthermore, the complex interactions between components of a rail network and the introduction of integer variables in associated tasks, can lead to search spaces which are non-convex. This means that the search space can contain *local maxima* as well as a *global optimum*. Figure 2 illustrates the difference between these by showing the value of an example objective function over the complete search space. There are two local maxima and a single global optimum. Many of the gradient based approaches for solving programming formulations can become 'trapped' in a local maxima of the search space, because in the case of maximisation problems, they always move towards greater values of $f(x)$, i.e. 'uphill'. This means that they do not guarantee to find the global optimum.

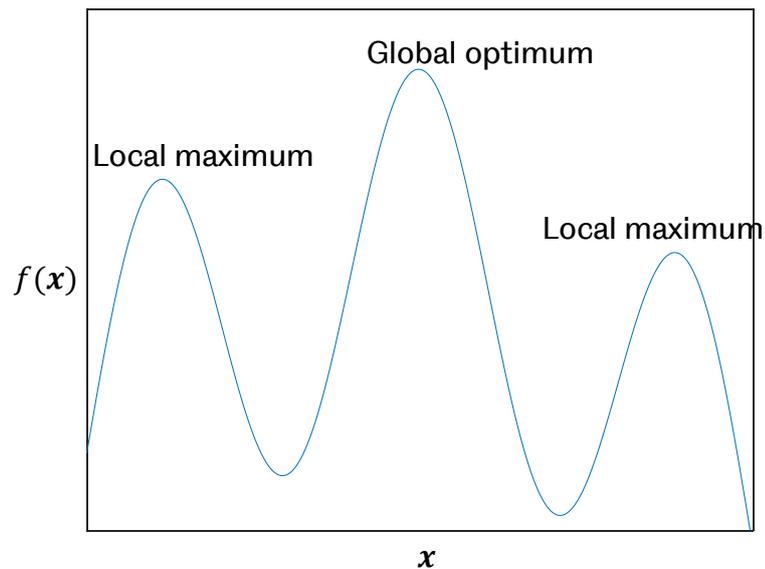


Figure 2 – An illustration of the value of the objective function, $f(x)$, over the complete search space. The maxima are labelled either 'global' or 'local'.

To overcome these difficulties associated with the programming formulations, some heuristic optimisation methods, e.g. Genetic Algorithms and Bayesian Optimisation, can be used which do not place any restrictions on the formulation of the task. However, heuristic optimisation methods do not guarantee to locate x^* , but instead aim to find a 'good' x for a 'reasonable' computational cost. Nonetheless, for many practical optimisation tasks x^* is not a necessity and a user-defined 'good enough' is acceptable. Consequently, heuristic optimisation methods have been widely applied for rail networks, e.g. for crew scheduling (Kokubo and Fukuyama, 2018), train control (Goodwin et al., 2016) and timetabling (Wei and Yuan, 2017). A heuristic method is chosen to be the most suitable for use in SUPREME because its adaptability means that the detail of the real-world challenge can be better captured for a wider range of challenges, thus improving the adaptability of the SUPREME framework and supporting the aims of this thesis. Generally, a heuristic method for rail network optimisation can be described as the combination of three modules: a model of the network dependent on x , a metric to assess the performance of the model network dependent on x , and an optimisation algorithm to select new candidates to evaluate. Sections 2.4, 2.3 and 2.5 respectively review the literature surrounding each one of these modules individually. As the network

assessment metric informs what elements of a network should be modelled, the section relating to the assessment metric is discussed prior to the section discussing the model.

2.3 Metrics for assessing rail networks

Within the methodology of a network optimisation tool it is necessary to make quantitative assessments of candidates. In many cases, the assessment of rail networks has been train focussed. For example, in the case of the timetable optimisation tool presented by Abid (2015), network performance is assessed as the aggregate of train journey times. However, reducing train journey times might negatively affect passenger experience if it leads to worse connections at transfer stations, for example. To fully capture the effect of network alterations upon passengers a network assessment metric should be used which explicitly consider passengers. In this section of the literature review possible methods to do so are discussed.

The aggregate of passenger end-to-end journey time has been used as a metric to assess network performance, for example by Vuchic and Newell (1968), Chang et al. (2000) and Cacchiani and Toth (2012). However, there is evidence that end-to-end journey time does not fully capture the passenger experience. For example Susilo and Cats (2014) show that, for public transport travellers, factors such as station environment, ease of transfer and service frequency are significant determinants of passenger satisfaction. Chen and Chen (2010) describe customer satisfaction as being influenced by customer experience, therefore in this thesis it is assumed that the satisfaction of a passenger is an indicator of their experience, and the effect of other factors such as ticket pricing is disregarded. Consequently, in this thesis, decreasing passenger dissatisfaction or disutility and increasing passenger satisfaction are considered to be equivalent to 'improving passenger experience'. The disconnect between passenger journey time and passenger satisfaction is evident in the results of a rail passenger survey by Transport Focus (2016) which showed that journey time has a smaller influence upon passenger satisfaction than punctuality of the service or its cleanliness. Therefore, to

better capture passenger satisfaction it is necessary to quantify a passenger journey in greater depth than journey time or punctuality alone.

A passenger journey can be modelled as the combination and repetition of specific activities, i.e. *journey stages*. For example, Wang et al. (2015) state that a passenger journey can be well represented with the stages: walking into and out of a station, waiting on the platform, riding on a train and transferring between platforms. However, they do not take into account the relative impact of time in each stage upon the whole passenger experience. Vansteenwegen and Van Oudheusden (2007) and Sels et al. (2016) describe a passenger journey using two stages ('In Station' and 'On Train') and capture the varying impact of time in different stages by weighting these times with a different Value of Time (VoT). The VoT concept has been developed in Transport Economics and describes, in monetary terms, the disutility experienced by a passenger over a time period. It can be thought of as the price a passenger would pay to reduce their travel time by one unit, hence a greater VoT indicates a worse experience for passengers. For further reading on the VoT concept, the reader is referred to Gronau (1970) and Watson (1973). As well as being sensitive to the journey stage of a passenger, a VoT can be sensitive to the mode of transport, journey purpose and distance, for example having different values for travel by car, bus, train or other public transport (ARUP et al., 2015). Wardman (2004) showed that the VoT is sensitive to the activity of the passenger, and Vansteenwegen and Van Oudheusden provide values showing that passengers rate 1 minute of waiting in a station to be equivalent to 2.5 minutes on a moving train. By modelling the amount of time passengers spend in both of these stages and weighting it by the VoT for each stage, Vansteenwegen and Van Oudheusden create a network assessment metric which can capture the relative effect on passengers of time savings in either stage. However, their metric does not capture the effect of *crowding* (i.e. the number of passengers on a train relative to the number of seats and standing space) which can reduce the personal space and comfort of passengers (Preston et al., 2017), causing additional disutility and hence increasing the VoT.

Horowitz (1978) showed that, as well as the journey stage, the 'environmental conditions' that a passenger experiences during a stage (referred to as *conditions* in this thesis) affect the VoT. As well as weather conditions, which are not considered in this thesis, Horowitz considered standing versus seated travel and crowding levels. Models to quantify the impact of crowding upon the VoT have been developed by Wardman and Murphy (2015) and Qin (2014), amongst others. Two metrics developed in different international systems to assess individual passenger journeys across journey stages and crowding levels are the Journey Time Metric (JTM) and the Disutility Metric (DM).

2.3.1 **Cost-Benefit Analysis**

Van Wee (2007) states that when conducting large rail projects it has been common in western countries to conduct Cost-Benefit Analyses (CBAs) as a method to compare the relative merits of proposed network changes. In their guidelines for conducting analyses of transport networks, The Department for Transport (2018a) describe that a CBA entails presenting, in monetary terms, *as many* of the impacts of a proposed scheme as is feasible. It is often necessary to employ modelling to predict the impact of different schemes. For example the MOIRA model described by Worsley (2012), or the commercial software Emme (INRO, c2020) – based on the model described by Florian (1977), could be used to determine the changes to travel demand. Similarly, commercial simulation packages such as RailSys and OpenTrack (described further in Section 2.4) might be used to determine changes in passenger travel times. Some of the 'standard' input parameters needed for modelling and some of the conversion parameters needed to translate the results of modelling to a monetary value, e.g. VoTs, are collected in documents such as the Rail Delivery Group's 'Passenger Demand Forecasting Handbook' described by Worsley and The Department for Transport (2009). However some values may be project specific and require estimation – which can be difficult. The results of a CBA are often presented as a benefit-cost ratio, a metric of investment performance, and investigation is often conducted into the

sensitivity of this ratio to different estimated values used in the calculations (Department for Transport, 2018b).

Because the CBAs used in the rail industry often take into account a large number of effects, they can be difficult to calculate. Consequently, they are often calculated for a 'small' (less than ten) number of investment scenarios, for example in the case of HS2 (Department for Transport, 2017b). The human brain can easily compare the cost-benefit ratios for this number of investment scenarios and search for the optimum – therefore a formal, computer aided, optimisation procedure is rarely employed. However, in the context of SUPREME where a formal optimisation procedure will be used and a larger number of network scenarios investigated, it is not feasible to conduct a full CBA for each candidate. Consequently this method in itself is discounted from further consideration for SUPREME, but the relevant information and concepts described in The Department for Transport's guidelines for conducting CBAs are used to inform the developed metric.

2.3.2 **Established journey metrics and network metrics**

The JTM has been developed by London Underground Limited and shared with the author by private communication, the most informative accessible documentation being the investigations of Chan (2007) and Hickey (2011). It describes passenger journeys using five stages 'Buying Ticket', 'Moving Through Station', 'On Platform', 'On Platform (Left Behind)' (where a passenger has not been able to board a suitable train because it is overly occupied) and 'On Train'. The effect of crowding conditions are considered in the 'On Train' stage by modifying the VoT with a *crowding penalty* that is dependent on the number of passengers, train capacity and seats. The DM has been developed in Japan and is documented in English by Kunimatsu et al. (2009, 2012). It takes a similar approach to the JTM, but resolves a journey using two stages ('On Train' and 'In Station') with weightings different to those used by the JTM. Similar to the JTM, the DM applies a crowding penalty for passengers in the 'On Train' stage that is sensitive to the same factors as the JTM crowding penalty, however a different formula is used. The DM is used again by Kanai et al. (2011) to assess

individual journeys as part of a network assessment metric used in a decision support tool for delay management. They discuss different methods of combining journey scores into a network score, however as presented, none of their methods normalise for the distance travelled by passengers, meaning that networks providing shorter journeys could compare favourably against networks providing longer journeys even if passenger experience is worse.

Moving from individual journey to network metrics, Ali et al. (2017) predict network performance by combining journey scores calculated using an individual journey metric with similarities to the JTM and DM. The network metric is demonstrated to predict observed simple qualitative relationships between timetable features and network performance, e.g. fewer train services result in worse network performance as determined by their metric.

The JTM, DM and the metric described by Ali et al. are the only metrics, found for this review, to capture the multi-stage nature of passenger journeys and weight the time spent in each stage including the effect of crowding. They therefore capture individual passenger journeys in more detail than the other metrics identified here which consider journey stages or crowding only.

However, the parameter values used within the metric of Ali et al. could not be retrieved so this is excluded from further analysis. Consequently, the JTM and DM are selected for investigation as the basis for a metric to assess network performance within the SUPREME framework. To the best of the author's knowledge, no publicly available documents describe the validation or comparison of the JTM and DM, or network assessment metrics based upon them. This gap defines the targets of Chapter 3, to make a comparison of the JTM and DM methods, and to develop a validated network metric based upon them for use within the SUPREME framework.

2.4 Network models

The network model must capture the network features required for the assessment metric. The previous section has explained that the SUPREME framework will use an assessment metric which involves assessing individual passenger journeys. Network models formulated from a program of equations,

i.e. equation-based models, are eliminated from consideration for use within SUPREME because there is evidence that they do not adequately capture real-world journeys. For example, both Yao et al. (2013) and Kanai et al. (2011) state that it is difficult to capture the complexity and interaction of passenger journeys with equation-based models. Similarly, well-established commercial software for simulating train movements, e.g. OpenTrack (OpenTrack Railway Technology Ltd, n.d) and RailSys (Rail Management Consultants GmbH, c2019) are eliminated because Caimi et al. (2017) point out that they do not model passenger movements.

An Agent-Based Model (ABM), sometimes referred to as a Multi-Agent System model, can be used to represent the movements of many individuals. For example the commercially available LEGION simulator (Bentley Systems, c2020) uses an ABM to capture the movements of individual pedestrians. To capture individual passenger journeys, Yao et al. (2013) use an ABM to encode passenger and train behaviours into virtual representations, i.e. agents. By simulating a day's operation of the Beijing network with two million passengers and two thousand trains, Yao et al. demonstrate that their model can capture the complexity of passenger journeys and scale to a large network. However, their model is not combined with a quantitative network assessment metric and therefore is not used for optimisation. The details of ABM are discussed further in Chapter 4 and the reader is directed to Bonabeau (2002) and Macal and North (2010) for more information. Kanai et al. also simulate the interactions of passengers and trains using separate models of the behaviour of each, but do not state whether it as an ABM. Kanai et al. combine their model with the DM to optimise delay management using a Tabu Search algorithm, however no validation is presented and the computational cost of the network model is not discussed. More recently, Chen et al. (2019) have also used an ABM network model in an optimisation of the train timetable, however they do not state the optimisation algorithm used or the computational cost of the network model and optimisation. Nonetheless, the tools of Kanai et al. and Chen et al. give a precedent for using an ABM within a heuristic optimisation.

The ABM method is adaptable and allows any number of agent types to be represented (within available computing power). For example, Li et al. (2011b) show that the ABM method can be extended to model passengers using a train and bus network, i.e. a multi-modal network. More recently, the author was involved in an industry project where ABM was used to simulate a train and bus network including the pedestrian movement of passengers in the station (University of Sheffield, 2017). From this review, the ABM method is considered to be the most suitable method for developing a model to capture individual passenger journeys.

2.5 Heuristic optimisation algorithms

In general, heuristic optimisation algorithms, sometimes also referred to as metaheuristics, use information from previous objective function evaluations to guide the selection of new candidates from the search space until some stopping criteria is reached, e.g. the number of objective function evaluations exceeds a user-defined limit. The x with the maximum objective function value encountered so-far is then returned as the solution. In an overview of heuristic optimisation Yang (2010) describes some of the different methods that can be used to interpret the information from previous evaluations to select new candidates. For example, within the field of rail network optimisation, methods including the following types have been used: Ant Colony (AC) (Sama et al., 2016), Particle Swarm (PS) (Fernandez-Rodriguez et al., 2015), Simulated Annealing (SA) (Burdett, 2015), Tabu Search (TS) (Kanai et al., 2011) and Genetic Algorithm (GA) (Zhu et al., 2017). Elbes et al. (2019) discuss the comparative advantages and disadvantages of the PS, TS, SA and GA methods for 'general engineering' challenges, showing that the relative effectiveness of any given method is task specific. For example, Fan et al. (2012) compare the AC, SA, TS and GA methods for use in a railway re-scheduling tool applied to different tasks. Their results show that the relative performance of solutions found with different methods, is sensitive to the task. Furthermore, the performance of the solutions found with the AC and TS methods are within 10% of those found with the GA method indicating that their effectiveness is

similar. Fan et al. measure the computational cost of the optimisation using computation time of the whole optimisation. However, this measure does not make a distinction between the computational cost of all the objective function evaluations and that of computing the algorithm to select new candidates. In other cases the cost of a single objective function evaluation, and the number required, may be different to that of Fan et al., therefore their findings cannot be generalised.

In particular the Genetic Algorithm method has been widely applied for challenges including timetabling (Xu et al., 2016), train control (Goodwin et al., 2016) and resource allocation (Wang et al., 2019). However, typically these tasks have required the evaluation of 10^4 to 10^5 candidates to find a 'good' solution. Although the performance of an optimisation algorithm is task-specific and sensitive to the value of any control parameters used, there is evidence, described below, to suggest that a similar number of objective function evaluations are required for the AC, PS and SA methods. As for GAs, the AC and PS methods evolve a 'population' of candidates at every iteration with the premise that the population members converge towards optimum areas of the search space. Often, elements of 'randomness' are introduced into implementations of these methods so that, if the population is converging towards a local maxima of the search space, the population can 'escape' the local maxima and 'discover' the area containing the global optimum. The total number of objective function evaluations is computed as the product of the population size and the number of algorithm iterations. For example Caceres et al. (2015) describe that typically AC is used with a budget 10^4 or more objective function evaluations. Similarly, Elbes et al. state that typically PS is used with a population size of approximately 50 and Kennedy and Mendes (2002) find that typically 10^4 iterations are required, meaning that to the order of 10^4 or more objective function evaluations are used. Although the TS method does not use a population, it often evaluates the objective function multiple times per iteration. Considering that Kokubo and Fukuyama (2018) state that the TS implementation they use for a crew scheduling tool requires to the order of 10^3 iterations, this indicates that the TS method also requires a 'large' number of

objective evaluations. Elbes et al. and Fouskakis and Draper (2002) describe that the SA method is poorly-suited to tasks with 'expensive-to-compute' objective functions, indicating that it also requires many evaluations. The term 'expensive-to-compute' is subjective, however, henceforth in this thesis it is used to describe objective functions which require more than a second of computation time.

Owing to the large number of objective function evaluations required, it may be intractable to optimise an expensive-to-compute objective function using the AC, PS, SA, TS or GA methods. Where these methods have been used for rail network optimisation, the computational cost of the optimisation procedure has often been kept reasonable by ensuring that the models used for evaluating candidates require less than approximately ten seconds of computation time. For example, Wei and Yuan (2017) demonstrate the use of a GA implementation for a task involving a comparatively small model network consisting of a single line and 13 stations. However, there are 2560 mainline stations in GB (Steer Davies Gleave, 2017a) and scaling a network model to this size might lead to substantial increase in the computational cost. An optimisation algorithm method which can find a 'good' solution using fewer objective function evaluations would allow models with a greater computational expense to be used.

Another heuristic method, Bayesian Optimisation (BO), only uses one objective function evaluation at every algorithm iteration and consequently might need fewer evaluations for the whole optimisation process than the previously mentioned methods. The BO method is discussed further in Chapter 5 and the reader is referred to Shahriari et al. (2016) for an overview. To the best of the author's knowledge, no publically available documentation exists describing the application of BO to optimise the performance of a model rail network. However, there are many examples of BO being used to select the hyperparameter values of expensive-to-compute machine learning algorithms (Snoek et al., 2012, Jordan and Mitchell, 2015). Applications outside the field of machine learning are less common but can be categorised into two purposes:

- To maximise the agreement between a model and observed data by optimally fitting model parameters.
- To maximise the performance of a real-world entity by optimising design and operational model parameters.

Within SUPREME, an optimisation algorithm is performing the latter activity. An example of this is the use of BO by Candelieri et al. (2018) to maximise the performance of a simulated water distribution network by optimising the pump schedule. Candelieri et al. state they use BO because of its advantages when applied with expensive-to-compute objective functions. For an alternative case, Lisicki et al. (2016) report that BO finds a solution which performs approximately 50% better than that found by arbitrarily selecting an equal number of candidates with uniform probability, i.e. a Random Search. Neither Candelieri et al. or Lisicki et al. make a quantitative comparison of BO against a sophisticated optimisation method such as GA and, to the best of the author's knowledge, the only identified application of BO in a transportation network setting is by Schultz and Sokolov (2018) who optimise the parameters of transportation network simulators to maximise agreement with observed data, i.e. the former activity listed above.

An explicit comparison between GAs and BO is presented by Trotter et al. (2017) to compare both approaches for maximising the performance of a distributed computing system. However, it can be inferred that this comparison is not made for an equal budget of objective function evaluations and is therefore difficult to generalise from. In another comparison of GAs and BO, Chandrashekar et al. (2016) report the number of candidate evaluations in a comparative optimisation of a speech recognition model. However, the use of atypical GA parameters limits the generalizability of the conclusions.

This section has identified three things. First, the GA method is a commonly applied heuristic within the field of rail network optimisation, however, it places restrictions on the computational cost of the objective function. Second, to the best of the author's knowledge, the BO method has not been used for optimising the performance of a rail network and it has the potential to allow an expensive-to-compute objective function to be used. Third, to the best of

the author's knowledge, there is a literature gap for a quantitative, like-for-like, comparison of the GA and BO methods.

2.6 Summary and research direction

This review has shown there is a diverse range of challenges associated with managing rail networks and there is a gap for a tool which can be adapted to many of these challenges. Developing an optimisation methodology that is adaptable is therefore one of the research aims of this thesis. Traditional programming methods for optimisation such as IP, ILP and MILP are excluded from consideration because of this adaptability requirement as well as evidence, described in Section 2.2, suggesting that they are poorly-suited to capturing passenger experience. Instead the heuristic method for optimisation is investigated because it can be applied to a wider range of optimisation tasks. The associated quantitative assessment of different network options can be calculated using a network assessment metric and model. Section 2.3 identifies that the JTM and DM are two metrics that can be used for capturing the experience of individual passengers. In the case of the JTM, to the best of the author's knowledge, there are no publically available documents describing its application for rail network optimisation. In the case of the DM only one has been identified. Furthermore, to the best of the author's knowledge, these two metrics have not been quantitatively compared when applied to the same network. Consequently, the research direction identified is to determine a whole-network assessment metric, based on individual passenger journey experiences captured with either of the JTM or DM, which can be used in the SUPREME framework. A comparison can then be made of the whole-network assessment metric when either the JTM or DM is used to assess individual passenger journeys. To support the use of this network assessment metric, a network model must be developed which can capture individual passenger movements as well as passenger-passenger, passenger-component and component-component interactions. Therefore the research direction also includes developing a suitable network model for use in SUPREME with the network assessment metric. Section 2.4 has identified that ABM techniques are

well suited to this. The heuristic method also requires an optimisation algorithm to explore the search space and Section 2.5 has identified that GAs are commonly used within rail network optimisation but may restrict the computational expense of the objective function. However the BO method might reduce this limitation and, to the best of the author's knowledge, has not been used for rail network optimisation. Consequently, applying a BO implementation for rail network optimisation within SUPREME is another research direction identified in this review. Separately applying a GA implementation within SUPREME will allow comparison of BO with a benchmarking measure and address the literature gap for quantitative, like-for-like, comparison.

Chapter 3

A metric to assess passenger rail networks

The previous chapter has described the need within SUPREME for a metric to assess the performance of model rail networks. It has also described the JTM and DM as two existing metrics for assessing individual passenger journeys. In this chapter a new method is developed which combines assessments of all individual passenger journeys, i.e. journey scores, within a network to give a network score. In a case study relating to the Victoria Line of the London Underground Limited (LUL) network, this whole-network assessment metric is validated against measured data from passenger surveys gathered by LUL (2018a). Additionally, international comparison is made by using the whole-network assessment metric with either the JTM or DM, which originate from different countries. The developed whole-network assessment metric gives a parameter summarising the overall network performance from the passenger perspective, enabling this to be effectively optimised and supporting the aims of this thesis.

Some of the work presented in this chapter is also presented in the author's paper: 'A rail network performance metric to capture passenger experience' published in the Journal of Transport Planning and Management (Hickish et al., 2019). However to the best of the author's knowledge, prior to the above paper, no publicly available documents describe the validation or comparison of the JTM and DM, or network assessment metrics based upon them.

3.1 Network assessment metrics that capture the passenger perspective

To assess a rail network, the metric developed in this chapter evaluates individual passenger journeys and captures the distribution of experiences. To evaluate modelled passenger journeys, the term *state* is introduced to describe a specific combination of journey stage and conditions. A passenger journey is decomposed into a sequence of states as shown in Figure 3, which illustrates an example four-state passenger journey. Shading is used to indicate

which journey stage the passenger is in ('On Train' or 'In Station'). Crowding is only considered in the 'On Train' stage and text is used to indicate this. The markers t_0 to t_4 indicate the times at which the passenger changed state. At t_0 , the passenger enters the origin station and their train at t_1 . At t_2 the train stops at an intermediate station where more passengers board making it crowded. The passenger journey *stage* does not change, but the *state* does. At t_3 the passenger reaches their destination station and exits at t_4 . The number of states in a passenger journey, S , is variable dependant on the journey and the index, s , is used to enumerate the sequence of states, $s = 1, 2, \dots S$.

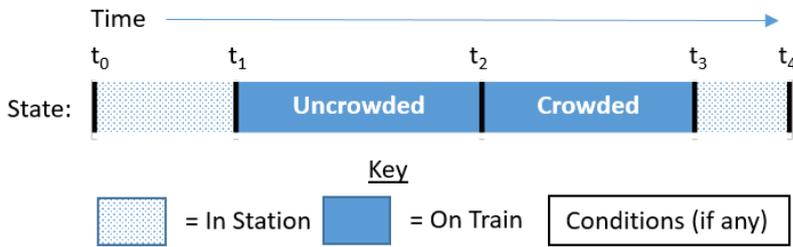


Figure 3 – An example passenger journey decomposed into four states. The journey is described with two stages: 'On Train' and 'In Station'. The shading of the state indicates the stage. Text is used to describe the conditions of the state. The markers t_0 to t_4 relate to the times when the passenger changed state.

The sum of VoT weightings across all states of a passenger journey can be used as an individual journey score. The following section describes how this is calculated when either the JTM or DM is used. The following section also compares how the JTM and DM calculate the crowding penalty. Section 3.1.2 then describes how the distribution of journey scores is evaluated to give a network score.

3.1.1 Calculating an individual journey score

A journey score calculated using the JTM is computed from the formula:

$$\psi^{JTM} = \sum_{s=1}^{s=S} T_s \Omega^{JTM}(\alpha_s^{JTM}, \beta_s^{JTM}, \omega^{JTM}) \quad (3.1)$$

where ψ denotes the journey score, T_s , the time (in seconds) spent in the s^{th} state, Ω , the VoT weighting function, α_s and β_s , respectively the journey stage

and conditions of the passenger's s^{th} state and ω the crowding penalty function. ψ^{DM} (given by (3.2)) is calculated similarly to ψ^{JTM} , but has an additional term to capture the relative disutility experienced by passengers changing train with a parameter for the number of times a passenger must change trains, ε , and a weighting factor, q_1 . A value of 600 is used by Kunimatsu et al. for q_1 , meaning that each train change has an associated disutility equivalent to 10 minutes (600 seconds) travelling on an otherwise unoccupied train. Table 2 provides the other parameter values for each metric.

$$\psi^{DM} = \sum_{s=S}^{s=S} T_s \Omega^{DM}(\alpha_s^{DM}, \beta_s^{DM}, \omega^{DM}) + q_1 \varepsilon \quad (3.2)$$

$\alpha_s^{JTM} =$	1	2	3	4	5
Description	On Train	On Platform	On Platform (Left Behind)	Moving Through Station	Buying Ticket
$\Omega^{JTM} =$	$1 + \omega^{JTM}(\beta_s^{JTM})$	2.5	3	2.7	2.5
$\alpha_s^{DM} =$	1	2			
Description	On Train	In Station			
$\Omega^{DM} =$	$1 + \omega^{DM}(\beta_s^{DM})$	3			

Table 2 – The VoT weighting, Ω , for both metrics dependent on the journey stage, α , of a passenger's s^{th} state. A description of the journey stage relating to α is also shown. The VoT weighting for passengers in the 'On Train' stage is dependent on a crowding penalty function, ω , calculated using the conditions of the state, β . For the JTM, these values have been shared with the author by personal communication from the Transport Planning department of London Underground Limited (Kelt, 2015)², and for the DM they are taken from Kunimatsu et al. (2012).

Table 2 shows the relative weighting both metrics put on each state (a lower value of Ω indicates a better passenger experience) and that the JTM describes a journey using five journey stages whereas the DM uses two. Both methods consider crowding only when passengers are in the 'On Train' journey stage. The JTM crowding penalty, ω^{JTM} , is determined with the formula given by (3.3),

² For more information, see the list of personal communications in the reference list.

using values given in Table 3, where δ denotes the number of passengers, μ , the number of seats on the train, δ_{\max} , the maximum passenger capacity, ϖ , the crush capacity and c_1 to c_3 constants.

$$\omega^{JTM} = \begin{cases} 0, & \delta \leq \mu \\ c_1 + c_2 \frac{\delta - \mu}{\varpi} - c_3 \frac{\delta\mu - \mu^2}{\varpi^2}, & \mu < \delta \leq \delta_{\max} \end{cases} \quad (3.3)$$

The crowding penalty formula given by (3.3) has been shared with the author by personal communication from the Transport Planning department of LUL (Kelt, 2015)³. The second term of (3.3) captures the number of standing passengers relative to the crush capacity of the train and the third term captures the effect of seated passengers also. The value of ϖ describes the theoretical maximum number of passengers that can fit into the train assuming seven passengers per square meter of standing floor space. However, LUL have determined that the practical maximum capacity of a train is less than ϖ and under 'normal operating conditions' the value of δ_{\max} is defined as 71% of ϖ . The DM crowding penalty, ω^{DM} , is determined with the formula given by (3.4) and requires computing the crowding factor, H , given by (3.5). The value of the constants q_2 to q_7 and c_1 to c_3 are shown by Table 3.

$$\omega^{DM} = \begin{cases} q_2 H, & H < 1 \\ q_3 H - q_4, & 1 \leq H < 1.5 \\ q_5 H - q_6, & 1.5 \leq H \leq 2 \end{cases} \quad (3.4)$$

$$H = \frac{q_7 \delta}{\delta_{\max}} \quad (3.5)$$

Name	c_1	c_2	c_3	q_2	q_3	q_4	q_5	q_6	q_7
Value	0.85	1.915	1.03	0.027	0.0828	0.0558	0.179	0.2	2

Table 3 - Constant values used to calculate the crowding penalty, ω^{JTM} and ω^{DM} , in (3.3), (3.4) and (3.5). For the JTM, these values have been shared with the author by personal communication and the DM constants q_2 to q_6 are taken from Kunimatsu et al. (2012). The value of q_7 is informed by Nippon (2018).

³ For more information, see the list of personal communications in the reference list.

The values of c_1 to c_3 have been derived by LUL and shared with the author by personal communication (Kelt, 2015). The values of q_2 to q_6 are listed by Kunimatsu et al. (2012). Although Kunimatsu et al. do not explicitly define η , they describe it as the ‘congestion rate of the train’, therefore it can be inferred as being proportional to δ/δ_{\max} . However because Nippon (2018) report the largest crowding factor (H) observed in Japan during 2017 as 2 (relating to when ‘bodies come into contact with each other and one feels considerable pressure’), the scaling factor q_7 is introduced into (3.5) and given a value of 2. The values of μ , δ_{\max} and ϖ are rolling stock specific and are defined by LUL for each fleet. For the LUL 2009 rolling stock (used on the Victoria Line and the subject of this investigation) their values are 288, 730 and 1028 respectively (Kelt, 2015).

For varying number of passengers (δ), Figure 4 compares ω^{JTM} , ω^{DM} and the minimum VoT weighting for passengers in the ‘On Train’ stage. Values on the y-axis are expressed as a multiple of the VoT for a passenger travelling on an uncrowded train, ζ . The number of seats on the train is shown by a vertical dashed line and reflects that when $\delta \leq \mu$, the JTM does not apply a crowding penalty. A crowding penalty is applied by the DM even at this level of occupancy, but it is small in comparison to the minimum VoT weighting for passengers in the ‘On Train’ journey stage (the dash-dot horizontal line). When $\delta > \mu$, the JTM applies a crowding penalty that is 4 to 8 times greater than the DM crowding penalty. For both metrics, the crowding penalty is always less than the minimum VoT weighting for the ‘On Train’ stage. Both the JTM and DM models of crowding assume that passengers are homogeneously distributed throughout the train and that passengers will always find and occupy a seat if one is available. Although this may not be realistic, it is the same for both models so the comparison is like-for-like. The VoT weightings (in Table 2) and crowding penalty function for the JTM and the DM have been derived for the LUL network and Japanese railway respectively. It is therefore expected for these values to capture local preferences and expectations.

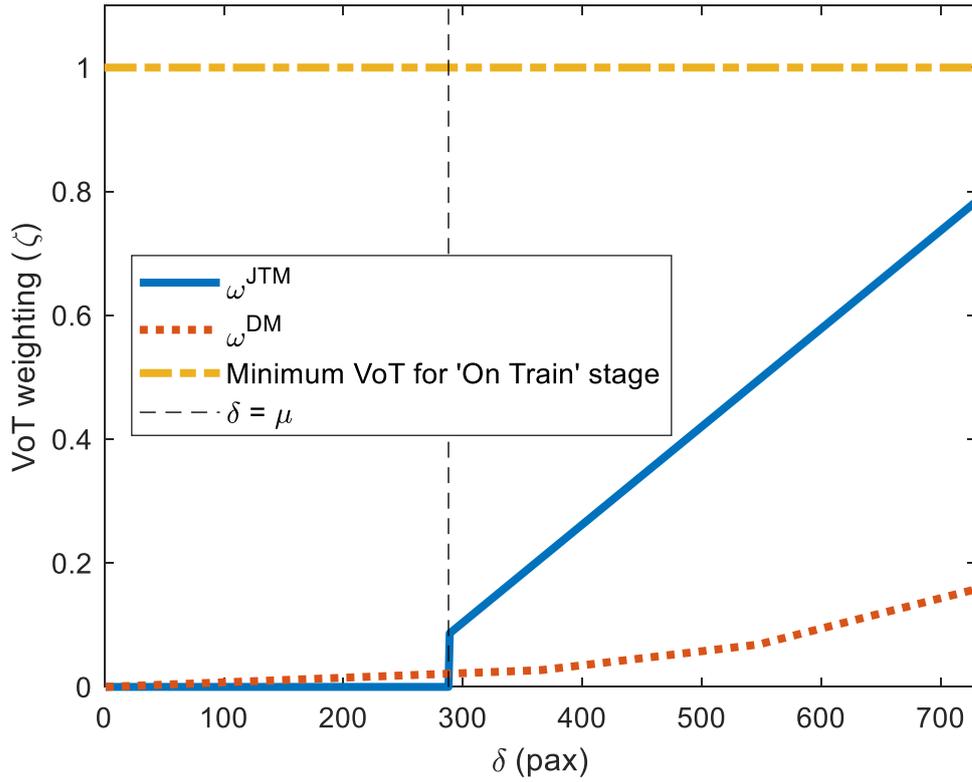


Figure 4 - The crowding penalty, ω applied by the JTM and the DM for different numbers of passengers, δ , in LUL 2009 rolling stock up to its maximum capacity. The number of seats, μ , is shown by a vertical dash line. The minimum VoT weighting applied by both metrics to passengers that are in the 'On Train' stage is shown by a horizontal dash-dot line.

3.1.2 Calculating a network score from journey scores

Networks provide journeys for multiple passengers so there is a distribution of journey scores. To ensure that journey scores only capture the quality of the service provided to the passenger by the network (and not the distance of the passenger journey which is a passenger choice), journey scores are normalised by the distance travelled. This allows like-for-like comparison of journey scores within the distance-normalised journey score distribution, Ψ , given by:

$$\Psi = \left[\frac{\psi_1}{d_1}, \frac{\psi_2}{d_2}, \dots, \frac{\psi_R}{d_R} \right] \quad (3.6)$$

where R denotes the number of passengers, ψ_r the journey score of the r^{th} passenger and d_r the distance travelled by the r^{th} passenger. Different features of Ψ can be used to provide the network score, ϕ , for all R passengers

conveyed by the network. It is desirable to capture the effect of passenger numbers upon crowding, but for the network score to be independent of the number of journey scores within Ψ . Consequently, an additional passenger-number normalisation step is included so ϕ is defined by:

$$\phi = \frac{1}{R} \sum_{r=1}^{r=R} \frac{\psi_r}{d_r}$$

(3.7)

Beyond this network score the characteristics of the distribution of Ψ can offer additional insight. For example, a manager wishing to examine the consistency of their service to passengers taking different journeys may evaluate the range of Ψ in addition to ϕ . This thesis focuses primarily on ϕ to study quality of service provided to all passengers within the network.

3.2 Validation and comparison

To validate the network assessment metric, ϕ values are calculated using either the JTM or DM (ϕ^{JTM} or ϕ^{DM}) for the Victoria Line of the LUL network. For the same network, a network score is determined from measured Customer Satisfaction Survey (CSS) data, ϕ^{CSS} . The predictive values of ϕ^{JTM} and ϕ^{DM} are compared against the measured ϕ^{CSS} values and the correlation between their changes relative to a baseline year is quantified. The predictive values are then compared to each other to determine a relationship between the network assessment metric when either journey score metric is used. To calculate ϕ^{JTM} and ϕ^{DM} , data describing the network was combined with data describing the number of passengers travelling between each station at different times, i.e. the passenger load, and captures the effect of varying timetables and passenger loads over ten years. For the Victoria Line in the period investigated, the formation, length and interior layout of rolling stock remain constant, therefore the frequency of trains (determined by the timetable) has the greatest effect upon the passenger carrying capacity of the network. Decreasing the speed of trains on a line slows travel but also reduces headway with potential to decrease intervals between trains, so typically there is a trade-off between journey times and frequency. To meet increasing

demand for travel, minimise crowding and generate more revenue, whilst maintaining competitive journey times against other transport modes, there is a pressure on LUL to balance this trade-off when updating their timetable.

3.2.1 Data sources

The data sources used in this investigation are: Victoria Line Working Timetable (WTT) numbers 31 to 41 (London Underground Limited, 2007, 2009, 2011, 2012a, 2012b, 2014, 2015a, 2015b, 2016b, 2016c, 2017), Access, Egress and Interchange (AEI) data provided by LUL (2016a), the Performance Data Almanac (PDA) (London Underground Limited, 2018a) and the Rolling Origin Destination Survey database (RODS) (London Underground Limited, 2018b). In the following section, the data is described in more detail.

3.2.2 Input data

The network operation data is taken from the WTTs and the AEI data. For each day, the WTTs provide the average train frequency and interstation run times for the three weekday operational periods on which the investigation concentrates: Morning Peak, Midday Off Peak and Evening Peak. Later operational periods are excluded because their timings are not consistent between the WTTs. The effect of this exclusion is unlikely to be significant because observing the RODS database indicates that this period is when the fewest passengers travel and so it has the least weighting on the network score. Weekends and holidays are not considered because they are more likely to be affected by events (e.g. sporting events or planned line closures for maintenance works) that affect passenger experience but are not captured in all the input data sources. The operational pattern described in the WTT is applied for every day the timetable was in effect (LUL update their timetable irregularly, but the date of introduction is provided by each WTT). The WTTs also provide the distance between adjacent station pairs. The AEI data describes the passenger travel time from station door to platform and vice versa, and platform to platform. The AEI data available relates to every four week period of the year beginning 2011 (the LUL reporting year begins on 1st

April), the annual mean for this period is 2.23 minutes. Because data is only available for one year, this is applied for all years of the investigation, implicitly assuming that personal mobility within the station remains constant over this period.

The passenger load data is a combination of two data sources: the PDA and RODS. RODS provides the proportion of passengers included within the database that travel between adjacent station pairs in an operational period, i.e. *line section loadings*. However, this data does not describe whole passenger journeys (i.e. an origin and destination with any transfer stations). The PDA provides the total number of passengers travelling on the Victoria Line each year, and the quarterly CSS data. To collect the CSS data, LUL use questionnaires to ask approximately 2,500 passengers per quarter to rate, on a scale of 1 to 10, their satisfaction with their travel on the line of the last leg of their journey. The mean of the ratings is then multiplied by 10 and reported for each line by LUL.

Customer satisfaction is affected by factors that are not captured in ϕ^{JTM} and ϕ^{DM} , and hence are not included in this investigation. For example, Paramita et al. (2018) describe that passengers will have a satisfaction associated with the fare they paid for a journey -this will affect their overall satisfaction with the journey. The results presented by Transport Focus (2016) indicate that train cleanliness and 'how train companies dealt with delays' are important drivers of passenger satisfaction and *dissatisfaction*, respectively, yet these also are not captured. Transport Focus also indicate that train punctuality is an important driver of passenger satisfaction. Whilst this factor can be captured by ϕ^{JTM} and ϕ^{DM} (passengers accrue greater journey scores whilst waiting for delayed trains), it is not included in this investigation. Journey length, train frequency, and crowding, are other important drivers of passenger satisfaction identified by Transport Focus which are captured in this investigation.

3.2.3 Experimental method

Figure 5 gives an overview of the data and processes used to calculate values of ϕ^{JTM} , ϕ^{DM} and ϕ^{CSS} for corresponding years so that a like-for-like comparison

could be made. In Figure 5, a rectangular box represents an action and a parallelogram represents a data item. To calculate ϕ^{JTM} and ϕ^{DM} , the line section loading data was scaled by the yearly passenger numbers data and used to disaggregate the journeys of passengers who travelled farther than the station adjacent to their origin, into a series of journeys between adjacent station pairs. For each operational period (Morning Peak, Midday Off Peak and Evening Peak) and line section, the number of passengers per train was calculated by dividing the number of passenger journeys in that period by the number of trains. Where demand for travel exceeded provision, the excess passengers were modelled as being 'left behind' by one train before catching the next. The frequency of trains was used to determine the total passenger time spent in the 'On Train', 'On Platform' and 'On Platform (Left Behind)' stages. The journey score metrics were used to calculate the VoT weighting for these states. To avoid over-counting, the AEI time and weighting was only applied twice for each whole passenger journey defined by the PDA data rather than the RODS data. The 'Buying Ticket' journey stage was disregarded because the use of pre-paid travel cards ('Oyster cards') and contactless payment at ticket gates is common for this network. For example, in 2012 Oyster cards were used for over 80% of public transport travel in London (Transport for London, 2012). The interstation distances were multiplied by the line section loadings so that the aggregate of the VoT weightings could be normalised by the total passenger distance travelled.

Figure 5 shows how the analysis of network performance was conducted for the Morning Peak, Midday Off Peak and Evening Peak operational periods of every weekday and was dependent on the daily timetable and yearly number of passenger journeys. To calculate the network score for that day, the values from the three operational periods of the day were summed. The year value was calculated as the mean of the year's day values. Because the CSS data is already normalised for passenger numbers and distance travelled, it is not relevant to normalise ϕ^{CSS} using (3.7).

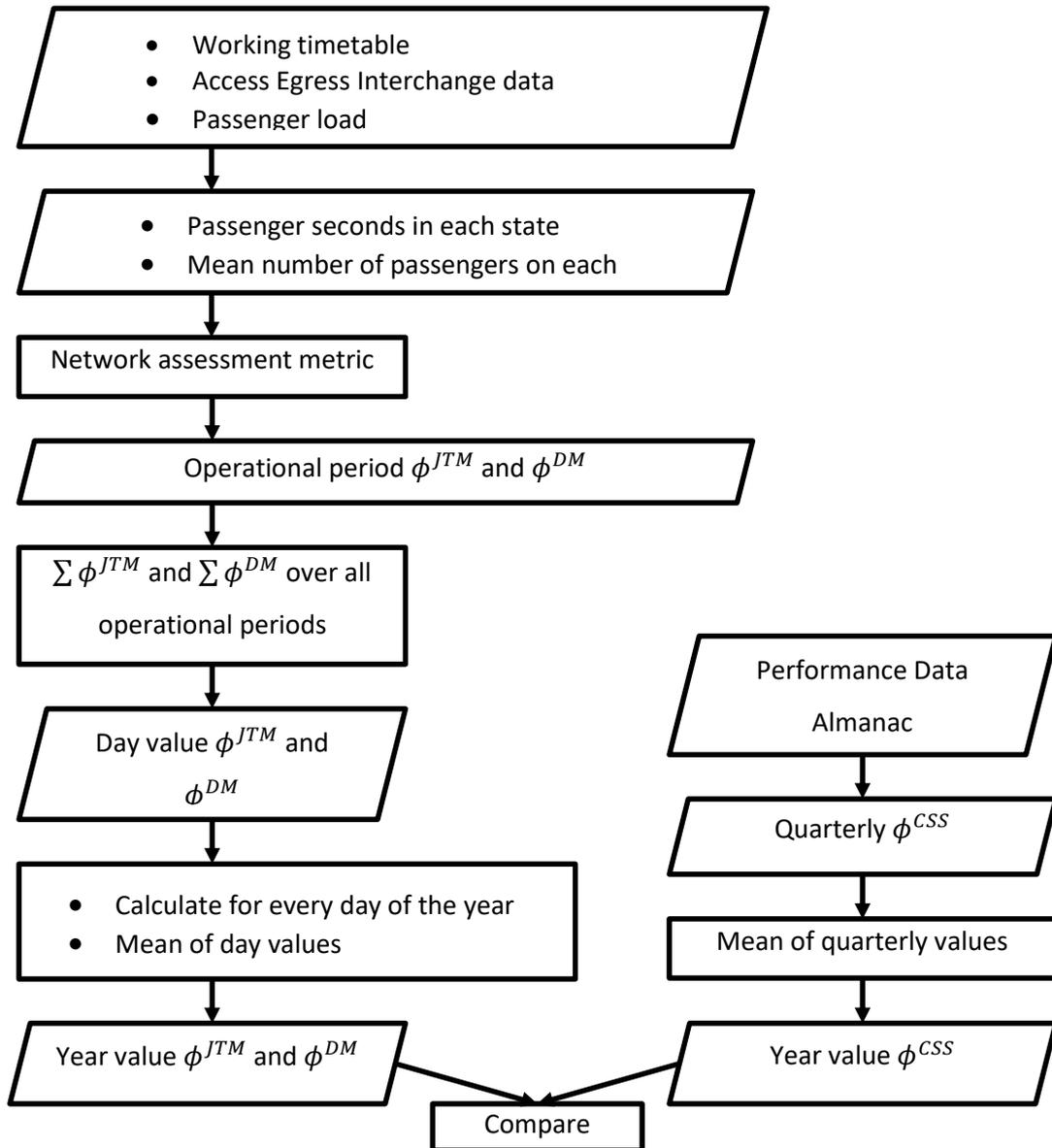


Figure 5 - The method for calculating the measured network score, ϕ^{CSS} , and predicted network score using the Journey Time Metric or Disutility Metric, ϕ^{JTM} and ϕ^{DM} .

3.2.4 Results

Figure 6 enables comparison of ϕ^{CSS} with ϕ^{JTM} and ϕ^{DM} , and also presents data where no distance or passenger normalisation is applied, $\phi^{JTM(UN)}$ and $\phi^{DM(UN)}$, for the years 2008 to 2017. The number of passengers, R , is also included in the plot. Upward-pointing bars with values displayed on the left ordinate are used for ϕ^{CSS} , while $\phi^{JTM(UN)}$, $\phi^{DM(UN)}$, ϕ^{JTM} and ϕ^{DM} are represented by downward-pointing bars with values displayed on the right ordinate. Because the prediction metrics measure dissatisfaction and ϕ^{CSS}

measures satisfaction, the right ordinate is inverted. A positive change in the vertical position of a bar-top for ϕ^{CSS} indicates a 'better' performing network. R is also represented by markers with values displayed on the right ordinate. To allow comparison of relative changes on different scales and using different units, all series have been normalised against their 2008 value. It can be seen that over time, in general, the measured network scores (ϕ^{CSS}) indicate improving network performance, with rising values relative to 2008. In general, this behaviour is successfully predicted by ϕ^{JTM} and ϕ^{DM} . However, $\phi^{JTM(UN)}$ and $\phi^{DM(UN)}$ predict deteriorating network performance and correlate with the increasing passenger numbers. It should be noted that, whilst the prediction metrics appear to give equal scores in 2008, this is because of the series normalisation process. The importance of normalising the predictive values by passenger numbers and distance travelled is clear if the metrics are to be compared over time.

To investigate the importance of applying VoT weightings to different passenger states, Figure 7 enables comparison of ϕ^{CSS} , ϕ^{JTM} , ϕ^{DM} and a simple end-to-end journey time metric, ϕ^{EE} . To ensure like-for-like comparison, ϕ^{EE} has been normalised for passenger numbers and distance. The ordinates are similar to Figure 6 with the right ordinate now displaying ϕ^{EE} normalised against the 2008 value. To quantify the level of agreement between predicted and measured performance, Kendall's rank correlation coefficient B , τ_B , is calculated between the series of ϕ^{CSS} with each series of: ϕ^{JTM} , ϕ^{DM} and ϕ^{EE} . For the series of ϕ^{CSS} with ϕ^{JTM} and ϕ^{CSS} with ϕ^{DM} a value of -0.82 ($P < 0.005$) is found (-1.0 indicates perfect negative correlation between prediction and measurement and 0 indicates no correlation). For the series of ϕ^{CSS} with ϕ^{EE} a value of -0.73 ($P < 0.005$) is found, indicating worse correlation and that the network assessment metric is improved by representing a passenger journey as a series of states and applying weighting to these.

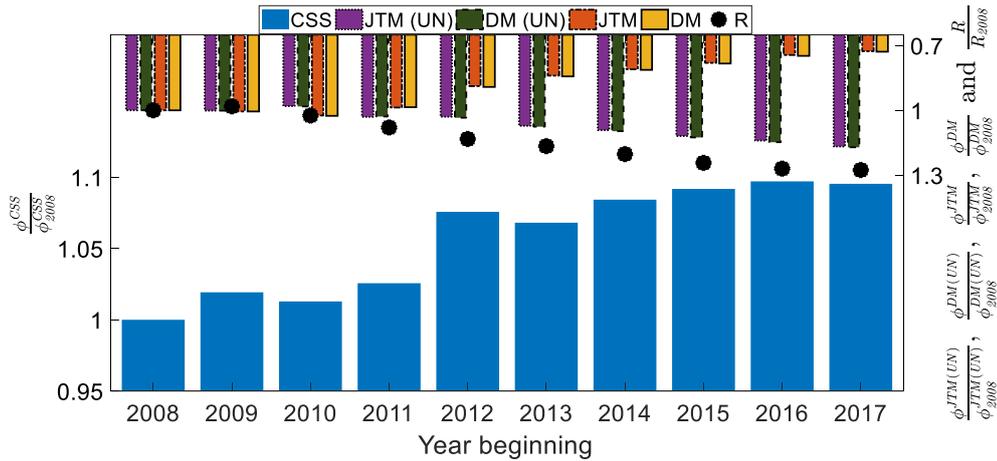


Figure 6 - Bar chart to compare predicted and measured network scores for different years and different prediction methods. Measured customer satisfaction scores, ϕ^{CSS} , are shown by the left ordinate. Predictions using the Journey Time Metric, ϕ^{JTM} , Journey Time Metric with no distance or passenger normalisation, $\phi^{JTM(UN)}$, Disutility Metric, ϕ^{DM} , and Disutility Metric with no distance or passenger normalisation, $\phi^{DM(UN)}$, are shown by the right ordinate which has been inverted. The right ordinate also displays the number of passengers, R . All values have been normalised against the corresponding 2008 value.

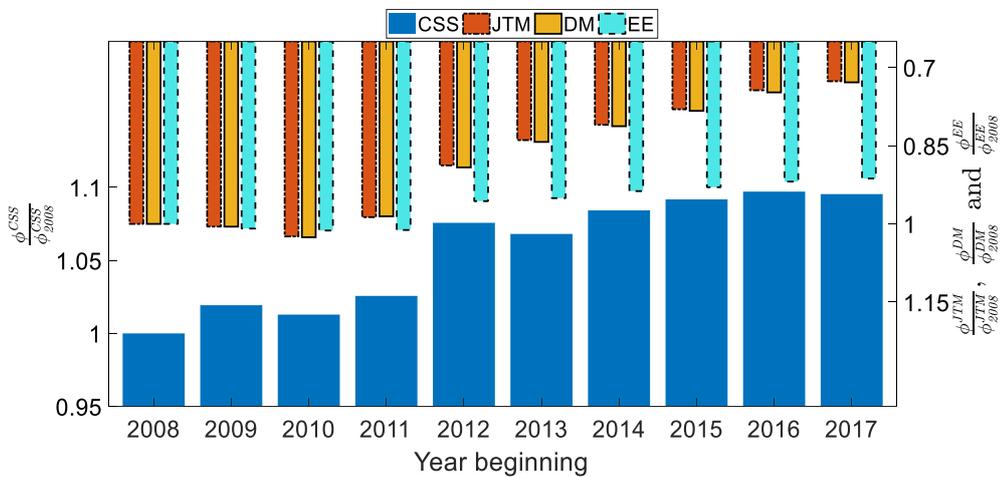


Figure 7 - Bar chart to compare predicted and measured network scores for different years and different prediction methods. Measured customer satisfaction scores, ϕ^{CSS} , are shown by the left ordinate. Predictions using the Journey Time Metric, ϕ^{JTM} , Disutility Metric, ϕ^{DM} , and end-to-end journey time, ϕ^{EE} , are shown by the right ordinate which has been inverted. All year scores have normalised against the 2008 value for the corresponding metric.

To explore the importance of the crowding penalty Figure 8 enables comparison of ϕ^{JTM} and ϕ^{DM} against the case where no crowding penalty has

been applied in the calculation, $\phi^{JTM(NC)}$ and $\phi^{DM(NC)}$. The y-axis displays the raw values of ϕ , i.e. they are not normalised against the 2008 value, which are displayed in units of the VoT for passengers travelling on an uncrowded train (ζ) per passenger per metre. To determine what proportion of the network score is contributed by factors other than the crowding penalty, the value of $\phi^{(NC)}/\phi$ is calculated when either journey metric is used. For the JTM and DM series respectively, a mean value of 0.91 and 0.99 is found both with a standard deviation less than or equal to 0.002. This behaviour is discussed in Section 3.3. Figure 9 plots ϕ^{DM} against ϕ^{JTM} for the data from the years 2008 to 2017. The strong linear relationship of the data ($\rho=1.00$, $P<0.005$) suggests that, in general, similar changes in network performance are predicted by the JTM and the DM. A linear fit to this data shows a gradient of 1.013 (95% confidence bounds of 1.012 and 1.015). The intercept has been forced to the origin because both metrics are zero under the same condition: when no passenger time is spent in the network. The gradient implies that ϕ^{JTM} is consistently approximately 1.3% greater than ϕ^{DM} , but both are reacting consistently to external change over the period investigated.

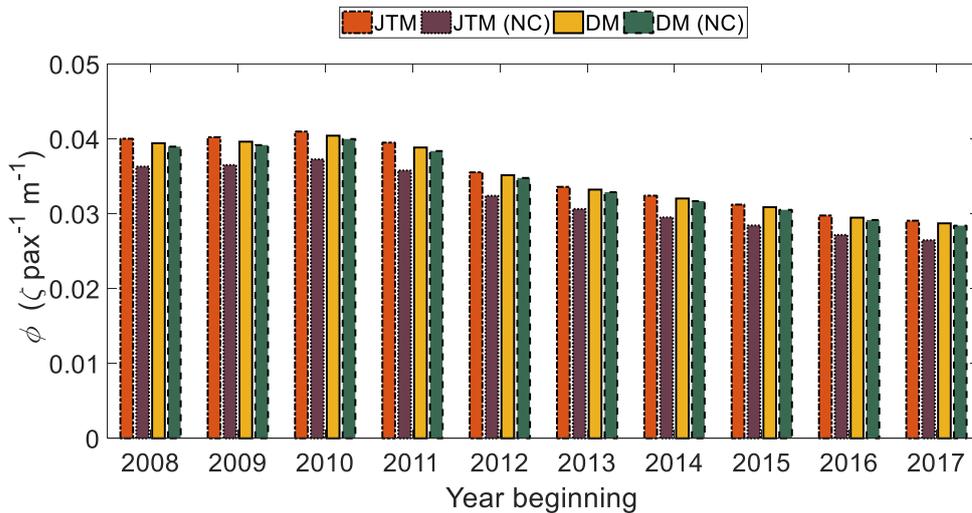


Figure 8 – Bar chart to compare the predicted network scores, ϕ , for different years and different prediction methods. Predictions using the Journey Time Metric, ϕ^{JTM} , and the Disutility Metric, ϕ^{DM} , are compared against the case where no crowding penalty is applied, $\phi^{JTM(NC)}$ and $\phi^{DM(NC)}$ respectively.

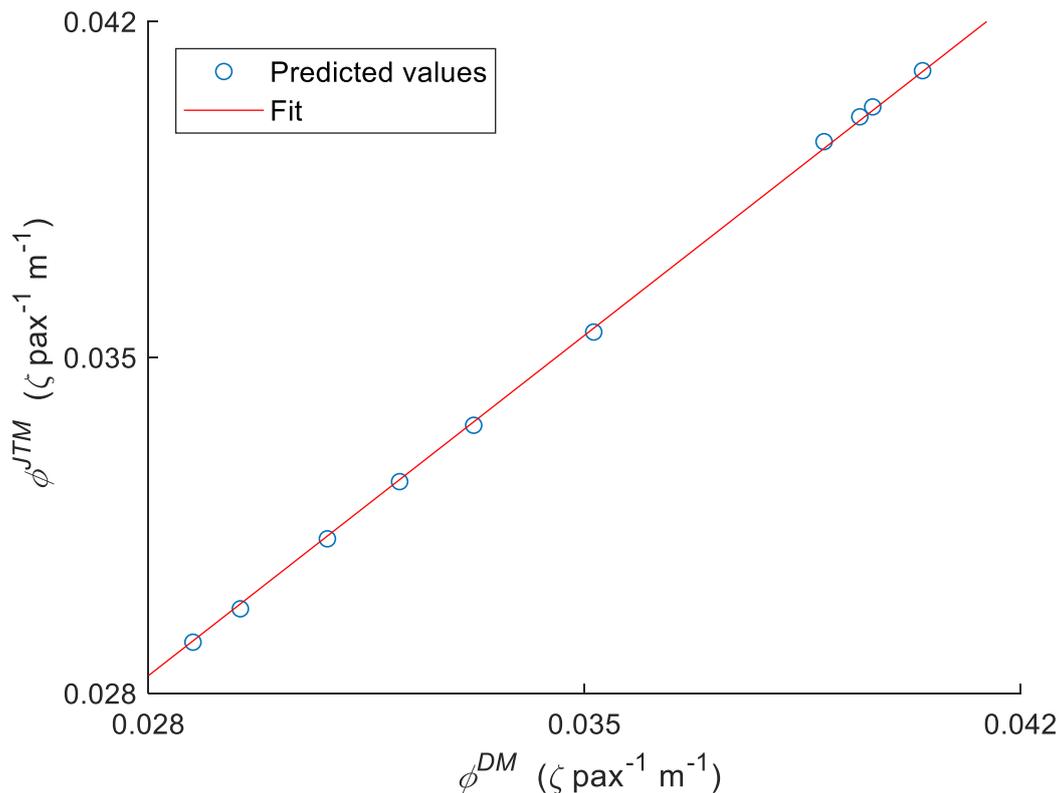


Figure 9- The relationship between the ten network score predictions for the Victoria Line from 2008 to 2017. The fit has an intercept forced to the origin and a gradient of 1.013.

3.3 Discussion

The results in Figure 6 indicate that, to successfully predict behaviour of ϕ^{CSS} , it is necessary to normalise the network assessment metric by the number of passengers and the distance they travel. In this investigation, the ratio between different line section loadings remains constant for all years therefore the value of R plotted in Figure 6 represents changes to passenger numbers and distance travelled. Consequently, the results in Figure 6 show that without passenger numbers and distance normalisation, the predicted network scores become sensitive to both. This effect is unwanted therefore including passenger number and distance normalisation within the network assessment metric is supported.

Choosing a typical significance level of 0.005, the results shown in Figure 7 are statistically significant evidence that the null hypothesis (that predicted and measured data are uncorrelated) can be rejected. Although the choice of

significance level is arbitrary (Wasserstein and Lazar, 2016), considering the JTM and DM have been developed from empirical studies of passenger preferences and there is evidence that end-to-end journey time influences passenger experience (Transport Focus, 2016), the alternate hypothesis is accepted, i.e. there is correlation between CSS data and predictions with the developed assessment metric when using the JTM, DM or end-to-end journey time. Because τ_B^{JTM} and τ_B^{DM} are closer to -1 than τ_B^{EE} , these results suggest that using the developed network performance metric with the JTM or DM better predicts relative changes to the CSS data than using end-to-end journey time. However, observing tables calculated by Walker (2016) indicate that even the 80% confidence intervals of τ_B^{JTM} , τ_B^{DM} and τ_B^{EE} are too large to determine a statistically significant difference between the values of τ_B^{JTM} , τ_B^{DM} and τ_B^{EE} . To determine a statistically significant difference by reducing the confidence interval without altering the significance level, more years of data for comparison are needed in the series of ϕ . It is unsurprising that τ_B^{JTM} and τ_B^{DM} do not equal -1.0 because, in this study, ϕ^{JTM} and ϕ^{DM} do not capture the effect of some factors, beyond the timetable and passenger load, which may affect ϕ^{CSS} , e.g. delayed trains. The network assessment metric using the JTM or DM can capture the effect of some of these other factors, but the limitation of data available to this study means that they are not well captured by the model of network operation used. Similarly, because of factors such as survey design and implementation, the CSS data may not fully capture influencers to passenger experience that distinguish ϕ^{JTM} , ϕ^{DM} and ϕ^{EE} , e.g. if the surveys were not conducted during times of high travel demand the effect of crowding will not be well captured. Consequently, not being able to determine a statistically significant difference in the accuracy of ϕ^{JTM} , ϕ^{DM} and ϕ^{EE} might also be a limitation of the measured CSS data.

Section 3.1.1 describes that for low passenger numbers, ϕ^{JTM} is insensitive to crowding (because no crowding penalty is applied), whereas ϕ^{DM} is. However when some passengers are standing (the normal operating regime for many GB services, e.g. 70% of services into London St. Pancras during the morning

peak (Peluffo, 2018)), ϕ^{JTM} will be more sensitive to crowding than ϕ^{DM} because it applies a crowding penalty four to eight times greater. This is confirmed by the results of Figure 8 which demonstrate that the contribution of the crowding penalty to the network score is on average 9% and 1% for the ϕ^{JTM} and ϕ^{DM} respectively. Section 3.1.1 also describes that the DM applies a greater VoT weighting than the JTM to passengers who are 'In Station'. Because the VoT weightings of the JTM and DM have been derived from surveying passengers, this may reflect local differences in passenger expectations where the metric was developed. For example, when used in the network assessment metric the JTM (developed in London) penalises crowding more and delay on the platform, less, than the DM (developed in Japan). This suggests that when considering a specific network, it is important to ensure the use of VoT weightings relevant to the passengers of that network. However, the similarity of the ϕ^{JTM} and ϕ^{DM} values in the results indicate that the difference in weightings placed on different passenger journey states approximately cancel out (for the study network in the years investigated). The results in Figure 9 show a high degree of correlation ($\rho=1.00$, $P<0.005$) between network scores calculated using the JTM and network scores calculated using the DM, despite their different formulations and countries of origin.

Considering all the results together suggests that the newly developed network performance metric, using either the JTM or DM, can predict network performance from the passenger perspective, and successfully aggregates across passenger states to capture effects such as crowding and different journey stages. There is evidence that the network assessment metric, using either the JTM or DM, better predicts changes to customer satisfaction than end-to-end journey time. Because the JTM, CSS data and network operation data are all related to LUL, this result might be considered special to this case where there is a 'closed-loop' between metric and validation. However, the DM has no connection to the LUL data but is demonstrated here to achieve similar outcomes. This indicates the result is not particular to the 'closed-loop' case.

3.4 Summary

Passenger journeys are multi-stage and the conditions of a journey stage, e.g. crowding when on a train, can vary. The term 'state' has been introduced to describe a specific combination of stage and conditions. A passenger journey can be described as a series of states and the literature discussed in Chapter 2 has shown that the relative time spent in each of these will have differing effects on the overall experience of the passenger. Measuring the passenger end-to-end journey time alone, or the train punctuality at final destination (as used in the GB Public Performance Measure) will not capture this. The JTM and DM are journey assessment metrics that can capture individual journey experience by applying a VoT weighting to time spent in each state. Both metrics sum the weighted time spent in each state, but they use different weightings, journey stages and the DM applies an additional penalty for train changes. Both apply a crowding penalty to capture the additional disutility caused to a passenger when traveling on a train with other passengers. For networks operating in the regime where some passengers cannot find a seat, the crowding penalty applied by the JTM is four to eight times greater than the DM. In this regime, the assessment of network performance using the JTM is more sensitive to crowding than when using the DM. Both the JTM and the DM can be used as part of the network assessment metric introduced here, where the network score is taken to be the aggregate of journey scores normalised by the distance travelled and the number of passengers. It is found that, for the Victoria Line of the LUL network from 2008 to 2017, there is a high degree of correlation ($\rho=1.00$, $P<0.005$) between the network scores calculated with the JTM and network scores calculated with the DM, despite their different formulations and countries of origin.

When comparing network scores against measured values of customer satisfaction for the same network (obtained from surveys) there is statistically significant evidence ($P<0.005$) to reject the null hypothesis that predicted and measured changes do not correlate. Considering other evidence from the literature, this thesis adopts the hypothesis that predicted and measured

changes are correlated which means the network assessment metric can be applied to predict the relative performance of different networks from the passenger perspective. For the data available, the network assessment metric using the JTM or the DM better predicted relative changes to customer satisfaction than end-to-end journey time. However, to determine a statistically significant difference more data for comparison is required.

In this thesis, from here on in, all network scores are calculated with the network assessment metric developed in this chapter, using the JTM for assessing individual journeys, and are denoted by ϕ (for brevity the superscript 'JTM' is omitted). The information required for the JTM, i.e. the passenger journey stages and crowding penalty parameters, informs what must be captured in the passenger rail network model developed in the next chapter. Because of the evidence discussed in Chapter 2, it is considered that minimising a network score calculated using the metric developed in this chapter is equivalent to maximising the satisfaction of passengers.

Chapter 4

A passenger rail network model

The SUPREME framework optimises a rail network model for the maximum satisfaction of virtual passengers travelling within it. This chapter discusses the rail network model developed and used for this purpose. As pointed out in Chapter 2, a model is necessary because of the impracticalities associated with testing, in the real world, the many network variations typically required by optimisation procedures. A requisite for this model is to capture the information required by the network performance metric developed in Chapter 3. Namely it must be able to: model individual passenger journeys as a series of states, capture the time a passenger spends in each state and capture the conditions relating to passenger crowding. Furthermore, Chapter 2 describes that the model must be adaptable to different network challenges and must be able to capture passenger-passenger, passenger-component and component-component interactions. Considering these requirements, Chapter 2 identifies that no pre-existing suitable models are known and available to the author. Therefore the Passenger Rail Model (PRaM), described in this chapter, has been developed by the author for use within SUPREME to model networks and their effect upon on passenger journeys. This model is discussed in the paper ‘Maximising passenger satisfaction through optimised train movements’ presented at the 2017 Stephenson Conference (Hickish et al., 2017). Given that accuracy and computational cost are important features of PRaM for its applicability in SUPREME, experiments to investigate these are described in this chapter.

4.1 The Passenger Rail Model

To meet the requirements listed in the previous section, PRaM uses Agent-Based Modelling (ABM) techniques described by Macal and North (2010) and Bonabeau (2002) to simulate each passenger and train individually but concurrently in a shared environment. Each train or passenger is represented

by a virtual *agent* with structure illustrated by Figure 10, adapted from Macal and North (2010).

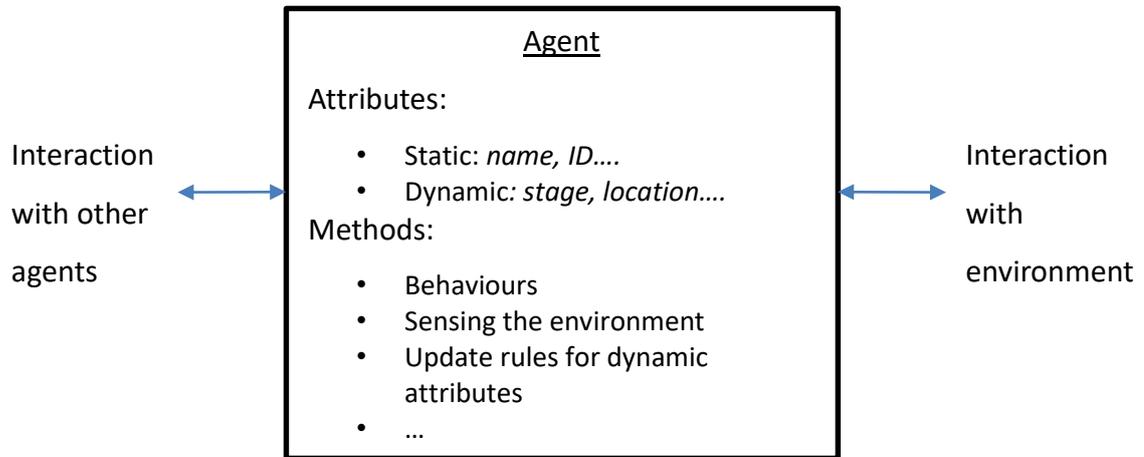


Figure 10 - An illustration of the structure of the agents used in PRaM. This is similar to the 'typical agent structure' described by Macal and North (2010).

Figure 10 shows that, as a virtual entity an agent has *attributes*, which maybe *static* or *dynamic*, as well as *agent-methods*. Some static attributes are unique identifiers, e.g. a name and an ID, whereas some describe a parameter modelling the real world, e.g. the number of seats in a train agent or the origin station of a passenger agent. Dynamic attributes are updated during the simulation, for example the number of passengers on board a train agent or the current location of a passenger agent. The *agent-stage* is a dynamic attribute which classifies the activity the agent is engaged in, e.g. 'At Platform' for train agents or 'On Train' for passenger agents. The agent-methods allow agents to function in the dynamic simulation without human input, i.e. be *autonomous*. Macal and North (2010) state that the *behaviours* of an agent 'relate information sensed by the agent [both internal and external] to its decisions and actions'. Furthermore, the behaviours of the agents capture their *goals*, e.g. for train agents to adhere to their timetable and for passengers to reach their destination, and their constraints, e.g. signalling constraints. Other agent-methods allow the agents to 'sense' the state of the environment or other agents as well updating their dynamic variables. The shared environment of the simulation is defined by the parameters of the network model representing the infrastructure, e.g. the stations and lines. The agents interact with each other and the environment according to their behaviours. Within an

agent type (e.g. passenger or train), the methods are identical, however the attributes may vary to represent different real-world instances of train or passengers. PRaM is a continuous simulation that models a user-defined period of operation, e.g. a whole day. Agents are sequentially updated, with a one second time step, based on the attributes of themselves, other agents and the environment at the start of the time step. The next sections describe the infrastructure environment, train agents and passenger agents in more detail.

4.2 Infrastructure environment

The simulation occurs in a two-dimensional environment with two infrastructure entities: stations and lines. Stations are represented as point-like with co-ordinates determined by real-world track distance separation. Multiple passenger and train agents can occupy the station location concurrently but stations have a defined number of platforms which limits the number of train agents. Lines connect the stations and are straight, one-dimensional, can have variable maximum speed limit along their length, and are either single or bi-directional. One line represents a pair of rails.

The signalling system within PRaM is a *moving block* system where the exact location and speed of all trains is known and used to determine for, each train, the furthest it can safely travel from its current location, i.e. the Limit of Movement Authority (LMA). This signalling system is chosen because the regulation of the European Traffic Management System (ERTMS) by the European Commission (2019), means that many European states (including the UK (Furness et al., 2017)) are developing towards moving block signalling, i.e. ERTMS Level 3. Appendix II includes a further discussion of this choice.

Generally, moving block signalling systems require constant communication of the LMA to the train, e.g. through wireless technology. Therefore, lineside signalling infrastructure is not included in PRaM because, when moving block signalling is used, the signalling infrastructure topography does not affect the movement of trains. PRaM captures the LMA communicated to a real-world train in the train agent behaviour (discussed in the next section).

4.3 Train agents

Train agents are single homogenous units which travel around the network and are used by passenger agents to make their journey. The maximum capacity, number of seats and current occupancy of a train agent are described by parameters stored as attributes. The goal of train agents is to adhere to the timetable which specifies the stations they must visit, via which lines, as well as the target arrival times and earliest departure times at each station. However, the behaviour of train agents ensure they avoid collision, obey safety rules (e.g. speed restrictions and dispatching) and operate within the limits of their performance, which means they can become delayed. There are two train agent stages, 'At Platform' and 'On Line', which are used to indicate when passenger agents may board and alight, and when train agents travel around the network. The movement behaviour of train agents in the 'On Line' stage is described in the next section. A train agent will change from the 'On Line' to 'At Platform' stage when it is motionless at the station it is calling at. When in the 'At Platform' stage, train agent behaviour is determined by the Dispatcher Model (described in Section 4.3.2) which governs when the train agent changes to the 'On Line' stage.

4.3.1 The movement behaviour of train agents

Whilst in the 'On Line' stage, the movement behaviour of train agents is determined by three models: the Signaller Model (SM), the Train Control Model (TCM) and the Train Dynamics Model (TDM). The SM determines the LMA for the train agent depending on its location and agent-stage as well as the location of other train agents and stations. Using the approach of Tao and Ke-Ping (2007), the SM applies standard kinematic equations to determine the train agent's braking distance at its current speed. The location of the LMA is calculated by considering the last location at which the brakes can be applied to avoid a collision, plus a 'safety margin'. If there is no Movement Authority for a train agent, its LMA is equal to its current position. The TCM determines the applied *traction control*, i.e. the control of traction or braking systems governing the magnitude of train acceleration or deceleration. The TDM

determines the position and velocity of the train agent when a traction control is applied.

Four types of traction control are used to capture the movement behaviour of trains: acceleration, cruising (speed holding), coasting and braking. Figure 11 is a schematic representation showing how two hypothetical train journeys can be described using this model. The vertical lines indicate the locations where the traction control phase changes, i.e. *control points*. In the left window the permissible line speed, i.e. the speed limit of the line section, is constant for the whole journey. The train agent accelerates to the permissible line speed, cruises at this speed, then coasts and brakes. In the right window, the permissible line speed is reduced for a midway section of the journey. For this journey, there are 7 control points separating a sequence of traction control phases where the acceleration, cruising and coasting phases are repeated. By combining and repeating traction controls it is possible to capture the distance-speed profiles of real world trains, as is shown by the results of Powell and Palacin (2015).

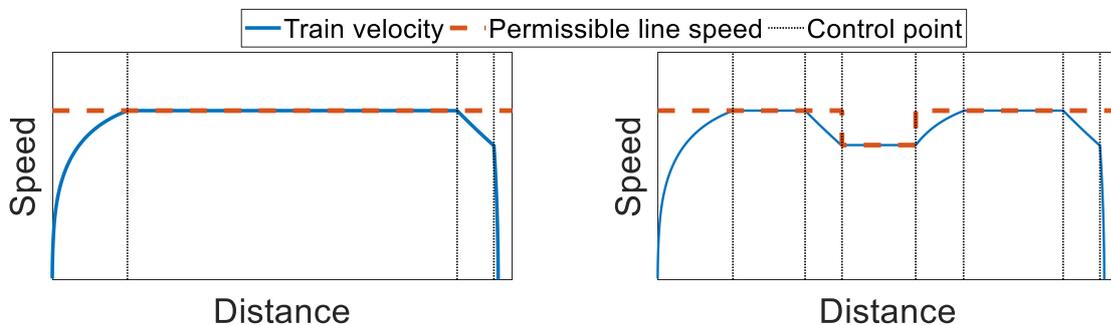


Figure 11 – Schematic distance-speed relationships and traction control phases for two hypothetical train journeys. The solid line represents train speed, the dashed line represents the permissible line speed and the vertical dotted lines represent the location of control points between traction control phases.

The TCM dynamically determines control points and the traction control. In the case of acceleration or cruising, the traction control is a percentage of the tractive effort available to the train. In the case of braking, the traction control describes a constant rate of deceleration relating to: ‘service brake’, ‘full brake’ and ‘emergency brake’.

To determine appropriate traction control, the TCM requires information from the TDM and SM. Figure 12 illustrates the flow of information between the SM, TCM and TDM as well as the input and output information. The TCM proposes a traction control and the TDM computes the resultant train agent position, velocity and stopping distance using the 'service brake'. This information is then compared against the LMA determined by the SM and the permissible line speed. If the proposed traction control does not meet the safety constraints it is rejected and a reduced traction control is proposed (less tractive force from the propulsion system or greater braking). This process iterates until an acceptable traction control is found. So that the movement behaviour of train agents captures the 'smooth' driving requirements of real world trains (i.e. no rapid changes between acceleration and braking), the TCM only proposes an increased tractive force after braking if: the LMA has been altered or the permissible line speed has been changed, or the train agent's velocity is less than 75% of the maximum that would meet the safety constraints.

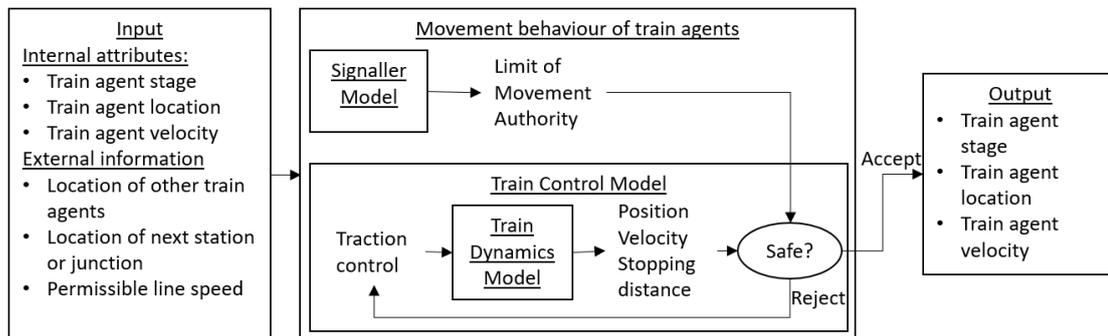


Figure 12 – An Illustration of the flow of information in the movement behaviour of PRaM train agents.

Because braking deceleration is modelled as constant, the TDM calculates the time-dependant position and velocity of the train agent using standard kinematic equations. However, when a tractive force is applied, the TDM uses Newton's second law of motion with the effective mass of the train agent and the total force acting upon it. Howlett and Pudney (1995) show that the mass of the train can be considered point-like for this calculation. The effective mass of the train agent, M_E , is given by the formula:

$$M_E = M_T(1 + r) + M_P \quad (4.1)$$

where M_T denotes the train stationary tare mass, M_p , the mass of the passengers and r is an empirically derived constant to capture the effect of rotating parts, i.e. the *rotary allowance*. The values of M_T and r are specific to the train being modelled and are provided by data. The total force is calculated by resolving the tractive force at the rail-wheel contact, the resistance force and the force due to gravity. The applied tractive force is calculated by multiplying the force available to the train, and the traction control percentage output by the TCM. The available tractive force is speed dependent and is calculated from the formula, given by (4.2), which models data relating to the speed-dependent power output of a real world train engine.

$$F_{available} = \begin{cases} F_{max}, v \leq v_0 \\ \frac{F_{max}v_0}{|v|}, v > v_0 \end{cases} \quad (4.2)$$

The symbol $F_{available}$ denotes the tractive force available to the train, F_{max} denotes the maximum tractive force available to the train when at rest, v denotes the train speed and v_0 denotes the boundary speed between constant tractive effort and constant tractive power. Similar to the approach of Bešinović et al. (2013), the formula of (4.2) uses a simple model of constant tractive power at speeds above v_0 , but could be replaced by another relationship trivially. The resistance force, $F_{resistance}$, is modelled using the Davis formula:

$$F_{resistance} = \mathcal{A} + \mathcal{B}v + \mathcal{C}v^2 \quad (4.3)$$

where \mathcal{A} , \mathcal{B} and \mathcal{C} are train-specific constants respectively relating to the static, rolling and aerodynamic resistance of the train with values determined from analysing real-world train motion, e.g. by Hansen et al. (2017). The force due to gravity, $F_{gravity}$, is calculated from the formula:

$$F_{gravity} = -g \cos(\theta) \quad (4.4)$$

where g denotes acceleration due to gravity and θ the angle between the rail and the horizontal with a positive value as uphill and a negative value as downhill. Because the total force acting upon the train is nonlinear, the

equations for position and velocity are approximated using the Euler method. The parameter values of M_T , r , \mathcal{A} , \mathcal{B} and \mathcal{C} are stored as static attributes of train agent. The value of M_P is a dynamic attribute recalculated everytime the train agent departs a station using a single passenger mass of 77kg (Matheson, 2010). The value of θ is determined by cross referencing the position and direction of travel of the train agent against a database of line gradients for the network.

4.3.2 The Dispatcher Model

On the GB network, for a real-world train to be dispatched: the train must have Movement Authority, all passengers must be clear of the doors and the doors must be closed (Rail Safety and Standards Board, 2017). To capture this, the Dispatcher Model returns a Boolean statement on whether a train agent can be 'dispatched' depending on if the train agent has Movement Authority from the SM and if the inequalities given by (4.5) and (4.6) are satisfied.

$$t \geq t_D \tag{4.5}$$

$$t \geq t_A + t_{DW} \tag{4.6}$$

The symbol t denotes the current simulation clock time, t_{DW} denotes the dwell time of the train agent, and t_D and t_A respectively denote the timetabled departure and arrival time of the train agent. The value of t_{DW} is determined for each train agent at every station stop and models the minimum amount of time required for the real-world passengers to board and alight a real-world train including door opening and closing times. To calculate t_{DW} , the formula developed by Weston (cited by Harris and Anderson (2007)) is used:

$$t_{DW} = C_0 + \left[1.4 \left(1 + \left(\frac{U}{35} \right) \left(\frac{W - Z}{Q} \right) \right) \right] \times \left[\left(U \times \frac{E}{Q} \right)^{c_1} + \left(U \times \frac{L}{Q} \right)^{c_2} + 0.027 \left(U \times \frac{E}{Q} \right) \left(U \times \frac{L}{Q} \right) \right] \tag{4.7}$$

$$U = \frac{\text{busiest door passenger flow}}{\text{mean door passenger flow}} \tag{4.8}$$

where, C_0 , denotes the total time in seconds for the train doors to open and close, W , the number of passengers remaining on the train, Z , the number of seats, Q , the number of doors, E , the number of passengers boarding, L , the number of passengers alighting, C_1 and C_2 , weighting constants and U the door factor. The purpose of U is to capture the heterogeneous distribution of passengers along the train and platform, with a value calculated from the formula given by (4.8). However, because PRaM does not model passenger locations on the train or platform, its value is set to 1. Weston proposed values of 15, 0.7 and 0.7 for C_0 , C_1 and C_2 respectively. These values are network-specific, but Harris and Anderson (2007) found that the value of C_1 and C_2 only needed to vary across a range of 0.45 to 0.9 and 0.8 to 0.9, for the model to have validity for 26 urban rail networks worldwide. Observing (4.7) it can be seen that there are terms to capture: the number of boarding and alighting passengers, their interaction with each other (at the door) and their interaction with other passengers in the train. Models such as that presented by Li et al. (2016) capture additional factors such as the dimensions of the door, train-platform gap and layout of the station and may be valid for a wider range of network types than Weston's. However, because data relating to these has not been available to the author for all networks investigated in this thesis, Weston's model is used.

4.4 Passenger agents

Passenger agents are defined with an origin and destination, and their route through the network is predetermined as the shortest path between these. The goal of passenger agents is to travel their route in the minimum time. Passenger agents can only move through the network using the train agents and so take the location of these or stations. Their journey is described by the sequence of agent stages that relate to the journey stages described in Chapter 3: 'Not In System', 'Buying Ticket', 'Moving Through Station', 'On Platform', 'On Platform (Left Behind)' and 'On Train'. The behaviours of passenger agents relate to changing their stage. Figure 13 illustrates the sequence of passenger agent stages for a journey and the behaviours which

connect them. The agent stages are shown in boxes and the behaviours as named arrows (the code 'PB' relates to 'Passenger Behaviour').

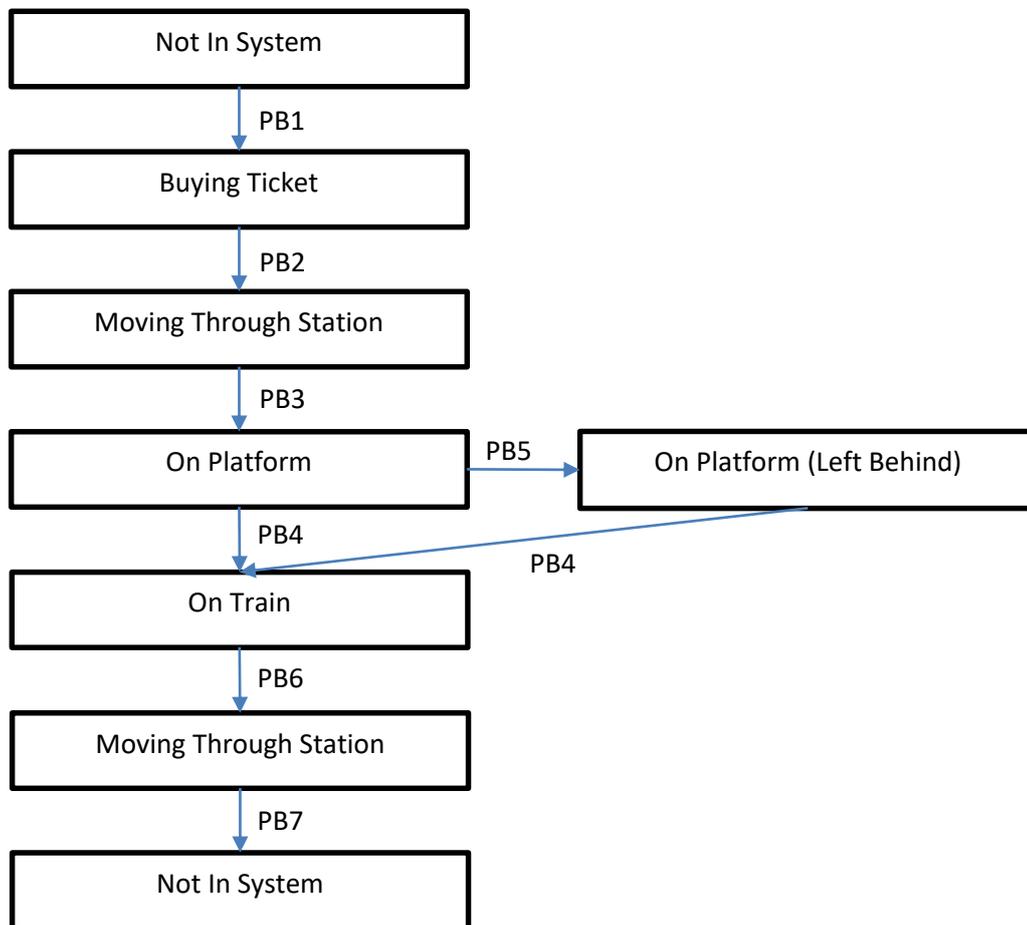


Figure 13 – Illustration of the possible passenger agent stages and the behaviours which connect them in the order of a journey. The stages are shown inside boxes. The behaviours are the named arrows connecting them. Passenger stages can only change in the direction shown by the arrow.

In Figure 13 it can be seen that the initial stage of passenger agents is 'Not In System' which means they have no location. PB1 occurs at a predetermined time for each passenger and they become in the 'Buying Ticket' stage with the location of their origin station. The passenger agent remains in the 'Buying Ticket' stage for τ_1 seconds before PB2 occurs and are then in the 'Moving Through Station' stage for τ_2 seconds before PB3 occurs. When in the 'On Platform' stage, PB5 occurs to move them onto the first train agent to call at their station whose next stop is on their route and has space. If a train agent with a suitable next stop is full and a passenger agent cannot board, then PB4

occurs and the passenger remains at the station but in the 'On Platform (Left Behind)' stage. The criteria are the same for a passenger agent to move from the 'On Platform' or 'On Platform (Left Behind)' stage to the 'On Train' stage, so PB5 connects both pairs. Passenger agents board a suitable train in order of priority given to the length of time they have been at the platform. A passenger agent will remain on the train agent, in the 'On Train' stage, until the train agent stops at a station and its next station stop is not on the passenger agent's route. In this case, PB6 occurs and the passenger agent alights the train, takes on the location of the station and is in the 'Moving Through Station' stage. If this station is the destination station of the passenger agent, it remains in this agent stage for τ_3 seconds before PB7 occurs and the passenger agent moves to the Not In System stage. If the station is not the passenger agent's destination, the passenger agent is in the 'Moving Through Station' stage for τ_2 seconds before PB3 occurs and the passenger goes back to the 'On Platform' stage to wait for another train agent. This loop continues until the passenger agent reaches their destination. The values of τ_1 , τ_2 and τ_3 are network specific but are constant for all passenger agents and are supplied as parameters to the PRaM simulation. The specific routes and departure times of all the passenger agents define the passenger load on the network.

To calculate the network score, Chapter 3 describes that passenger journeys are captured as a series of states. The state of a passenger agent changes each time their journey stage or the crowding conditions they are experiencing change. The crowding conditions that a passenger agent is experiencing are determined using information stored in the attributes of the associated train agent, e.g. maximum capacity, number of seats and number of riders. The VoT weighting function for each state a passenger agent experiences is calculated using (3.1) and (3.2). The simulated duration of the passenger state and the VoT weighting function is stored for each passenger agent so that journey scores can be calculated. The distance travelled by each passenger agent is determined from their route and the network topography.

4.5 Validation of train movement

To validate the TCM and TDM, measured train movement data was compared against model predictions of train velocity and displacement. To validate the TCM, the location of control points for the predicted motion of the train were compared against the measured location of changes in train acceleration for a real-world train. The TDM must be validated independently of effects from the TCM, therefore the TDM was validated by comparing the predicted and measured time-distance and distance-speed relationships during a single phase only. Because the TCM is dependent on the TDM, the TDM was validated first.

The measured data was collected by Fletcher (2015)⁴ using a GPS device on board the 17:57 InterCity 125 High Speed Train service as it travelled from Stevenage to Grantham on the East Coast Mainline on 10/06/2015. The GPS device recorded the position, altitude and speed of the train at approximately one second intervals. The TCM and TDM predictions were calculated using the parameter values corresponding to an InterCity 125, shown by Table 4, with line gradients calculated from the altitude data collected by Fletcher.

Parameter	Value
Effective mass, M_E	489 tonnes
Traction system phase boundary speed, v_0	55 km/hour
Maximum tractive force, F_{max}	170 kN
Davis formula \mathcal{A}	3.22 kN
Davis formula \mathcal{B}	113 Ns/m
Davis formula \mathcal{C}	7.8 Ns ² /m ²

Table 4 – Parameter values used to describe an InterCity 125 in the validation experiment of the Train Dynamics Model.

The effective mass of the train shown in Table 4, has been calculated using (4.1) with a tare mass (M_T) of 393 tonnes (provided by personal communication with Nicholson (2016)⁴), a rotary allowance (r) of 0.08 (Steimel, cited by Zhao (2013)), 630 seats in the train (Angel Trains, c2019b) and assuming they are all

⁴ For more information, see the list of personal communications in the reference list.

occupied (as is common for trains departing London between 17:00 and 18:00 (Peluffo, 2018)). The other values in Table 4 have also been provided by Nicholson. To calculate the line gradients, first the altitude data was smoothed by taking the altitude at a location to be the mean of all data points in an interval 300m either side. The line gradients are then calculated by considering the change in altitude between each location. No data could be retrieved for the deceleration of an InterCity 125 at different braking levels therefore values, provided by Gill (2016)⁵ in a personal communication, for the 'Desiro' trains were used: 0.85 ms^{-2} , 1.0 ms^{-2} and 1.1 ms^{-2} respectively for service, full and emergency brake.

To isolate it from the effects of the TCM, validating the TDM requires comparing measurements and predictions of train movement for the same traction control. Although there is no measured data for the traction control applied for the train, Powell and Palacin (2015) state that generally during the acceleration phase, a traction control is continuously applied that is close to 100%. Consequently, Figure 14 compares the predicted and measured distance-speed relationship of the train for the initial acceleration phase, which was assumed to be from stationary at Stevenage until the first reduction in the train's speed (in this case, the first 12km of the journey). Measured values are shown by markers and predicted values are shown by lines, all with values relating to the left ordinate. Although the GPS device reports an accuracy of to within 10m for the data collected, it applies processing to the data which obscures the uncertainty in speed readings. Consequently, no error bars are plotted for the measured speed data, however, it can be seen to be 'smooth' indicating a small (approximated to be within $\pm 5 \text{ km/hour}$) uncertainty relative to the speeds involved. For clarity, no error bars are plotted for the distance data because in the scale of the x-axis a 10m uncertainty is indistinguishable. Predictions when a 100% and 80% traction control have been continuously applied are shown. Observing Figure 14 shows the TDM correctly predicts the

⁵ For more information, see the list of personal communications in the reference list.

train to accelerate to near top-speed with decreasing acceleration. The predicted top speed reached in 12km is dependent on the traction control applied. Corresponding changes in the gradient of the altitude data and distance-speed speed relationship indicate that the model is capturing the effect of line gradients, e.g. at approximately 4 km of the journey the downhill line gradient increases, causing the model to predict an acceleration which is also present in the measured data. It can be inferred that the distance-speed behaviour of the train is well predicted by the TDM with a traction control between 100% and 80%.

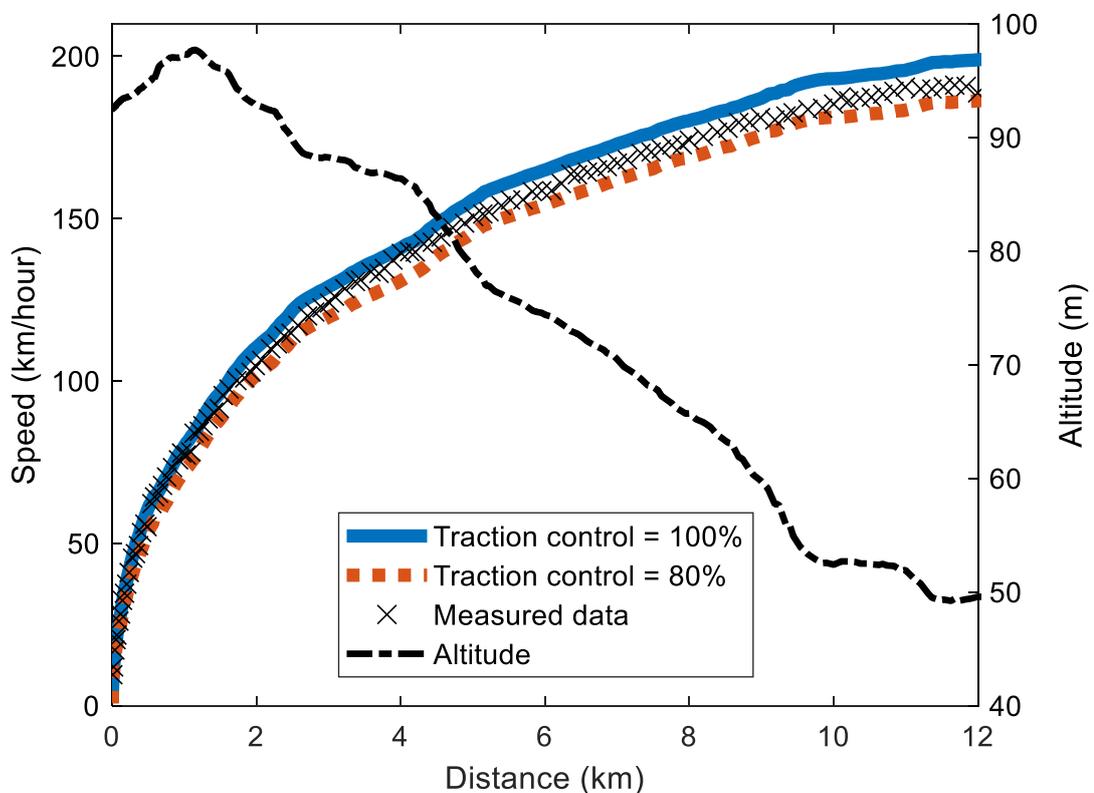


Figure 14 – Measured and predicted values of an InterCity 125's speed. The x-axis displays the distance from Stevenage. The left ordinate displays train speed for measured values and predicted values when different traction controls have been applied. The right ordinate shows the smoothed altitude at that point in the train journey.

Figure 15 compares the measured and predicted time-distance relationships for the initial acceleration phase. With a traction control of 100%, the TDM very closely predicts the location of the train at all times so the markers and line are almost indistinguishable. To travel 12km, the train is measured to take 344

seconds. The model shows that with 100% and 80% traction control, 342 seconds and 372 seconds are predicted respectively.

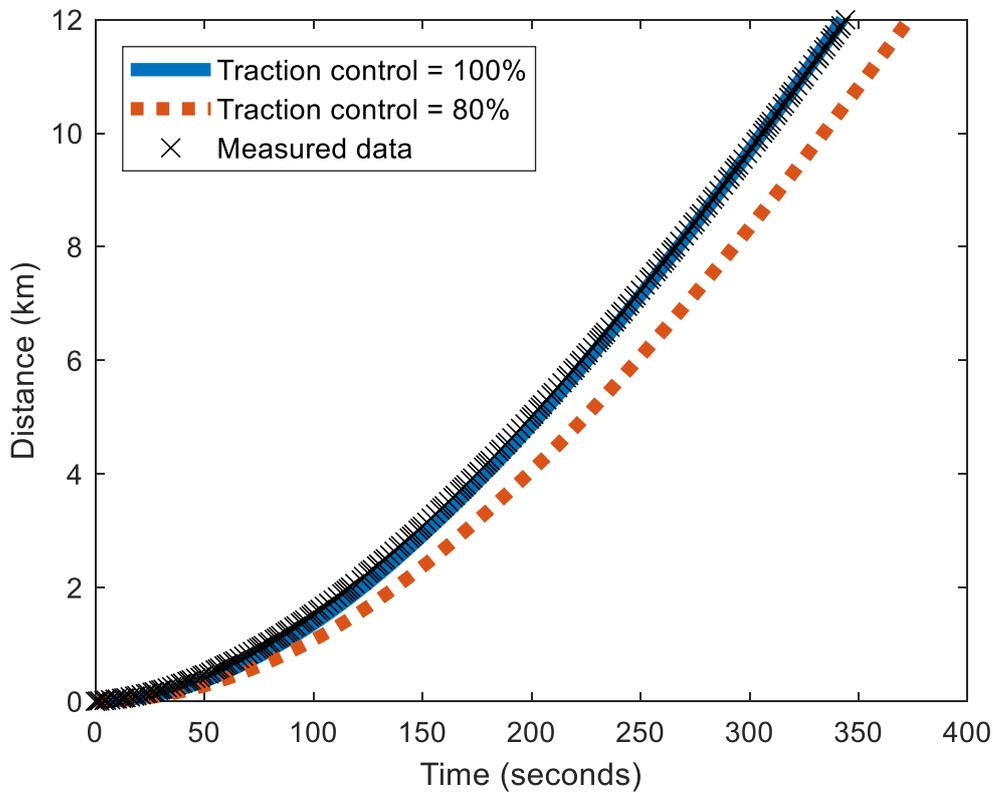


Figure 15 – Measured and predicted values of an InterCity 125’s distance-time relationship during the initial acceleration phase of its journey. Predicted values have been calculated with the TDM when a traction control of either 100% or 80% have been applied continuously.

Assuming that the TDM is valid, the TCM is validated by comparing the measured and predicted motion of the train for the whole journey when the traction control input to the TDM is variable and determined by the TCM. Figure 16 plots the measured and predicted speed of the train for the whole 125km journey. Observing the measured data, it can be seen that the real-world train does not exceed 195km/hour although train capability is 201km/hour and the Sectional Appendix published by Network Rail (2013, 2015) either side of the journey indicates that the permissible line speed for the journey was also 201km/hour. Consequently to maintain a like-for-like comparison, the predictions were calculated using a permissible line speed of 195km/hour. Vertical markers have been plotted at predicted control points.

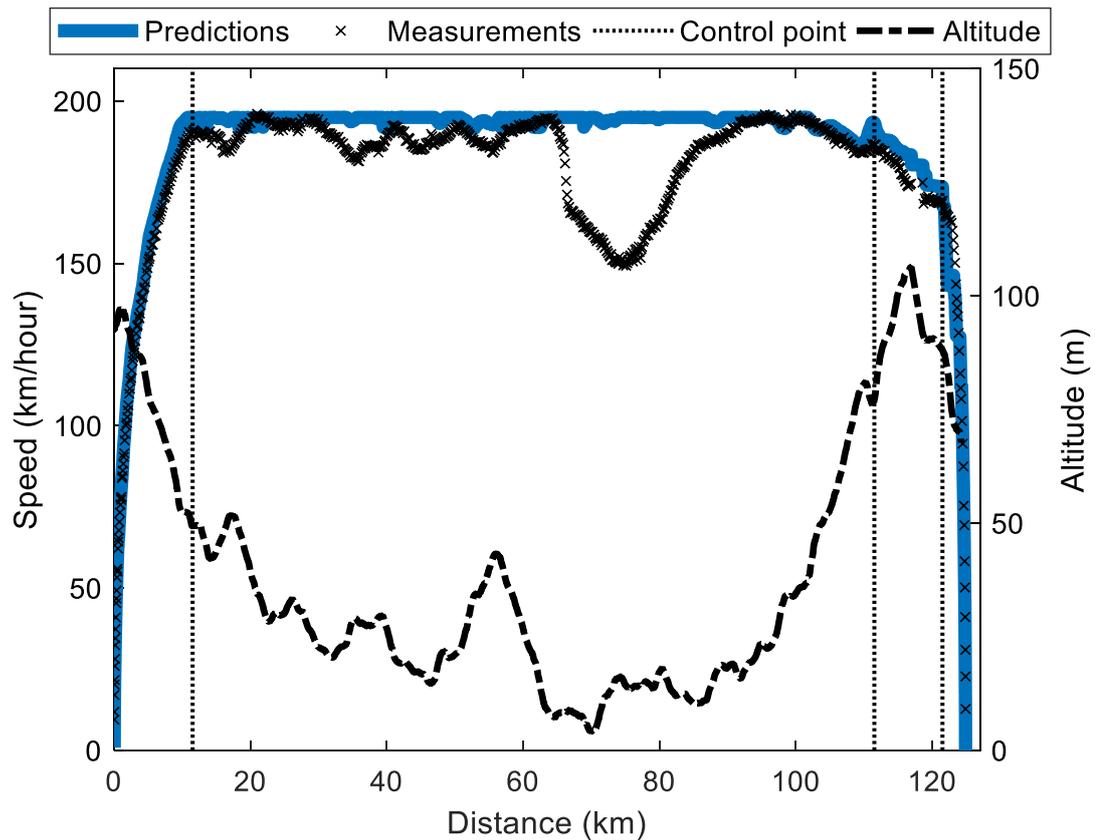


Figure 16 – The measured distance-speed relationship of an InterCity 125 travelling from Stevenage to Grantham on the East Coast Mainline, and predictions of the same journey made using the Passenger Rail Model (PRaM). The measured altitude of the train is also shown.

The results of Figure 16 show that during the initial acceleration phase there is good agreement between measurements and predictions. The control point from acceleration to cruising is predicted to occur at 11km and, by observing the first measured data point where the real-world train speed acceleration became less than or equal to 0, is measured to occur at 12km. During the cruising phase of the predicted train journey the model correctly predicts the train to maintain a speed of up to the permissible line speed, however, there are substantial variations between predicted and measured values.

Furthermore, the measured data indicates that the real-world train had an additional braking and acceleration phase with control points at respectively 64km and 75km. This might be because of an event that can be captured by PRaM, but is not captured in the validation simulation (e.g. a temporary speed restriction or congestion in the network). There are also small changes in the train speed during the cruising phase of the predicted journey, but these are

not caused by additional control points. By comparison with the altitude data it can be seen that the small changes in predicted speed during the cruising phase correlate with steep gradients. Consequently, the small changes in speed during the cruising phase of the predicted journey are attributed to an increased effect from the force due to gravity at parts of the journey with a steeper gradient. The TCM predicts the coasting phase control point to occur at 112km which corresponds in a measured negative acceleration beginning at 111km. The braking phase is predicted to begin at 124km which corresponds with an increase in the measured train deceleration at 123km. The mean percentage error between measured and predicted speeds is 6% with a standard deviation of 9%. The standard deviation indicates a positive skew on the distribution which might be caused by the model overestimating train speed in the cruising phase where not all control points have been predicted. The whole journey is predicted to require 2499 seconds, but is measured to take 2610 seconds.

4.5.1 Sensitivity analysis of Train Dynamics Model parameters

To investigate the sensitivity of the TDM to the input parameters, each was varied using a factor multiplier, ζ , and changes to the time predicted for the train to travel the first 12km of the journey were recorded. Each parameter was varied independently whilst others maintained their original value. Figure 17 plots the value of ζ on the x-axis which was varied to reflect up to an order of magnitude error in TDM parameter values. The percentage change in time, relative to the case when all TDM input parameters have the value shown in Table 4, is plotted on the y-axis. All series have the same value when $\zeta = 1$ because this the case where no input parameters are affected. Linear fits have been plotted for the first four series. For the last two series of data (F_{max} and M_E) fits of the form $y = dx^h + c$ have been used because in the case of constant acceleration the relationship can be shown analytically to be: $y = (M_E - \sqrt{M_E})\zeta^{1/2}$ and $y = (F_{max} - \sqrt{F_{max}})\zeta^{1/2}$ where F_{max} and M_E have their original values (see Appendix II). Table 5 shows the value of the fit parameters

and their upper and lower confidence bounds when a 95% confidence level is used.

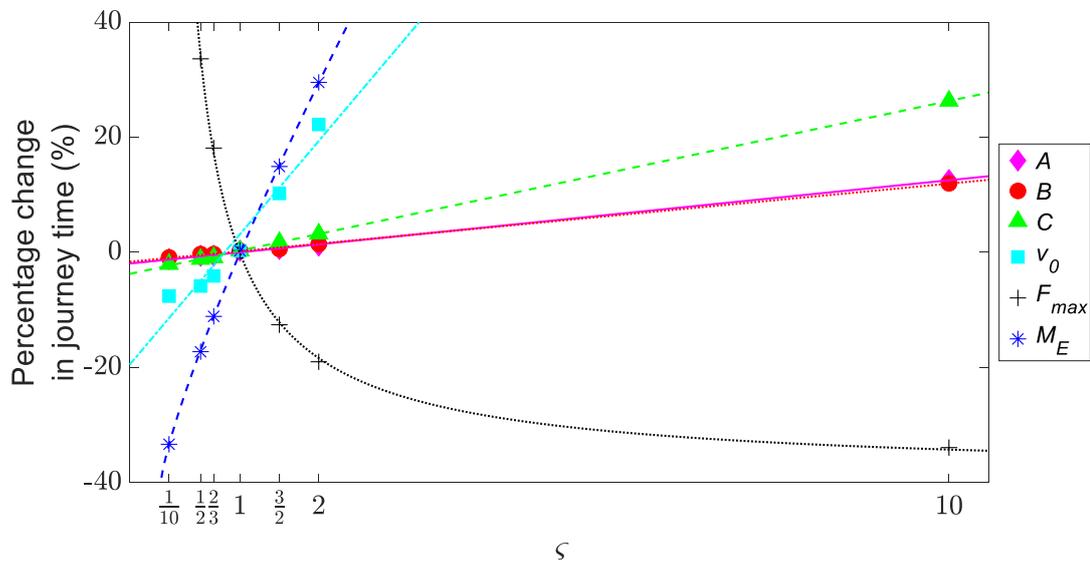


Figure 17 – The percentage change in time, to travel the first 12 km of the journey, calculated by TDM when different input parameters are multiplied by, ς , shown on the x-axis. The percentage change in journey time relative to when all parameters have the value shown in Table 5 is shown on the y-axis.

	<i>m</i>	LCB	UCB	<i>c</i>	LCB	UCB			
<i>A</i>	1.39	1.33	1.45	-1.38	-1.61	-1.15			
<i>B</i>	1.31	1.26	1.35	-1.10	-1.23	-0.91			
<i>C</i>	2.89	2.84	2.93	-2.58	-2.76	-2.40			
<i>v</i>₀	16.9	10.8	21.6	-13.0	-19.2	-6.8			
	<i>d</i>	LCB	UCB	<i>h</i>	LCB	UCB	<i>c</i>	LCB	UCB
<i>F</i>_{max}	38.9	33.9	43.8	-0.91	-1.04	-0.79	-39.0	-43.1	-34.9
<i>M</i>_E	39.4	37.3	41.6	0.81	0.76	0.86	-39.6	-41.4	-37.7

Table 5 - Parameters of the fits plotted in Figure 17. The series the fit relates to is shown in the first column. Fits of the form $y = mx + c$ and $y = dx^h + c$ are used. The Lower Confidence Bound (LCB) and Upper Confidence Bound (UCB) are shown for each fit parameter when a 95% confidence level is used.

Observing Figure 17 shows that the TDM is much more sensitive to changes in v_0 , F_{max} and M_E than to the Davis formula coefficients A , B and C . Relative to other parameters, changes to A and B , have very similar effects, with the markers and fits almost being indistinguishable. The results show that an error of 50% in a single input parameter will result in an error in time that is less than 20%.

4.5.2 Sensitivity analysis of time step used

To investigate the sensitivity of the TDM to the time step used, a simulation of the train applying 100% traction control from stationary for 340 seconds was computed using different time steps. The distance-time profiles were then considered relative to a baseline case. For each profile, the distance travelled at 20 second intervals was compared with the corresponding baseline distance. The differences were expressed as an absolute percentage of the baseline distance and the mean calculated for each profile. Because the baseline has been computed with a small time step (10^{-4} seconds), it is taken to represent the 'true' value of the simulated train movement and therefore Figure 18 plots the mean absolute percentage error for three different time steps. Reducing the time step from 10 seconds to 1 second, and from 1 seconds to 0.1 seconds, reduces the error by a factor of 26 and 3 respectively. These results indicate that at a time step of 1 second or less, errors in the TDM introduced by discretised time become unsubstantial. Furthermore, there become diminishing reductions in this error at smaller time steps. These results are discussed further in the context of the computational cost in Section 0.

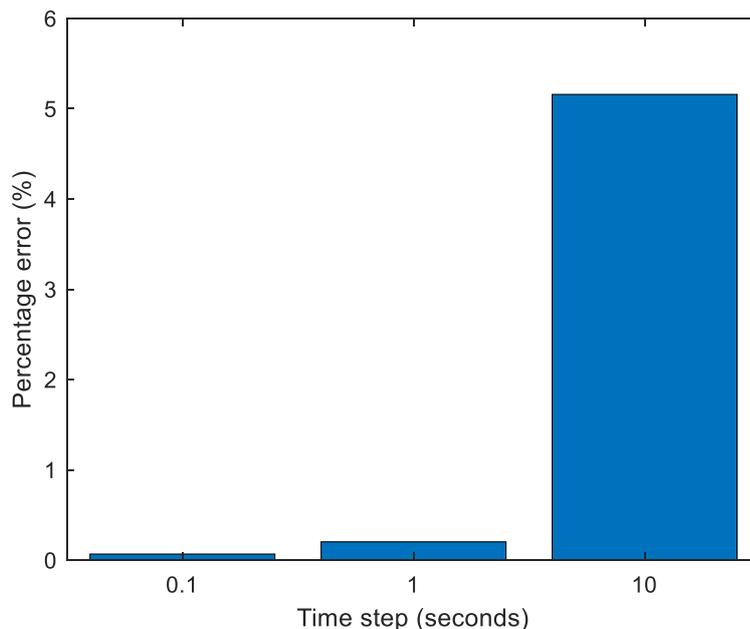


Figure 18 – The percentage error of the TDM when three different time steps are used.

4.5.3 Discussion of results

For modelling the time required by a train to travel a distance under constant traction control, the results of Figure 15 show that the TDM is accurate to within 0.6% of the measured time. This result has been found through considering an accelerating train only, but by symmetry it is reasonable to expect it to hold for decelerating trains also. Although this result is based on comparison with one train journey only, the similarity between the approach of other researchers, e.g. Sumpavakup and Kulworawanichpong (2015), and the TDM gives further confidence that the TDM is accurate enough for use in PRaM where larger errors can be introduced by the TCM.

The results of Figure 16 show that the TCM controls train speed to, on average, within 6% of the measured value and that the control points predicted by the TCM are within 1km of corresponding measured control points. For the range of speeds involved, it takes 18 to 23 seconds to travel 1km and given that human train control behaviour is varied, e.g. Dorrian et al. (2007) have shown that train driver performance is affected by fatigue, it is possible that this difference in control points is within the bounds of realistic train control behaviour. To determine the bounds of realistic driving behaviour, it is necessary to compare the speed profiles of human driven trains for the same journey (e.g. the results of Mott MacDonald (2006, p.55)).

To quantify the error expected from the TDM for predicting times for a whole journey, the results from Figure 15 and Figure 16 can be combined. The TCM predicts this train journey to have non-zero acceleration for 35% of the total measured journey time. An error of 0.6% during this time (because of inaccuracies in the TDM) causes an error of 0.2% in the total journey time. However, because acceleration and deceleration also occur in the cruising phase, this is likely to be an underestimate. The upper bound on the error in journey time will be 0.6% in the case that the train always has non-zero acceleration. Therefore, an error of 0.2% to 0.6% is expected for predicting journey times, however, this would increase if more than one acceleration,

coasting or braking phase occurs. The error would also increase further in the case that TCM does not predict all traction control phases.

The accuracy of journey time predictions is dependent on the parameter values used in the TDM. For example, the results of Figure 17 show journey times predicted with the TDM vary by more than 30%, if input parameters relating to the effective mass of the train or the traction system are not accurate to within an order of magnitude. However, for all input parameters, an error of 50% in an individual input parameter will result in a variation in predicted journey time that is at most 20%. Obtaining parameter values to this level of accuracy is realistic (see Appendix II) and a 20% variation in journey time is acceptable for the purposes of PRaM within the SUPREME framework. Because the results of Figure 17 show that the accuracy of the TDM is relatively insensitive to changes in the Davis formula coefficients, and the results of Hansen et al. (2017) indicate that the value of \mathcal{C} , representing aerodynamic resistance, changes by less than an order of magnitude for trains in a tunnel (where aerodynamic resistance rises), this means that the TDM can still be applied for networks involving these. Only the effect of an error in an individual TDM parameters has been investigated, but in some combinations, errors in multiple parameters will interact, e.g. effective mass and maximum tractive force.

4.6 Validation of passenger agent behaviour

Having validated train movement in the previous section, this section discusses the validation of passenger agent behaviours. In Great Britain, there has been a historical shift in general service provision from high capacity trains to smaller, more frequent trains. Whilst discussing the use, in this shift, of DMUs to replace locomotive-hauled trains, Ford (1986) describes the rail industry's belief that increased service frequency increases the number of passengers. Consequently, to validate passenger behaviour, different networks reflecting this shift in service pattern are simulated and the changes to indicators of attractiveness to passengers are compared against the statements of Ford.

Two networks are simulated with the same topography and train performance, and carrying the same passenger load, but with different timetables and passenger capacity of the trains. Figure 19 illustrates the topography of the two networks simulated with a circle representing a station, the letter the station name, and a connecting edge representing a line. The outer stations are all connected to the central station by two single direction lines with opposite directions, i.e. an 'up' and 'down' line. All lines are the same length and are flat. At each station, there is one platform serving each line. There are 12 different combinations of origin and destination stations, i.e. routes, in this network and the passenger load is distributed homogenously between all of them. The journeys of 1.7×10^4 passengers over a 14 hour day are simulated. The arrival time of passenger agents at their origin station (the time at which PBI for that passenger agent occurs) is distributed homogenously from 09:00 to 21:00 and equivalently between the routes. The time to buy a ticket (τ_1), move through the station (τ_2) and exit the destination station (τ_3) was 120 seconds.

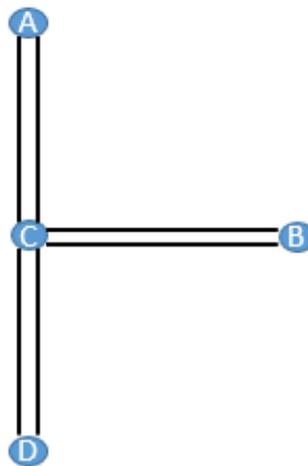


Figure 19 - The network topography used for the investigation. A circle represents a station with the letter as its name. A single edge represents a line so there are two lines connecting each outer station to the inner station.

For both networks, the train agents are timetabled to travel back-and-forth between a single outer station and the central station. The network operating a service of less frequent, larger trains is referred to as 'Network X' and the network operating a service of more frequent, smaller trains is referred to as 'Network Y'. Network X has one train travelling between each outer and central

station pair. Table 6 shows the first three hours of the timetable and final arrival time of all the trains in Network X. It can be seen that all the trains travel synchronously and that there is an hourly service between all stations, which continues until the trains arrive at the outer station at 23:00.

Train 1			Train 2			Train 3		
Station	Arrive	Depart	Station	Arrive	Depart	Station	Arrive	Depart
A	09:00	09:03	B	09:00	09:03	D	09:00	09:03
C	09:30	09:33	C	09:30	09:33	C	09:30	09:33
A	10:00	10:03	B	10:00	10:03	D	10:00	10:03
...
A	23:00		B	23:00		D	23:00	

Table 6 – The first three hours of the timetable for all the trains in Network X and half the trains in Network Y.

All the trains in Network X are simulated to be InterCity 125s with a passenger seating capacity of 630 (Angel Trains, c2019b) and a standing capacity of 63. The standing capacity has been determined as 10% of the seating capacity, because data published by the UK government (2003) shows that this is the case for the British Rail (BR) Class 442 which uses the same coach design as an InterCity 125. The formula given by (4.1) shows that the number of passengers in the train affects the effective mass of the train, but all other parameters values are the same as stated in Section 4.6.

In Network Y, an additional train agent is simulated operating between each outer and central station pair, making a total of six train agents. Three train agents have the same timetable as for Network X. The first three hours of the timetable and the final arrival time of the other three train agents is shown by Table 7 which shows that the additional train agents are timetable to travel synchronously with each other, but by comparison with Table 6, asynchronously with the other three train agents. For Network Y, there is a half-hourly service between all connected stations. All trains agents in Network Y have the performance parameters of an InterCity 125, but with half the seating capacity, standing capacity, tare mass, maximum tractive force and Davis formula resistance coefficients.

Train 4			Train 5			Train 6		
Station	Arrive	Depart	Station	Arrive	Depart	Station	Arrive	Depart
C	09:00	09:03	C	09:00	09:03	C	09:00	09:03
A	09:30	09:33	B	09:30	09:33	D	09:30	09:33
C	10:00	10:03	C	10:00	10:03	C	10:00	10:03
...
C	23:00		C	23:00		C	23:00	

Table 7 - The first three hours of the timetable for the additional trains in the smaller, more frequent trains network.

Both networks were simulated and the total passenger time in two different journey stages recorded. Figure 20 plots the total passenger time spent either on the train or in the station for both networks. The results show the total passenger time on a train is the same because the timetabled train journey times are the same for both networks. Since Network Y has double the frequency of train agents, passenger agents must wait less time before a suitable train arrives causing the time in station to be reduced by 53% compared to Network X. For both networks, all passengers successfully make their journeys.

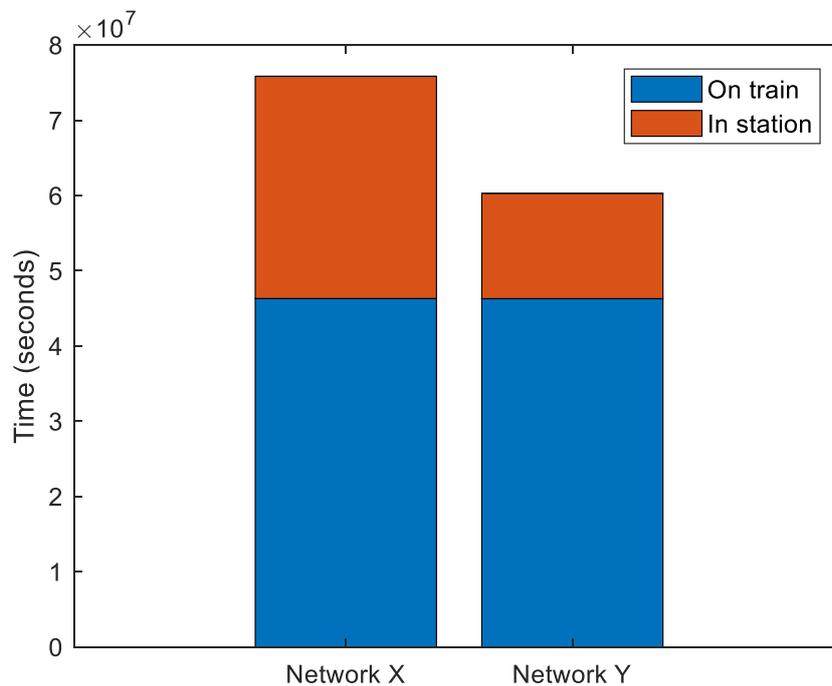


Figure 20 - The total passenger time spent either on a train or in a station for Network X and Network Y. For both columns, the lower portion represents time on the train and the upper portion represents time in the station. Network X has a service of less frequent, larger trains. Network Y has a service of more frequent, smaller trains.

To investigate the effect of high passenger demand, the number of passengers was doubled to 3.5×10^4 whilst the other parameters describing the networks remained constant. The passenger load remained evenly distributed between routes and homogenous in time. Figure 21 shows the total passenger time spent in different journey stages for both networks. The passenger time spent on a train remains equal between the two networks and the time in a station is less for Network Y than Network X. As a result of the increased passenger demand, not all passengers can fit on the first train and some passengers are left behind at the platform. The total amount of time left behind is equal for both networks.

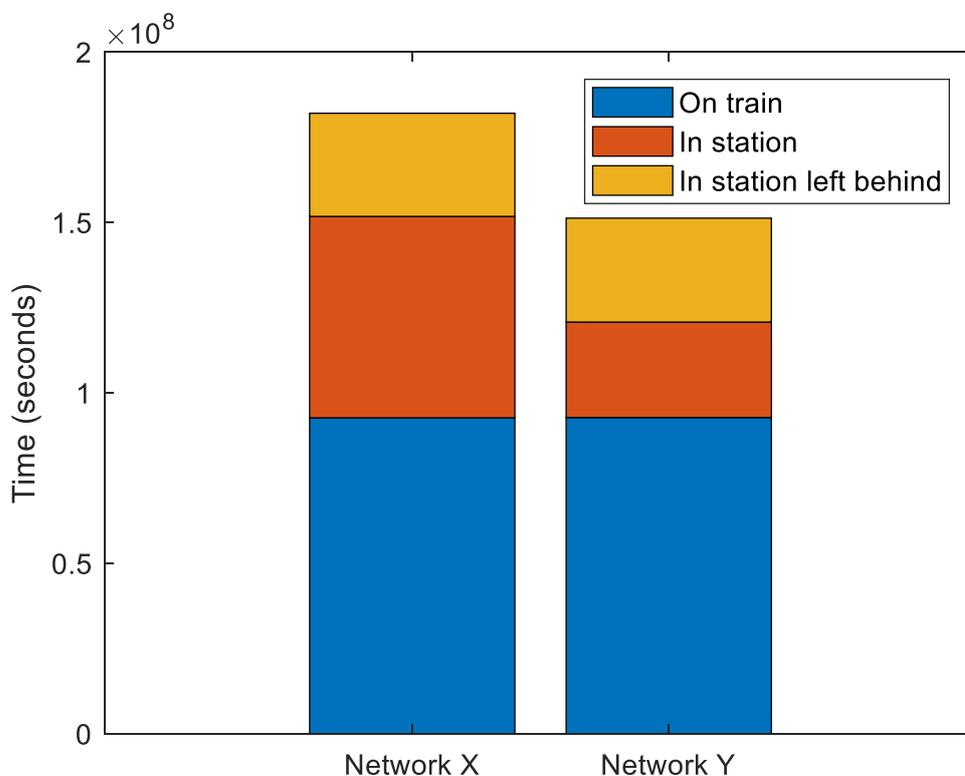


Figure 21 - The total passenger time spent either on a train, in a station or left behind in a station for Network X and Network Y. For both columns, the lower portion represents time on the train, the middle portion represents time in the station and the upper portion represents time in the station 'left behind'. Network X has a service of less frequent, larger trains. Network Y has a service of more frequent, smaller trains.

Owing to the fact that all passenger agents complete their journey, the results of Figure 20 indicate that passenger behaviours PB1, PB2, PB3, PB6 and PB7 allow a passenger agent to travel through the network. The results of Figure 20

also show that a more frequent service reduces the total journey time of all passenger agents, this indicates that behaviour PB5 only occurs when a suitable train agent is available. The results of Figure 21 show that behaviour PB4 captures the effect of passengers not being able to board an overly crowded train. However, this effect is sensitive to the standing capacity of the train agent which may not fully capture the real world because situation dependent factors might not be fully included in an assessment of standing capacity, e.g. a specific group of passengers' willingness to 'crush', or variations in luggage volume. Furthermore, PRaM models passenger distribution through the train as homogenous and assumes that passenger agents will locate and occupy any available seat, which is not always realistic. Both sets of results indicate that Network Y is more attractive to passengers than Network X because as would be intuitively expected, on average, passengers spend less time waiting for a train (with all other factors affecting service quality remaining constant). Network Y relates to smaller, more frequent trains and the results agree with the statements of Ford and the historical trend in service patterns.

The investigation in Section 4.6 has assumed the same passenger demand for both Network X and Network Y despite different service frequencies. This assumption has been made in order to isolate the effect of service frequency on passenger experience from secondary effects so that this primary relationship can be clearly validated. However, in reality passenger demand is sensitive to service frequency – as has been described by Balcombe et al. (2004). Increased service frequency provides a more convenient service for most passengers, and hence is likely to lead to an increase in demand for travel. If it were desirable to include the effect of service frequency upon travel demand, this could be done by calculating the elasticity of demand to service frequency, i.e. comparing the percentage change in demand divided by the percentage change in frequency for a series of observed values.

4.7 Testing the performance of the Passenger Rail Model

By recording the wall clock time required for different simulations, the computational cost of PRaM is investigated for its sensitivity to: the number of

train and passenger agents as well as the time steps used for both agent types. A virtual network is created using the same topology described in the experiments of Section 4.6. The number of train agents is incremented in threes and these are distributed evenly between the outer and central station pairs. Passenger agents are also distributed evenly between the possible routes and the arrival times to their origin station are homogenously distributed over the simulated day. Train agents have parameters relating to an InterCity 125 and a timetable with a one-hour period. The simulation terminates when 12 hours of network operation have been simulated. The simulations have been computed using MATLAB R2017b on an Intel Xeon Dual Processor @ 2.4 GHz. This is an unremarkable ‘server’ machine with performance likely similar to machines available to organisations for whom SUPREME is intended for, i.e. network managers.

Figure 22 plots the factor increase in computational cost and agent numbers compared to the case of 3 train agents and 1.7×10^4 passengers, which required 17 seconds of computation time. A time step of 1 second was used for both agents. Linear fits have been plotted for the data with parameters and their 95% Lower Confidence Bound (LCB) and Upper Confidence Bound (UCB) values shown in Table 8.

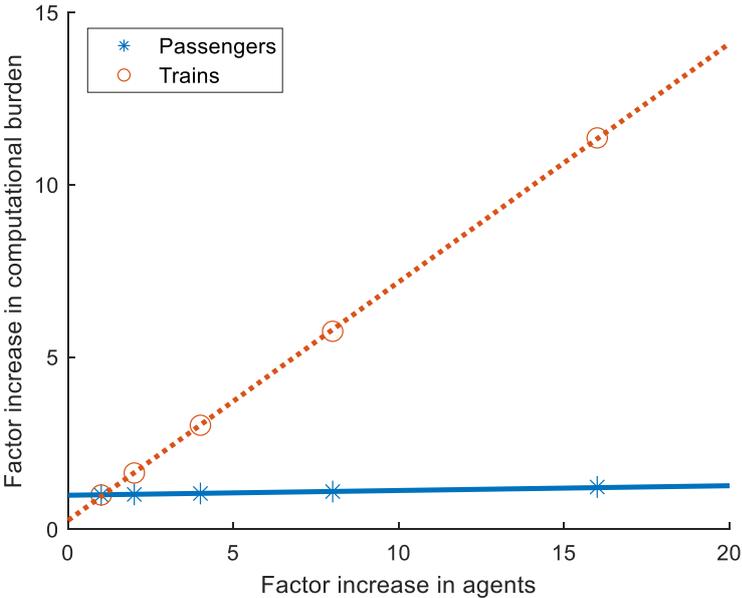


Figure 22 - The relationship between increases in agent numbers and computational cost

Series	Gradient	LCB	UCB	Intercept	LCB	UCB
'Passengers'	0.014	0.014	0.014	0.99	0.98	1.02
'Trains'	0.69	0.68	0.70	0.26	0.17	0.36

Table 8 – The fit parameters for the two series of data shown in Figure 22.

The strong linear fits and gradients of the results in Figure 22 indicate that the computational cost of PRaM increases proportional to increases in agent numbers with a constant of proportionality of 0.014 and 0.69 for passengers and trains respectively. The computational cost is more sensitive to increases in train agents because the code for train agent calculations runs serially, whereas the passenger agent code uses MATLAB's vector coding functionality. Vector coding can substantially speed up computation (Mathworks, 2019), therefore the passenger agent code may be computationally cheaper than the train agent code. This would mean that changes to computational cost of the agent code are unsubstantial compared to the train agent code, which is observed in the results. Since the intercepts of the fits do not equal zero, the results confirm that there is a computational overhead of the simulation regardless of the number of agents.

To investigate the effect of changing the time step to 60 seconds for passengers and 0.1 seconds for trains, Figure 23 plots the increase in computational cost compared to the same simulation but with a time step of 1 second for both trains and passenger. A linear fit has been plotted for the data the 'Passengers' series, the fit has a gradient of -3.5×10^{-3} (95% lower and upper confidence bounds of -4.1×10^{-3} and -2.9×10^{-3}) with intercept of 0.95 (95% lower and upper confidence bounds of 0.92 and 0.98). As the 'Trains' series does not support a linear fit but there is no reason to expect a different relationship, no fit has been plotted. The intercept of the 'Passenger' series indicates that changing the time step used by passenger agents from 1 second to 60 seconds reduces the computational cost by 5%, however this reduction becomes gradually greater at higher passenger numbers. The computational cost of train agents is more sensitive to changing the time step, but this reduces with greater numbers of train agents. For a factor increase of 16 times more train agents, the factor increase in computational cost is less than 1. Therefore these results imply that once a threshold of train agents is reached it

is faster to use a 0.1 second time step than 1 second, which is surprising. Consequently, this is an area for further investigation. In the context of the relationship of TDM accuracy and time step, discussed in Section 4.5.2, these results indicate that a 1 second time step offers the best compromise between computational cost and accuracy. Increasing the time step to 10 seconds would reduce computational cost but result in a 5% error in train movement calculations, which is substantial if modelling a congested network. Reducing time step to 0.1 seconds does bring a small increase in accuracy (<1%), but is outweighed by the substantial increase in computational cost.

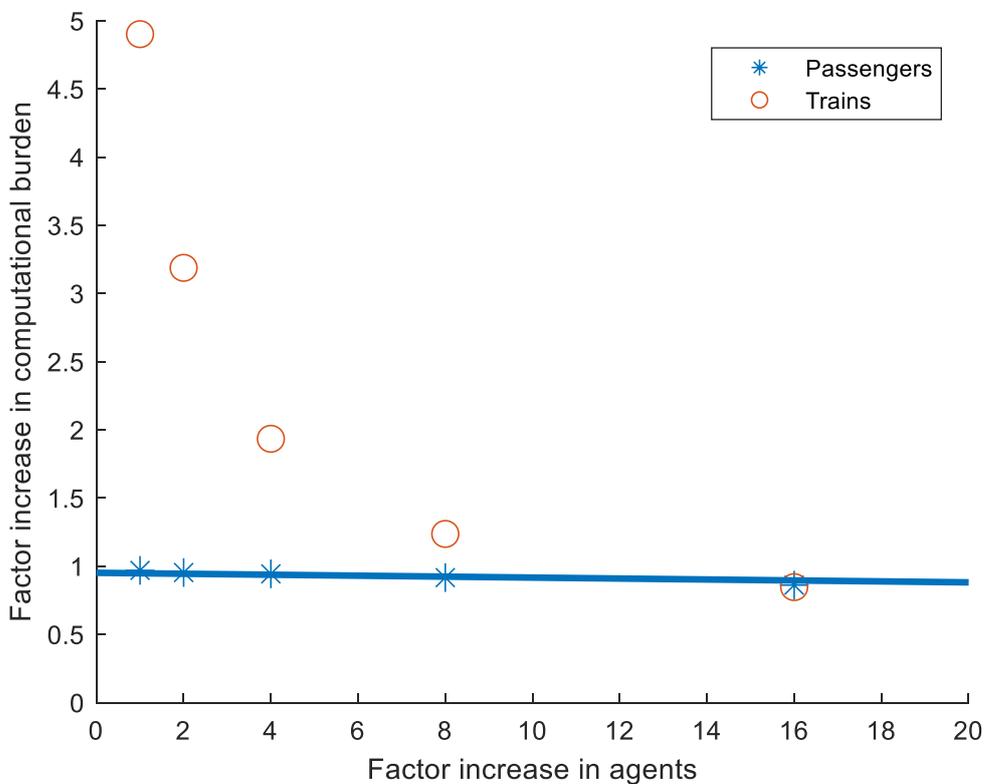


Figure 23 – The relationship between the increase in computational cost, relative to the value for the same number of agents but with a time step of 1 second for passengers and 1 seconds for trains, at different numbers of agents, calculated with a time step of 60 seconds and 0.1 seconds for passengers and trains respectively.

4.8 Summary

Within SUPREME, PRaM is used to model simultaneous passenger journeys so that they can be assessed using the metric developed in Chapter 3. In this way network performance can be determined dependent on the parameters

describing it. PRaM is an ABM where train and passenger agents are simulated in the network with their actions determined by 'behaviours'. Because of the likelihood of future implementation of moving block signalling within the GB network, PRaM only models the stations and track sections of a network infrastructure. Including junctions is identified as an area for further work which can be informed by existing models. Train agents have behaviours with the goal of adhering to their pre-defined timetable, whilst capturing the constraints of obeying safety restrictions. Experimental comparison shows that the journey times of trains can be modelled to within 0.6% when the traction control is known, and that traction control can be predicted to within 23 seconds. Further comparison between predicted and measured journeys is needed to determine the statistical significance of this result and whether this difference is within the range of variation in human driver control. The accuracy of train movement predictions is sensitive to the input parameters describing the performance of the train. It is shown that an error of 50% in a single input parameter will lead to a maximum of 20% error in train journey times, however the magnitude of the effect of error in multiple parameters has not been investigated. A time step of 1s is used to simulate trains and the ratio of proportionality between train agent numbers and computational cost is 0.7. Passenger agents are simulated individually with behaviours determined by the goal of reaching their destination station as quickly as possible. The journeys of passenger agents relate to the journey stages of the network assessment metric described in Chapter 3, therefore the time passenger agents spend in each state, and the conditions, can be determined from the simulation. The behaviours of passenger agents allow them to travel through the network but captures the constraints of passengers not being able to board trains that do not take them towards their destination or are overly crowded. The computational cost of PRaM scales with the number of passenger agents with a constant of proportionality of 0.02.

Considering all the errors discussed for the movement of train agents, PRaM can predict train journey times in a network with a maximum error on the order of 10%. This means that the time passengers spend on a train is modelled

with on the order of 10% maximum error, however, because of the connected nature of rail networks, this might have a non-linear effect on the error in total journey time. For, example the error in train journey time may be the difference between a passenger agent making a connection or not. The position of passengers is not modelled within stations or trains and as a consequence there is also error introduced in determining the time passengers spend moving through a station and their level of crowding on a train. Nonetheless, for this stage in the investigation of the SUPREME concept, this level of accuracy is considered acceptable. More accurate models of rail networks and pedestrian movement have already been developed and these can be used to inform the development of PRaM once the overall SUPREME concept has been evaluated.

Chapter 5

Selecting an optimisation algorithm

The SUPREME framework requires an optimisation algorithm to locate ‘good’ solutions for a ‘reasonable’ computational budget that places no restrictions on the task formulation. Chapter 2 described that Genetic Algorithms (GAs) have been popular for this purpose in existing rail network optimisation methodologies where objective function evaluations are cheap-to-compute. However, the requirement for cheap-to-compute objective function evaluations might limit the challenges to which SUPREME can be scaled to. Consequently, GAs might be poorly suited for use within SUPREME. Optimisation algorithms using the Bayesian Optimisation (BO) method have the potential to find similarly ‘good’ solutions to a GA, but in fewer model computations because they use a predictive model of the search space to target the selection of new candidates for evaluation. However, calculating the predictive model has a computational overhead that increases substantially with the number of optimisation variables. Therefore, whilst GAs might be well-suited to tasks with cheap-to-compute models, BO has the potential to be computationally cheaper than GAs for tasks with an expensive-to-compute model and few (many established implementations currently perform well for less than approximately 25) optimisation variables. This reduction in computational expense would allow SUPREME to be applied to larger scale challenges. Consequently, following a brief review of both methods in the next section, in this chapter the BO method is compared experimentally against the GA method as a benchmarking measure and its suitability for use within SUPREME discussed.

Some of the work presented in this chapter is also presented in the author’s paper: ‘Investigating Bayesian Optimization for rail network optimisation’ which has been accepted for publication in the International Journal of Rail Transportation. Prior to this, to the best of the author’s knowledge, no publicly available documents describe the application of BO for optimising the

performance of a rail network model, or a like-for-like comparison of BO and GA.

5.1 Genetic Algorithms and Bayesian Optimisation

A brief description of both GAs and BO is given in the following subsections. The general optimisation task described in Chapter 2 by (2.1), (2.2) and (2.3) is considered as the basis for an equitable comparison of BO and GAs. More comprehensive descriptions of GAs can be found in Goldberg (1989) and Mitchell (1996), and of BO in Shahriari et al. (2016). Note that within the GA literature, the objective function is usually referred to as a ‘fitness function’, however ‘objective function’ is used here for both cases for consistency.

5.1.1 Genetic Algorithms

A GA is a simple computational model of the process of natural selection in an evolving population. At every iteration, a GA evaluates the objective function for every candidate within a population. By selecting the candidates with the best objective function scores and ‘mating’ them, the population at later generations exhibits more characteristics of candidates with ‘good’ objective function scores and converges towards the optimum. ‘Mutation’ is used to allow a GA to ‘explore’ the search space. The GA method can be applied to many types of optimisation task with any number of optimisation variables and, because it requires many (typically $>10^4$) evaluations of $f(x)$, it is well suited to tasks where $f(x)$ is cheap-to-compute. Algorithm 1 presents pseudo-code for a simple GA. An important control parameter of the algorithm is the population size, P . The algorithm iterates depending on a conditional statement at line 4 that is often related to the objective function scores of the candidates found so-far, or, the computational resources used. The number of algorithm iterations used by a GA, θ^{GA} , is the final value of q in Algorithm 1. At every iteration all of the candidates in a generation are evaluated so the number of objective function evaluations, η^{GA} , required, is given by:

$$\eta^{GA} = \theta^{GA} \cdot P \tag{5.1}$$

Algorithm 1: Genetic-Algorithm(P)

1. Initialise population, G_1 , with P candidates
2. Evaluate objective function for every candidate in G_1
3. $q = 1$
4. **while** Is-Not-Terminated(q, G_q)
5. $G_{q+1} = \text{Evolve}(G_q)$ // select, mate and mutate candidates
6. Evaluate objective function for every candidate in G_{q+1}
7. $q = q + 1$
8. **end**
9. Return best candidate evaluated so far

5.1.2 Bayesian Optimisation

To estimate the global maximum of an objective function, a BO implementation creates an approximation of it, called a *proxy function* (also referred to as a *surrogate model* or *response surface*). In comparison to the objective function, the proxy function is cheaper to compute and is continuous so it is 'easier' to find its maximum. To create the proxy function, a probabilistic model is inferred from previous evaluations of the objective function's value at different locations in the search space. For this model, it is common (Snoek et al., 2012) to use a Gaussian Process (GP) regression model as is the case considered here. The proxy function and its uncertainty are respectively the mean and variance of the GP model – an overview of how these are calculated is given in the next section. At each iteration of a BO algorithm, the objective function is evaluated and the new data is used to update the probabilistic model and, hence, the proxy function. Information from the proxy function is used to create an *acquisition function*, whose global maximum indicates where in the search space the objective function should next be evaluated. The acquisition function is important to the success of BO because it 'guides' the search, but finding its maximum increases the computational expense of the whole process, particularly for tasks with more than approximately 25 dimensions (Kandasamy et al., 2015). Nonetheless, for tasks with less than approximately 25 dimensions, it is often cheaper than evaluating an expensive-to-compute objective function and consequently the BO method is often well suited to task

of this type. For higher dimension tasks, BO implementations are still being developed that keep the cost of maximising the acquisition function reasonable (Li et al., 2017).

Algorithm 2 presents pseudo-code outlining BO and is supplemented by Figure 24 which illustrates the early stages of a BO procedure. The x-axis represents the value of the optimisation variable, x , and the left ordinate represents both the value of the objective function, $f(x)$, its proxy function, $\mu(x)$, and the uncertainty about the proxy function, $\mu(x) \pm \sigma(x)$. The right ordinate represents the value of the acquisition function, $\alpha(x)$. The first stage is to evaluate $f(x)$ at the initial candidate, x_1 , whose location is shown by a square marker (line 2 of Algorithm 2). It should be noted that where the vector, x , is indexed this refers to its order in the list of all candidates evaluated during the optimisation, and is different to the indexing of scalar optimisation variables, e.g. x_i , introduced in Chapter 2 and also used later in this thesis. Stage 2 shows that this information is used to create an initial $\mu(x)$ that models what is known about $f(x)$ at this stage (line 4 of Algorithm 2). A corresponding $\sigma(x)$ is also calculated and combined with $\mu(x)$ to calculate the acquisition function. Stage 3 shows that $\alpha(x)$ increases further away from objective function evaluation because of the increased uncertainty on the value of $f(x)$. Stage 3 also shows the maximum of the acquisition function with a triangle marker. The value of this maximum is x_2 (line 7 of Algorithm 2) which is then used to evaluate $f(x)$ in Stage 4, shown by a square marker (line 9 of Algorithm 2). Stage 4 shows a circle marker to represent that data from previous samples remain in the probabilistic model and because there is now more data, $\mu(x)$ better approximates $f(x)$ than in Stage 2 with reduced $\sigma(x)$ in the region around the second observation. Stages 5 and 6 demonstrate that maximising $\alpha(x)$ and calculating $\mu(x)$ and $\sigma(x)$ are two important processes that are repeated in the ‘while loop’ of Algorithm 2. Furthermore, Algorithm 2 also demonstrates that the BO method only uses one function evaluation per iteration, therefore the number of expensive-to-compute evaluations, η^{BO} , is equal to the value of q at the end of the algorithm.

Algorithm 2: Bayesian-Optimisation()

1. Initialise candidate, x_1
2. $y_1 = f(x_1)$ // sample objective function
3. $D = [x_1, y_1]$ // data set of corresponding x and y values
4. Calculate proxy model, $\mu(x)$, and uncertainty, $\sigma(x)$, using D
5. $\varrho = 1$
6. **while** Is-Not-Terminated(ϱ, D)
7. Create acquisition function, $\alpha(x)$, using $\mu(x)$ and $\sigma(x)$
8. $x_{\varrho+1} = \arg \max_{x \in X} \alpha(x)$
9. $y_{\varrho+1} = f(x_{\varrho+1})$
10. $D = \{D, [x_{\varrho+1}, y_{\varrho+1}]\}$ // augment new data to data set
11. Calculate $\mu(x)$ and $\sigma(x)$ using D
12. $\varrho = \varrho + 1$
13. **end**
14. Return best candidate evaluated so far

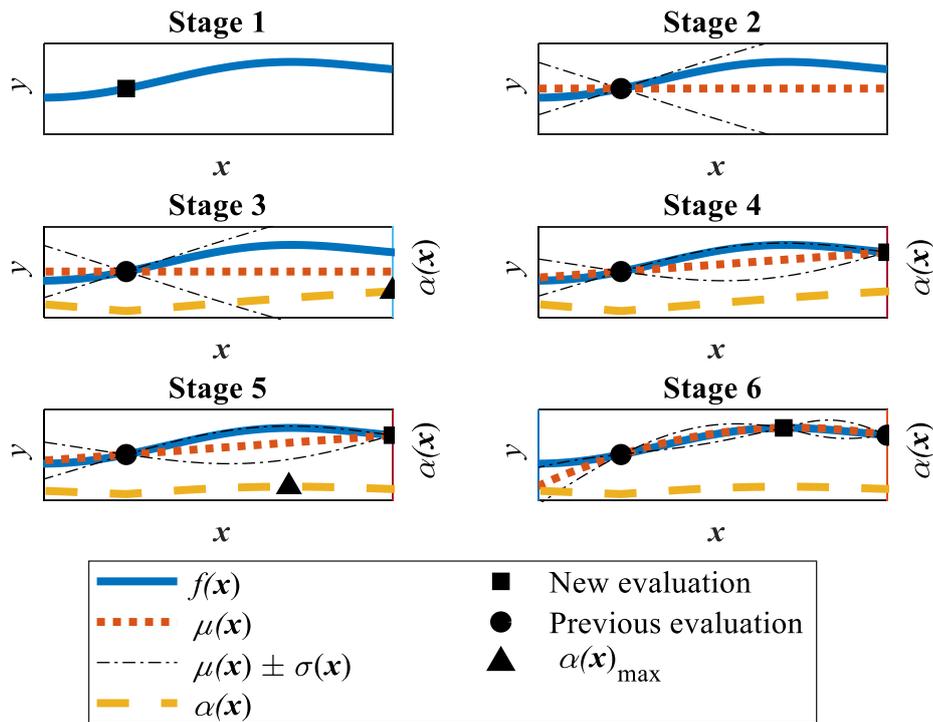


Figure 24 - Illustration of six of the early stages in the BO approach. The objective function, $f(x)$, is shown by a solid line, the proxy function, $\mu(x)$, is shown by a dotted line and the uncertainty about the proxy function, $\mu(x) \pm \sigma(x)$, is shown by a dash-dot line. The values of $f(x)$, $\mu(x)$ and $\mu(x) \pm \sigma(x)$ are plotted on the left ordinate. The acquisition function, $\alpha(x)$, is shown by a dashed line and its value plotted on the right ordinate. The x -axis represents the variable, x , being optimised.

5.1.3 Gaussian Process regression modelling

Here an overview on GP regression modelling is given, for further descriptions the reader is referred to Rasmussen and Williams (2006) and Seeger (2004). The following descriptions concentrate on a noiseless scenario because this thesis considers modelling the output of a deterministic simulation.

A GP is a collection of random variables, any finite number of which have joint Gaussian distribution, and is fully specified by a *mean function* and *covariance function*⁶. A GP can be used to describe a distribution over functions, in a function space defined by a covariance function. An unknown target function can be modelled as the mean function of a GP. A convenient outcome of this approach is that the variance of the function distribution provides explicit information about the uncertainty around the target function's value. GP regression modelling uses Bayesian inference to calculate the mean function of a GP that models the target function – hence estimating the target function. Including target function observation data in the GP model improves the estimation of the target function and reduces the variance of the GP model. In the context of Bayesian inference the GP model which includes previous assumptions or knowledge (such as a choice of covariance function or observation data) is referred to as the *prior*, and the GP model to be used for estimation is referred to as the *posterior*. In summary, GP regression modelling is a non-parametric approach to regression modelling which uses prior information when updating its model of the objective function, i.e. a Bayesian approach.

In the context of BO, the 'target function' described above is the 'objective function' referred to in Section 5.1.2 and throughout this thesis. The prior GP model is updated with additional data iteratively. The mean function and variance of the posterior GP model are used in the acquisition function to determine where the next sample of the objective function should occur. The

⁶ Sometimes referred to as a *kernel*.

information from both of these can be combined to reflect a user choice balancing exploitation and exploration.

The effectiveness of GP regression modelling is dependent on the choice of covariance function used – it being an assumption about the class of basis functions for the target function, and one of the first prior assumptions made. Covariance functions are required to be positive semi-definite functions, i.e. for all points in the search space the resultant covariance matrix should be Hermitian and its eigenvalues positive, and typically aim to reflect that the closer together two points are in the input space – the stronger they should be correlated. It is also possible to encode specific prior knowledge or assumptions about the target function into the mean function of the GP model, however it is common to apply a zero valued prior mean function when such knowledge does not exist – as is the case here. The specific covariance function, mean function and hyperparameter values used in these experiments are further described in Section 5.2.5.

5.1.4 Comparing the computational cost of Genetic Algorithms and Bayesian Optimisation

For this investigation the total computational cost of each method is decomposed into the cost of all the objective function evaluations, Γ , and the cost of computing the algorithm (excluding objective function evaluations), Π . Γ is the product of the number of evaluations, η , and the cost of a single evaluation, γ . Note that when variables apply to both methods, superscripts are used to denote variables specific to an method. Because of the multiplier, P , in (5.1), it is expected that $\eta^{GA} > \eta^{BO}$ and consequently that $\Gamma^{GA} > \Gamma^{BO}$. However, because BO involves the expensive step of maximising an acquisition function at every iteration, $\Pi^{GA} < \Pi^{BO}$ resulting in a trade-off between the low algorithm cost, high objective function evaluation cost of GAs and the high algorithm cost, low objective function evaluation cost of BO. The case of BO being cheaper than a GA is captured by (5.2) which demonstrates that the value of γ is important for determining the best method. Note that γ is constant between the approaches and is assumed to be constant for all x .

$$\Pi^{BO} + \gamma \cdot \eta^{BO} < \Pi^{GA} + \gamma \cdot \eta^{GA} \quad (5.2)$$

Because the number of objective function evaluations and the number of algorithm iterations are related, the value of Π is dependent on η . In the case of a GA this is a linear dependency. However in the case of BO with a GP model, calculating the model with ϱ data points requires inversion of a square matrix of dimension ϱ . The computational cost of this is $O(\varrho^3)$. Consequently, the value of Π^{BO} may become fourth order for large η^{BO} . However, developing methods to compute large matrix inversions at reduced cost is an active area of research (Li et al., 2011a, Ballard et al., 2016) so this is not seen as a fundamental limitation of BO. Furthermore, although not utilised with BO, Gardner et al. (2018) present a method using parallel computing techniques that can reduce the cost of computing a GP model to $O(\varrho^2)$.

5.2 Experimental comparison of Genetic Algorithms and Bayesian Optimisation

For an experimental comparison, specific GA and BO implementations were applied to a range of test-tasks involving an expensive-to-compute objective function that simulates passengers using a rail network and captures their satisfaction. For an unambiguous comparison, the globally optimal solutions (\mathbf{x}^*) must be known. For this reason, the test-tasks have been chosen so that \mathbf{x}^* can be calculated analytically.

5.2.1 The test-tasks

Two examples of demands in rail network operation are:

- The allocation of finite rolling stock between scheduled train paths (Abbink et al., 2004).
- The choice of which areas of the rail network should receive investment for increased permissible line speed (Network Rail, 2018).

These general demands are synthesised with a family of rail networks to create a family of test-tasks that involve the allocation of a limited number of identical carriages between trains and the setting of permissible line speeds around the

network. The number of carriages allocated to a train determines its passenger capacity but, for these test-tasks, does not affect any other characteristics of the train. For the purpose of the test-tasks, any number of carriages greater than zero can be allocated to a train provided that the limit on the total number of carriages available is not exceeded. The passenger capacity of a train affects the performance of the network because it is related to the comfort and duration of passenger journeys. A line linking adjoining stations in the test-task network is homogenous and bi-directional. In these test-tasks, the permissible line speed can be one of two alternatives: a 'basic' level and an 'upgraded' level. The permissible line speed affects passenger journey times and hence the network performance. For the test-tasks, the carriage allocations and permissible line speeds are the optimisation variables whose values are chosen to maximise network performance calculated using the metric developed in Chapter 3.

The family of networks used in the test-tasks are chosen to have a high degree of symmetry so that the global optimum can be calculated analytically. A radial network design is used where the central station is connected to outer stations by two lines. The networks are named B2, B3 and B4 where the number refers to the number of links in the network. Figure 25 shows the topographies of the networks with a circle representing a station and a connecting edge representing a railway line. Each line within the network has one train operating upon it and the trains do not transfer between lines. Table 9 displays the number of lines and trains in the test-networks and how this controls the number of optimisation variables. The number of trains within the network is equal to the number of lines within the network so, from here on in, only the number of trains are stated. The trains operate to a symmetrical timetable and all lines have a dedicated platform at their connected station. The passenger load has a symmetrical origin-destination-time matrix. It is recognised that these networks are simplistic and idealised relative to real-world networks, however their properties are sufficient to test the relative performance of the GA and BO methods in preparation for application to more realistic cases. Furthermore, although the BO implementation tested cannot be scaled to tasks

representing the whole GB network in this way, i.e. every individual line represented by an optimisation variable, the task demonstrates the type of network parameters which managers may wish to optimise. To address a GB network sized task, the network model could be reformulated so that each optimisation variable represents multiple lines or a BO implementation that allows more optimisations variables might be used, e.g. the implementation presented by Li et al. (2017).

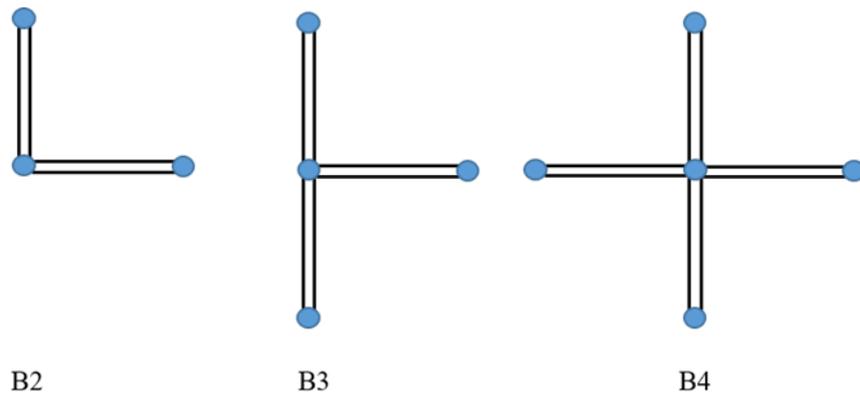


Figure 25 – Topographical representation of three different sized networks from the family of networks used for test-tasks. A circle represents a station. An edge between two stations represents a bi-directional line upon which trains travel. The name of the network topography is displayed below each network.

Network topography name	B2	B3	B4
Number of lines	4	6	8
Number of trains	4	6	8
Number of optimisation variables	8	12	16

Table 9 - The number of lines and trains for the three different sized networks used in test-tasks.

5.2.2 Formal definition of the test-tasks

Here the general definition of constrained optimisation given by (2.1), (2.2) and (2.3) is modified to the family of test-tasks described in the previous section.

The tasks can be described as optimising the distribution of \aleph identical carriages amongst T trains and selecting the permissible line speed of each of the \mathcal{L} lines from \wp discrete choices. The vector, x , is $T + \mathcal{L}$ dimensional with x_1, x_2, \dots, x_T describing the number of carriages allocated to trains 1 to T and

$x_{T+1}, x_{T+2} \dots x_{T+\mathcal{L}}$ describing the permissible line speeds of lines 1 to \mathcal{L} . The form of \mathbf{x} is therefore shown by:

$$\mathbf{x} = [x_1, x_2, \dots, x_T, x_{T+1}, x_{T+2} \dots x_{T+\mathcal{L}}] \quad (5.3)$$

The general objective function, $f(\mathbf{x})$, is modified to the specific objective function $F_1(\mathbf{x}; \lambda, \theta)$ which quantifies network performance, relative to other candidate networks, from the passenger perspective where λ describes the network parameters that are constant for this task (e.g. station locations, train performance, timetable) and θ describes the passenger load. Neither λ or θ are optimised. The test-tasks have no constraints placed on the choice of permissible line speed (i.e. all lines can have the maximum permissible line speed). Furthermore, there is no penalty associated with increasing the permissible line speed, meaning that the globally optimum solution has all permissible line speeds maximised. Although it may not be realistic to have no constraint or penalty placed on increasing permissible line speed, this does ensure that the global optimum to the task is known and therefore is appropriate for the purpose of the test-tasks. However, there is a constraint on the carriage allocations that a maximum of \aleph carriages can be distributed between all the trains, captured by:

$$\sum_{i=1}^{i=T} x_i \leq \aleph \quad (5.4)$$

A negative number of carriages cannot be allocated, this is captured by (5.5), and the limit on the number of choices of permissible line speed is captured by (5.6).

$$0 \leq x_i \quad \text{for } 1 \leq i \leq T \quad (5.5)$$

$$1 \leq x_i \leq \wp \quad \text{for } T + 1 \leq i \leq T + \mathcal{L} \quad (5.6)$$

$F_1(\mathbf{x}; \lambda, \theta)$ captures the performance of a network, relative to other candidate networks, on a percentage scale from 0% to 100%. 100% relates to the known global maximum, 0% relates to the known global minimum, i.e. where no carriages are allocated and all permissible line speeds are at their minimum.

Consequently, the challenge of maximising network performance relates to the task of maximising $F_1(\mathbf{x}; \lambda, \theta)$. As a result, the formal definition of the family of test-tasks can therefore be written as (5.7) subject to (5.4), (5.5) and (5.6).

$$\mathbf{x}^* = \underset{\mathbf{x} \in X}{\text{arg max}} F_1(\mathbf{x}; \lambda, \theta) \quad (5.7)$$

5.2.3 Calculating the value of the objective function

To calculate the value of $F_1(\mathbf{x}; \lambda, \theta)$, the formula given by (5.8) is used:

$$F_1(\mathbf{x}; \lambda, \theta) = \frac{\phi(\mathbf{x}; \lambda, \theta) - \phi(\mathbf{x}^*; \lambda, \theta)}{\phi(\mathbf{x}^-; \lambda, \theta) - \phi(\mathbf{x}^*; \lambda, \theta)} \quad (5.8)$$

where $\phi(\mathbf{x}; \lambda, \theta)$ is the network score (calculated using the method described in Chapter 3) relating to the PRaM simulation (described Chapter 4) of the network defined by \mathbf{x} and λ , carrying the passenger load defined by θ . The symbol \mathbf{x}^- denotes the candidate where no carriages are allocated and all permissible line speeds are minimised, and the symbol \mathbf{x}^* denotes the candidate where each train is allocated \aleph/T carriages and all permissible line speeds are maximised.

The experimental parameter, F_1^* , is introduced to describe the *target performance* of the solution. F_1^* is the smallest value of $F_1(\mathbf{x}; \lambda, \theta)$ that must be found by an implementation before terminating with a result. Candidates (\mathbf{x}) for which $F_1(\mathbf{x}; \lambda, \theta) \geq F_1^*$ are referred to as *acceptable*.

5.2.4 Features of the search space

For an optimisation task with integer variables, such as the ones considered in this chapter and the next, the number of candidates in the search space, N_C , is generally calculated by multiplying the range of each optimisation variable. For the family of test-tasks considered in this chapter, the formula given by (5.9) can be used.

$$N_C = \wp^L (T + 1)^\aleph \quad (5.9)$$

Candidates in the search space which satisfy the constraint functions, e.g. for this task (5.4), are referred to as *feasible*. For these test-tasks, the number of

feasible candidates in the search space, N_F , can be calculated with the formula given by (5.10).

$$N_F = \wp^{\mathcal{L}} \sum_{a=0}^{a=T} \frac{(a + \aleph - 1)!}{a! (\aleph - 1)!} \quad (5.10)$$

5.2.5 The Genetic Algorithm and Bayesian Optimisation implementations

The BO implementation used was a modification of a proprietary MATLAB 2017b function, here denoted '*bayesopt*'. The BO algorithm was seeded with two initial observations, relating to when all line speeds are maximised and all the carriages are distributed evenly amongst the trains, and when all line speeds are maximised and no carriages are distributed. Evaluating these candidates and updating the data set, D , replaces steps 1 to 3 in Algorithm 2 in Section 5.1.2. These evaluations are included in the value of count of objective function evaluations used. Within the GP regression model an ARD Matern 5/2 kernel was used, $k(\mathbf{x}_m, \mathbf{x}_p)$ - given by:

$$k(\mathbf{x}_m, \mathbf{x}_p) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5r^2}{3} \right) e^{-\sqrt{5}r} \quad (5.11)$$

where \mathbf{x}_m is the vector relating to the m 'th observation, \mathbf{x}_p is the vector relating to the p 'th observation, and σ_f is the signal standard deviation parameter and had a value of the standard deviation of the objective function values observed so-far, divided by $\sqrt{2}$. The value of r is given by:

$$r = \sqrt{\sum_{i=1}^n \frac{(x_{m,i} - x_{p,i})^2}{\sigma_i^2}} \quad (5.12)$$

where $x_{m,i}$ is the i 'th element of \mathbf{x}_m , $x_{p,i}$ is the i 'th element of \mathbf{x}_p , n is the number of optimisation variables within \mathbf{x} , and σ_i is the kernel length scale parameter, which controls how 'far apart' two values of \mathbf{x} need to be for the objective function values to become uncorrelated. Observing (5.12) it can be

seen that a different kernel length scale was used from each dimension of x . The value of this parameter was calculated as the standard deviation of the x_i values where observations have been made so-far. The GP model was initialised with a zero value mean function. The *bayesopt* function was applied with the ‘Probability-of-Improvement’ acquisition function, i.e. the probability that an observation will ‘better’ the ‘best’ observation observed so far, and an exploration ratio of 0.5. Clearly there is opportunity for investigating and even tuning the parameter values, location and number of seed points, or function types used here. However, considering that the aim of this thesis is to investigate SUPREME and hence Bayesian Optimisation for a range of different challenges and that tuning for one challenge might diminish its performance for another, this is considered out of scope.

The first modification made to *bayesopt* was the removal of a portion of code that checked whether the task submitted to *bayesopt* had constraint functions that could be satisfied. The removed code compared 10^6 candidates, arbitrarily selected from the search space, against the constraint functions. If no candidates from this selection satisfied the constraint functions, the unmodified *bayesopt* determines that the constraint functions cannot be satisfied and terminates the optimisation procedure. However, using (5.10) and (5.9) with parameters values of 2, 8, 48 and 8 for φ , \mathcal{L} , \mathcal{R} and \aleph respectively, which are also values used in these experiments, gives N_F/N_C to the order of 10^{-40} . This meant that the unmodified *bayesopt* often, incorrectly, determined the constraints of this task as ‘impossible to satisfy’. Removal of this code meant that the modified *bayesopt* can be used for this task without early termination. In the context of these test-tasks, there is little reason for the removal of this code to alter the effectiveness of the BO implementation other than the desired effect of preventing it to terminate unnecessarily.

The second modification was to alter the number of local optimisations used to approximate the global maximum of the acquisition function (line 8 of Algorithm 2). The local optimisations begin from arbitrarily selected candidates within the search space. This means that there is no guarantee that the

acquisition function maximum found by this process relates to a feasible solution. If N_C is 'large', the number of local optimisation seeds is 'small' and N_F/N_C is 'small', this causes *bayesopt* to regularly find infeasible acquisition function maximums. This leads to a waste in computation because the objective function is evaluated at the candidate relating to the acquisition function maximum at every iteration, regardless of whether it satisfies the constraint functions. Increasing the number of local optimisation seeds increases the likelihood of locating a maximum of the acquisition function that relates to a feasible candidate, and hence increases the *bayesopt*'s effectiveness. Consequently, the number of seeds was increased to the 10^7 , which was the maximum that allowed MATLAB to remain stable owing to increases in the system memory required.

The GA implementation used was a modification of a proprietary MATLAB 2017b function, here denoted '*ga*'. The default *ga* settings were used with a cross-over rate of 0.5, a mutation rate proportional to the initial range of values in the population (shrinking to zero at the final generation) and a uniform stochastic selection function. In the literature, GA implementations are used with population sizes typically ranging from 30 to 100. Here, the minimum of this range has been used so that, with a budget of objective function evaluations in the order of 100 (a typical number required by the BO method), *ga* has the opportunity to converge on a solution. A population size that did not allow the *ga* algorithm to iterate before exceeding the number of evaluations use by the BO method would be equivalent to a Random Search algorithm and would not be representative of GA effectiveness.

5.2.6 Experimental method

The number of objective function evaluations used by the GA and BO implementations was measured for eight 'jobs', i.e. a specific combination of test-task and F_1^* that is input to a function. A test-task is defined by the number of trains in the network and the number of carriages to be allocated. To collect experimental data, a job is submitted to a function and the algorithm iterates until an acceptable solution is found or a limit on the number of objective

function evaluations is reached. Because the algorithms are non-deterministic it is necessary to repeat, independently, each experiment multiple times and describe distributions. For comparison between the GA and BO methods, identical tasks are submitted to the algorithms. Because the control parameters of each implementation differ, attainment of identical terminal objective function values is used to ensure like-for-like comparison. The algorithm computational cost is measured using the computation time. This is machine specific, but gives an indication to the behaviour of algorithm and is comparable across jobs since they were performed on the same machine, an Intel Xeon Dual Processor @ 2.4 GHz.

5.3 Results

Figure 26 is a box and whisker plot comparing the number of evaluations required by *ga* and *bayesopt*, η^{GA} and η^{BO} , for eight jobs. Each box and whisker represents, on a logarithmic scale, a distribution from 32 repeats of an experiment. Distributions for the same job are plotted next to each other and separated by vertical dashed lines to allow easy comparison. The job and method that each distribution relates to is displayed on the x-axis. The notation 'J1', 'J2', 'J3' etc. can be cross-referenced against Table 10 to observe the experimental parameters defining the job. A full factorial design for three experimental parameters and two levels has been used. An arbitrary limit of 1.6×10^4 objective function evaluations was used which corresponds to approximately 48 hours of objective function computation time per experiment. This limit only affects the median and quartile results of 'J6,GA', but still allows a discernible difference with 'J6,BO'. The box notches indicate a 95% confidence interval of the median. When comparing distributions for the same job but different methods, there are no cases in which the notches overlap indicating that, on average, $\eta^{GA} > \eta^{BO}$ with approximately 95% confidence (Chambers et al., 1983). The features of the 'J6,GA' plot are indiscernible because for 31 of the experiments, the GA implementation did not find an acceptable solution within the limit of objective function evaluations. Comparing all eight jobs, the mean of the factor differences between η^{BO} and

η^{GA} is 43 with a standard deviation of 76. Furthermore, excluding J6 where the comparison is not valid, the inter-quartile ranges of the BO distributions are narrower (the log scale of the y-axis means this is true even for J7), indicating that BO is more consistent in the number of objective function evaluations required to find an acceptable solution.

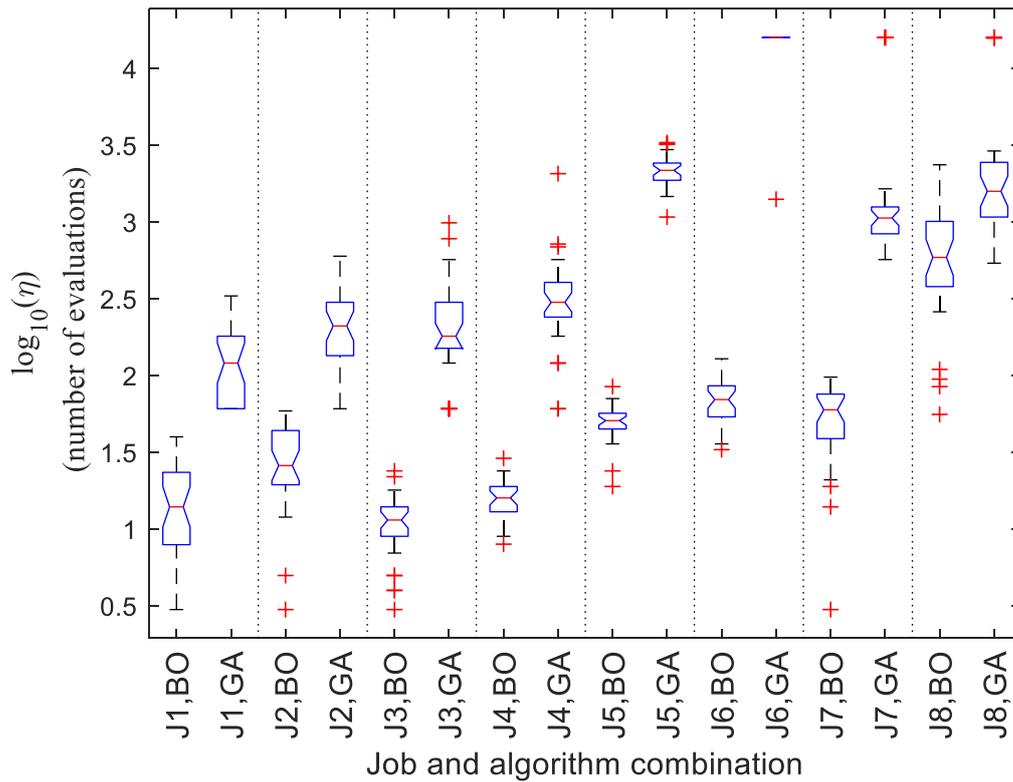


Figure 26 - Box plots showing the distribution of the number of objective function evaluations required, η , for eight different jobs. The y-axis is on a \log_{10} scale. Each box plot represents a distribution of 32 repeat experiments. Table 10 displays the experimental parameters of each job described by the 'J number'. The whiskers extend to a maximum of the inter-quartile range below and above the 25th and 75th percentiles respectively. Data outside this range is considered an outlier and is shown by a cross.

	J1	J2	J3	J4	J5	J6	J7	J8
Trains	4	4	4	4	8	8	8	8
Carriages	8	8	48	48	8	8	48	48
F_1^* (%)	90	95	90	95	90	95	90	95

Table 10 – The experimental parameters of the eight jobs for which the BO and GA methods are compared in Figure 26.

5.3.1 The number of objective function evaluations and target performance

Following the comparison to a GA, the relationship between the number of objective function evaluations required by *bayesopt* and the target solution performance was investigated. Five different tasks were considered and the y-axis value of Figure 27 is the logarithm of the median from 24 repeats of the experiment.

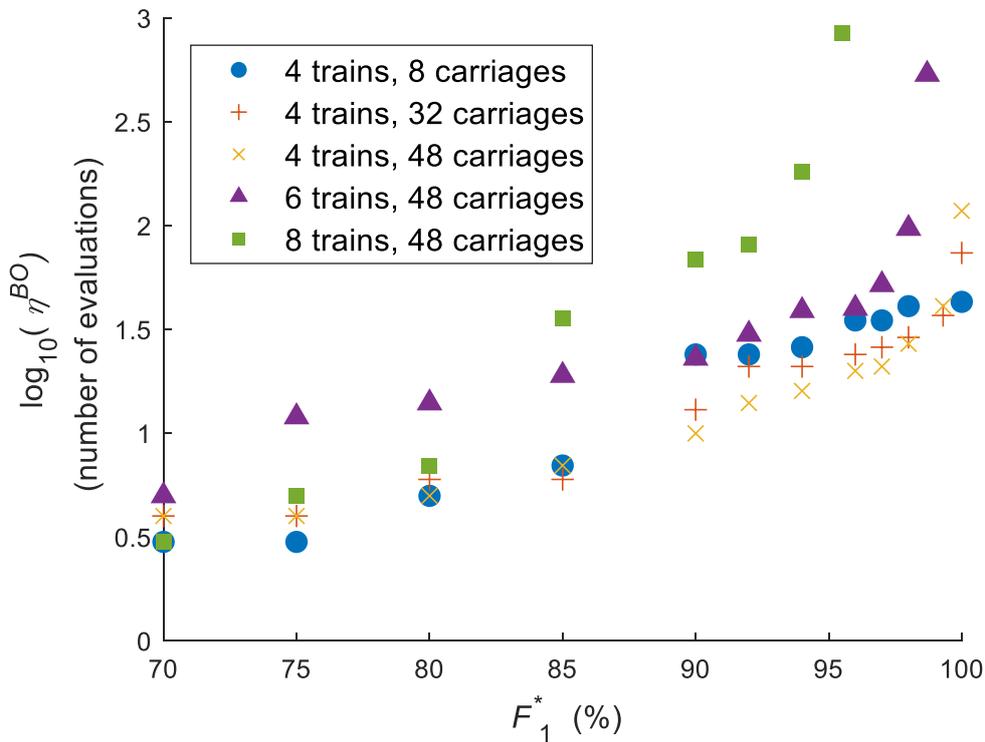


Figure 27 – Scatter plots to show the number of objective function evaluations required by Bayesian Optimisation, η^{BO} , at varying target performance, F_1^* . Data is shown for five different tasks. The values plotted are the median of a distribution of 24 repeat experiments. The y-axis displays η^{BO} on a \log_{10} scale.

The data shows a positive relationship between F_1^* and $\log_{10}(\eta^{BO})$ that is at least linear. This means that the relationship between F_1^* and η^{BO} is at least exponential. The increase in η^{BO} is more sensitive to the number of trains in the task than the number of carriages. It can be seen that for the two most difficult problems (six and eight trains, 48 carriages) there is no data for $F_1^* > 99\%$ and 95.5% respectively. This is because after 4 days of computation time F_1 had not been increased and the experiment terminated. While this appears limiting it is

worth nothing that, in the case of the 8 train task, relaxation of the target performance to 95% enables the target solution performance to be reached in only 1 hour and 20 minutes of computation time. The rapid increase in η^{BO} for solutions close to the optimum is believed to be because of the GP model and not inherent to BO. This is further explained in the Section 5.4.

5.3.2 Algorithm computation time

To investigate the relationship between the measured total computation time for the *bayesopt* algorithm (Π^{BO}) at varying iterations, Figure 28 plots Π^{BO} and η^{BO} for different numbers of trains. The computation time required was found to be insensitive to the number of carriages so this is not displayed. For each task the data plotted is the median from eight repeat experiments.

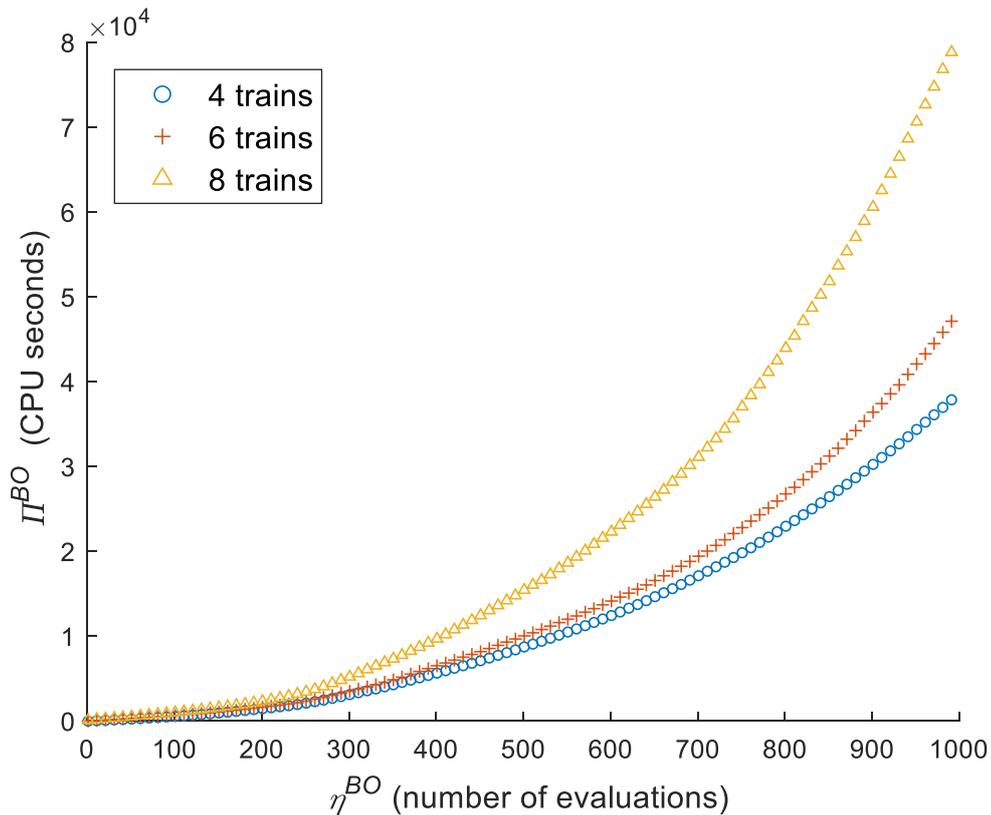


Figure 28 – The total algorithm computation time of the Bayesian Optimisation implementation, Π^{BO} , at increasing objective function evaluations, η^{BO} . This relationship is investigated for three different tasks. The value displayed by each marker is the median from eight repeat experiments.

Figure 28 shows that the computational cost of the *bayesopt* implementation grows rapidly at later iterations of the algorithm. This is thought to be specific

to the use of a GP model within BO and the associated matrix inversion, rather than inherent to the BO approach. The data also shows that the algorithm computation time increases faster for tasks with more trains. This is because increasing the number of trains in the task increases the GP model matrix dimensions and the associated cost of inverting it. Taking the logarithm of Π^{BO} and η^{BO} , to determine the order of computational cost as a function of η^{BO} , $O(\eta^b)$, gives $b = 2.73$. However, testing to larger values of η^{BO} would be required to confirm b is not larger.

5.3.3 Computation time of the whole procedure

Values can be substituted into the inequality given by (5.2) to investigate the smallest computation time for a single objective function evaluation (γ), for which it is satisfied. This represents the ‘threshold’ at which it becomes computationally cheaper to use BO over GA. The data for Π^{BO} from Figure 28 is used as well as the median values of the data shown in Figure 26 for η^{BO} and η^{GA} . The value of Π^{GA} is assumed to be the product of η^{GA} and the computational cost of a single iteration of the GA algorithm excluding objective function evaluations, π^{GA} . To reflect that π^{GA} may vary between GA implementations, here it is given values of 0.001 seconds and 5 seconds to capture reasonable lower and upper bounds inferred from results published by Kohmoto et al. (2003) and Zhu et al. (2017). Figure 29 plots the threshold γ for all of the jobs shown in Figure 26. It can be seen that for most of the jobs, the threshold time is less than a second, regardless of π^{GA} . For Job 6 and $\pi^{GA} = 5$, there is no positive value of γ which does not satisfy inequality (5.2), indicating that in $\Pi^{GA} > \Pi^{BO}$ for all tasks, i.e. the BO method is computationally cheaper. This is because of the large number of objective function evaluations required by the GA implementation for this job. For Job 8, the threshold value of γ is substantially greater but the sensitivity to changes in π^{GA} is still less than a second.

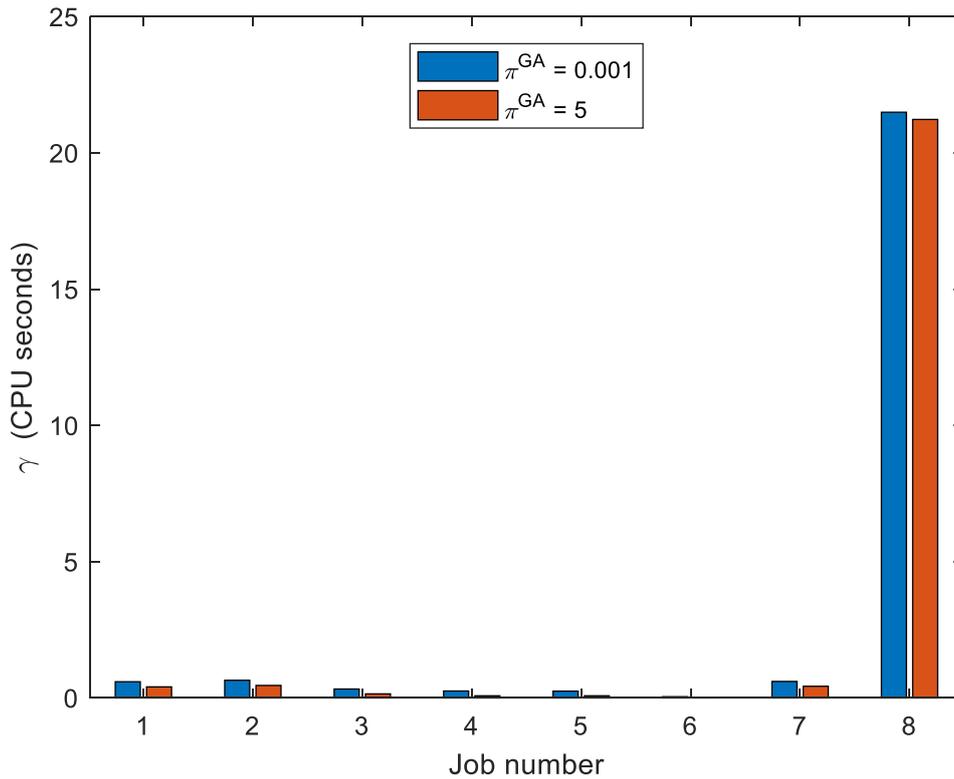


Figure 29 – The computation time for a single objective function evaluation (γ) at which the total computational cost of the optimisation becomes equal, for the eight jobs in the results of Figure 26 and under two different assumptions of the algorithm computational cost of one GA implementation iteration (π^{GA}).

5.4 Discussion

The results of Figure 27 indicate that increasing the target performance of the solution leads to an at least exponential increase in the number of objective evaluations required by BO. This is thought to be because in general, for jobs with a high target performance, a proxy function which accurately models the objective function is required. This demands a higher density of evaluations but when many evaluations become clustered in one region the ability of BO to effectively select new candidates is reduced (McLeod et al., 2018). In addition to the increase in objective function evaluations, the results in Figure 28 indicate that there is a super-linear relationship between the number of evaluations and the algorithm cost. This is thought to result from the matrix inversion when calculating a GP. Both these effects are thought to be a consequence of implementing BO with a GP model rather than inherent in the BO method. Taken together, the results of Figure 27 and Figure 28 indicate that

when a GP regression model is used as the probabilistic model within BO, increasing the quality of the solution significantly increases this total expense. However if accelerated techniques for calculating GP models, such as the one developed by Gardner et al. (2018), can be implemented with the BO method, this effect might be reduced. As identified by McLeod et al. (2018) it is likely that in applications where the target performance for the solution is high, a multi-strategy optimisation method would be most effective, i.e. switching from BO to GA (or another alternative).

The results of Figure 26 indicate that for certain tasks BO may find 'good' solutions in significantly fewer objective function evaluations than a GA. For tasks involving expensive-to-compute objective functions, this leads to a reduction in total computational expense. Figure 29 shows that the threshold computational cost of one evaluation (γ) for which BO is cheaper than GA was less than one second all but one of the jobs in this chapter and for the remaining job it was approximately 20 seconds. Given that the results of Chapter 4 have shown that the objective function associated with a small network requires approximately 20 seconds of computation, but this increases with the size of the network, the results of Figure 29 indicate that the BO method is more suitable for use in SUPREME when applied to realistic sized networks.

5.5 Summary

GAs are a well-established optimisation method. However, there is a pressure to keep the computational cost of an objective function evaluation low because they typically require, for real-world applications, in the order of 10^4 objective function evaluations or more. BO uses information from all previous evaluations of the objective function to guide the selection of new candidates so that the most beneficial ones are targeted. This means that BO has the potential to find solutions of a similar quality to a GA, but in fewer objective function evaluations, and to be computationally cheaper for tasks with expensive-to-compute objective functions. This was experimentally confirmed using a range of test-tasks where the mean factor difference between the

numbers of evaluations required by the methods was 43 with standard deviation of 76. However, due to the overhead in the algorithm of the BO implementation tested, a super-linear relationship was found between the total algorithm cost and the number of objective function evaluations required. Furthermore, the relationship between the number of objective function evaluations required and improving solution quality is at least exponential. This is thought to be an effect of using a GP model within the BO algorithm and not inherent to the BO method itself. This means that the BO implementation tested is better applied to tasks involving expensive-to-compute objective functions where approximate answers are satisfactory and the budget for computational expense is small. Two approaches which may improve the solutions found by BO are to either: improve the probabilistic model used within BO, or, sequentially combine BO with GA for a multi-strategy optimisation method.

The BO method is selected for use within SUPREME because the results in this chapter suggest it is well suited for with tasks with expensive-to-compute objective functions and the results of Chapter 4 indicate that the objective functions encountered within SUPREME will be of that type. For the remainder of this thesis, the BO implementation used in this chapter is used as the optimisation algorithm within SUPREME and symbols defined within this chapter that can apply to GA or BO, e.g. η , are used without the superscript but relate to BO.

Chapter 6

Applying SUPREME

Chapter 3, Chapter 4 and Chapter 5 have discussed the three main components of SUPREME in isolation. This chapter describes the application of the SUPREME framework to three case-study examples, and evaluates its effectiveness. To demonstrate the adaptability of the SUPREME framework, the case studies involve different networks and different optimisation tasks. The first case study involves optimising a timetable which is typically undertaken with a planning window of the order of months. The second case study considers the situation where equipment failure has occurred and resources must be reallocated in a time window of the order of hours. The third case study considers the optimisation of network design, where the time window may be of the order of years. In the case studies, due to the data available to the author, some assumed values are used to represent the real-world challenges. Although this might reduce the accuracy of the task, it is emphasised that the focus of this chapter is to demonstrate the application of the SUPREME framework rather than ‘solve’ the real-world case studies investigated.

6.1 Case study one – optimising the timetable of the regional train network around Derby

The first case study considers the part of the regional network which connects Crewe, Nottingham, Matlock and Derby in GB. This network is part of the East Midlands rail franchise which has been managed by East Midlands Trains (EMT) from 2007 to the time of writing in 2019. Figure 30 A) illustrates the approximate geographical location of the network. Figure 30 B) uses network map data, provided by the Association of Train Operating Companies (c2015) and Network Rail (2016), to illustrate how these stations are connected to each other and the wider GB network. Intermediate stations are also included in the case study and Figure 30 C) illustrates these as well as their connections with

either one bi-directional line, e.g. between Crewe and Alsager, or two single direction lines, e.g. between Alsager and Kidsgrove.

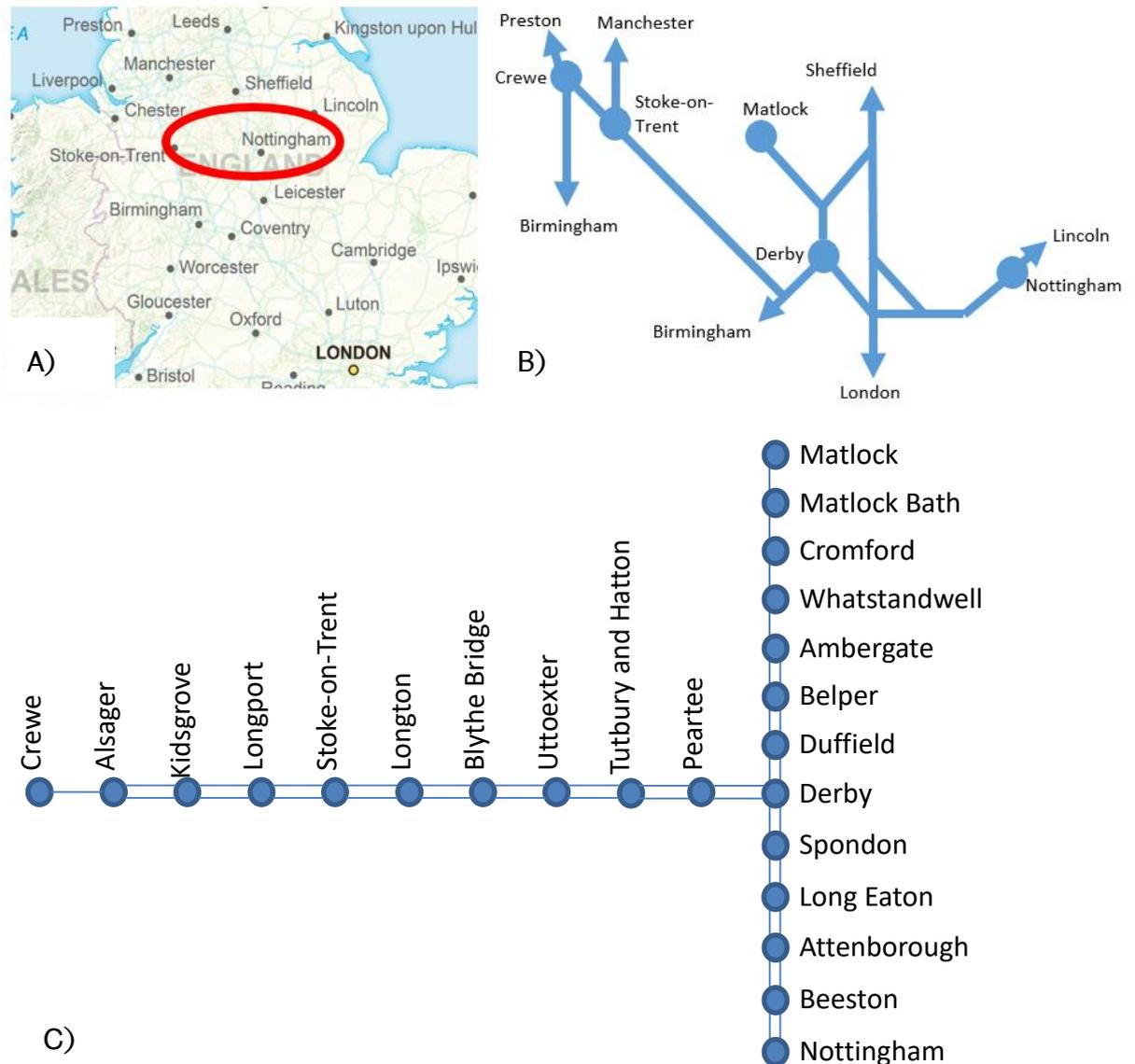


Figure 30 – A) displays with an ellipse the geographical area of the case study network and contains OS data © Crown copyright and database right (2019)⁷. B) uses circles to show five of the stations included in the study network and displays how these are connected to other cities. Connections are informed by maps published by the Association of Train Operating Companies (c2015) and Network Rail (2016). C) displays all the stations included in the model network and their connections with either one bi-directional line or two single-direction lines.

⁷ This is used under an Open Government Licence, a copy of which can be found at the URL: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

The Department for Transport (2017a) identified challenges for this network which included overly-long passenger waiting times for connecting trains and overcrowding. Assuming a punctual service, the timetable controls the waiting time for a passenger making connections between trains, i.e. the *connection time*. Furthermore, because it also controls the interval between trains and hence the accumulation of passengers waiting for a train, the timetable also affects the level of crowding. The SUPREME framework captures the effect of the timetable upon passenger connection times and crowding, as well as the overall effect on passenger disutility. It should be noted that in this investigation these are the only effects of altering the timetable which are captured - other effects are not captured, for example that passengers often prefer a 'clock face' timetable (Johnson et al., 2006). In the study network, the trains operate between Crewe and Derby, and Nottingham and Matlock (calling at Derby), therefore passengers make connections at Derby. Consequently, in this case study the SUPREME framework is used to optimise the arrival and departure times of regional trains at Derby station. The timetabled interstation journey times and number of trains are based on the current timetables for the network provided by EMT (2019) and, together with the dwell times, are not altered. The movements of services that the Department for Transport (2017a) describes as 'Intercity' or 'Inter-urban', which also pass through the real-world network, are not modelled and assumed not to constrain the optimisation.

6.1.1 Formal definition

The optimisation task models three trains operating between Crewe and Derby as well as three trains between Nottingham and Matlock. Because the timetable follows an hourly pattern throughout the day, where interstation run times and dwell times do not vary, it is modified by adding a time supplement to all arrival and departure times of a train. This is done for each train independently with integer minutes. The notation discussed in Chapter 2 can be adapted to this task where the vector x captures the time supplement to all of the six trains with form shown by:

$$x = [x_1, x_2, x_3, x_4, x_5, x_6]$$

(6.1)

where the element x_i denotes the time supplement for train i . The formal definition of the task can therefore be written as (6.2) subject to (6.3), (6.4) and (6.5).

$$\mathbf{x}^* = \underset{\mathbf{x} \in X}{\operatorname{arg\,min}} F_2(\mathbf{x}; \lambda, \theta) \quad (6.2)$$

$$-29 \leq x_i \leq 30 \quad \text{for } i = 2,3,5,6 \quad (6.3)$$

$$-10 \leq x_i \leq 30 \quad \text{for } i = 1 \quad (6.4)$$

$$0 \leq x_i \leq 30 \quad \text{for } i = 4 \quad (6.5)$$

The value of the objective function, $F_2(\mathbf{x}; \lambda, \theta)$, is equal to the network score (calculated using the method described in Chapter 3) relating to the PRaM simulation (described in Chapter 4) of the network defined by \mathbf{x} and λ , carrying the passenger load defined by θ . The objective function cannot be expressed on the same percentage scale as used in Chapter 5 because the global maximum and minimum are not known. Consequently, the raw network scores calculated using the metric described in Chapter 3 are used. Because these relate to disutility, improving the network score is a minimisation task. The purpose of the constraints shown by (6.3), (6.4) and (6.5) is to preserve the order of trains as well as controlling the operational hours of the network to be within 6am to 1am. The search space contains 4.7×10^{10} candidates.

6.1.2 Constant parameters describing the network

To describe the topography of the network, the Origin Destination Matrix (ODM) supplied by Steer Davies Gleave (2017b) provides the interstation line distances, and the Sectional Appendix supplied by Network Rail (2019) provides the permissible line speeds.

The trains are modelled as British Rail (BR) Class 153 DMUs because Pettit (2017) describes that Class 153, 156 and 158 units operate on the study network (of which only the Class 153 uses a single carriage formation), but Furness (2017) reports that trains with 'one carriage' operate 'most of the time' on the Crewe-Derby route and The Friends of Derwent Valley Line (2019) report that

two out of three services on the Matlock-Nottingham route are single carriage. The number of seats for a Class 153 is provided by Angel Trains (c2019a) as 75. To determine the standing capacity of a Class 153, data provided by the UK government (n.a, 2003) was used to calculate the standing capacity as a proportion of the total capacity for Class 150 and Class 156 trains, which are both similar to a Class 153. The mean of these values is 31% which when applied to the Class 153 gives a standing capacity of 23. Although this is likely an underestimate of the maximum number of standing passengers encountered in the real world, it represents the ‘recommended’ maximum and so is used here.

The speed-dependent tractive force available ($F_{available}$) to a Class 153 was provided by extracting data from the Birmingham Rail Virtual Environment (BraVE) simulator described by Umiliacchi (2016). Because this data did not fit well to the relationship described by (4.2), $F_{available}$ was determined during PRaM by linear interpolation of the extracted data. Other parameters describing the performance of a Class 153 were also extracted from the BraVE simulator and have values shown in Table 11. The Davis formula coefficients were determined by fitting a second order polynomial to extracted speed and resistance data, this is further discussed in Appendix III.

Parameter	Value
Tare mass, M_T	45 tonnes
Rotary allowance, r	0.17
Maximum speed	121 km/hour
Davis formula \mathcal{A}	1.09 kN
Davis formula \mathcal{B}	1.56×10^{-10} Ns/m
Davis formula \mathcal{C}	$4.76 \text{ Ns}^2/\text{m}^2$

Table 11 – Parameter values used to describe a British Rail Class 153, derived from data extracted from the BRaVE simulator described by Umiliacchi (2016).

6.1.3 Capturing the passenger load

The passenger load input was determined from three data sources:

- The Origin Destination Matrix (ODM) (Steer Davies Gleave, 2017b) describes the passenger flows between each station pair in the study

network. This was shared through private communication with the Information and Analysis team of the Office of Rail and Road (ORR).

- The Estimates of Station Entry Exit (ESEE) data describes estimates of the number of passengers to enter and exit each station, published by the ORR (Steer Davies Gleave, 2017a).
- Hourly variations in the number of passenger arrivals and departures in London (Department for Transport, 2012). No data specific to the case-study network was available to the author, however it is assumed the London data approximates the temporal variation in demand for travel and could easily be replaced by data specific to the network if this became available. To capture the variation in total demand for rail travel, the arrivals and departures were combined.

The ODM underestimates the passenger flows because it only captures passengers whose origin *and* destination is inside the study network. For example, passengers travelling from Alsager to Derby before making an onward connection to London would not be captured. To quantify this effect, the total entries and exits in the ESEE data was compared with the total number of flows in the ODM beginning or ending at that station. To reduce the effect of passengers captured in the ESEE who do not travel on the services included in this study, this comparison is done for stations that are only called at by services included in the study. The factor difference between the ESEE and ODM data was recorded for each station compared, and the mean calculated as 14%. The mean value was then used to scale all passenger flows between station pairs where at least one station is called at by services not included in this investigation. When determining the time that each passenger arrives at their origin station the daily total number of passengers in each flow was scaled to the hourly variation in passenger demand data. Within each hour, the 'ideal' time that passengers would wish to travel was assumed to have a uniform distribution. However, Ingvardson et al. (2018) show that at train intervals of five minutes or more, at least 43% of passengers plan their arrival time considering the departure time of a train. They report the 'average' difference between passenger arrival time and train departure time, i.e. *the*

waiting time, for train intervals from five minutes to one hour but do not state what statistical measure they mean by ‘average’. The median of this distribution across the train intervals is six minutes. Consequently, in the model discussed here, passengers are modelled to arrive at their origin station (the time PB1 occurs in Figure 13 of Chapter 4) six minutes before the departure of their train. Although Ingvardson et al. also show that for a given train interval there is distribution of waiting times, this effect is not included in the PRaM input because the distribution of passenger agent arrival times would remain constant for all timetables and hence will not affect the relative performance of different candidates. The passenger load data therefore captures: passengers who travel on the services studied, the time-varying demand for travel throughout the day, and that passengers plan their arrival time depending on the timetable.

6.1.4 Experimental method

To compute the task defined by (6.2) to (6.5), an initial feasible candidate was arbitrarily chosen and the objective function evaluated using the method described in Section 5.2.3. The BO implementation discussed in Chapter 5 was then used with the ‘Expected-Improvement-Plus’ acquisition function, and one arbitrarily chosen feasible x used to seed the algorithm. A fixed budget of objective function evaluations, η_{max} , were used after which the algorithm terminates and results are returned containing the value of the objective function at every evaluation number (η). There was no restriction upon the objective function value of the solution.

For a given value of η_{max} , the solution returned by SUPREME is sensitive to the initial candidate. Consequently, the optimisation task was repeated, independently, 40 times with different initial candidate and the distribution of values relating to the results were examined. All experiments were computed on an Intel Xeon Dual Processor @ 2.4 GHz.

6.1.5 Results

The BO implementation function returns the ‘best’, in this case the minimum, network score found so-far at every algorithm iteration. Figure 31 displays two plots, both showing data relating to the distribution of ‘minimum network score found so-far’ values at different values of η . For both plots, the network score is expressed in units introduced by Figure 8 in Chapter 3. The left plot shows a trace for every optimisation with the maximum and minimum traces at $\eta = 1$ emphasised to show that there is a weak relationship between the network score at $\eta = 1$ and $\eta = 100$, relative to other repeats of the optimisation. For all optimisations, the network score at $\eta = 100$ is less than at $\eta = 1$, indicating that the network score has been improved. The median improvement in network score, compared to the corresponding initial score, is 19% with standard deviation of 7%. The right plot uses markers to show features of the distribution and also plots the network score of the unmodified timetable, i.e. $x = [0\ 0\ 0\ 0\ 0\ 0]$. When $\eta = 100$, the difference between the minimum and upper quartile marker is 3% of the difference between the median marker at $\eta = 1$ and $\eta = 100$, indicating that 75% of the optimisations found a solution ‘close’ in performance to the best one found out of all the optimisations, hereafter referred to as the *best solution*. The value of the median marker at $\eta = 100$ is 13% less than the network score of the unmodified timetable. Taken together the results shown in Figure 31 indicate that the BO implementation is effectively searching the search space regardless of the candidate it is initialised with.

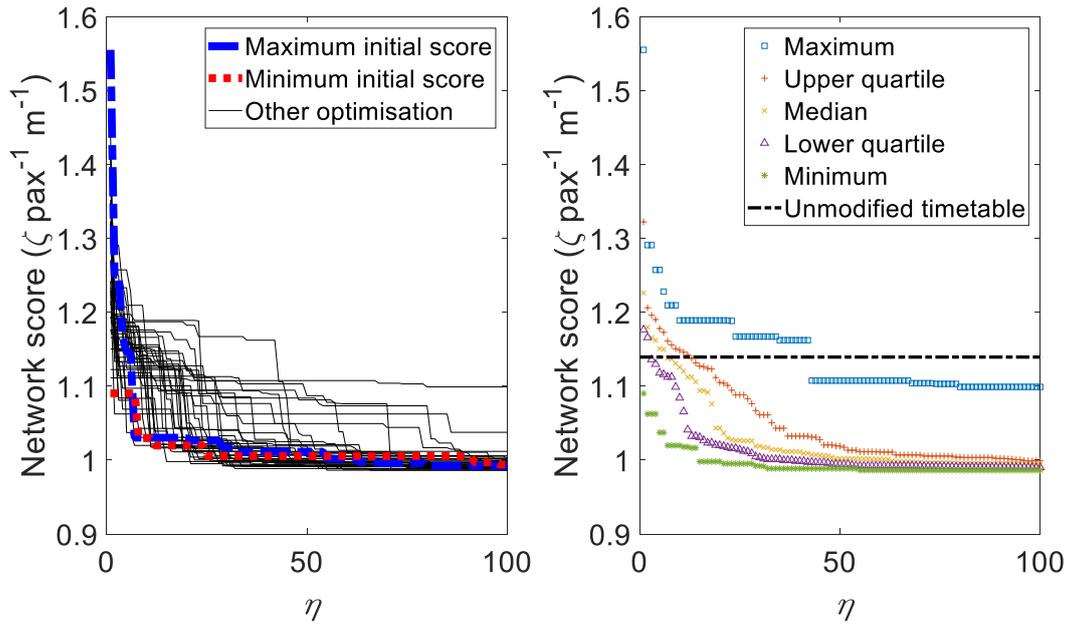


Figure 31 – Data relating to the distribution, from 40 independent repeats of the optimisation, of ‘minimum network score found so-far’ values at different numbers of objective function evaluations (η). The left plot displays a trace for each optimisation with the trace emphasised for the maximum and minimum network score at $\eta = 1$. The right plot uses markers to show features of the distribution and the network score of the unmodified timetable.

Figure 32 displays x relating to the ‘minimum network score found so-far’ values at $\eta = 1$ and $\eta = 100$, for all forty repeat optimisations. The value of each element of x has been plotted with the index (i) on the x-axis and the optimisation variable’s value on the y-axis. For the $\eta = 100$ plot, the filled markers relate to the best solution out of all the forty repeat optimisations. Observing the $\eta = 1$ plot it can be seen that the optimisation variable values distributed across close to their full feasible range. This indicates that, as expected, x is being initialised with uniform distribution, therefore the solutions found by the optimisation are free from any bias in initialisation. Where the optimisation variables have not taken a value in the feasible range is because of the small (40) sample size of the distribution. There are no x_1 and x_4 values less than -10 and 0 respectively because of the constraints capturing the operational hours of the network.

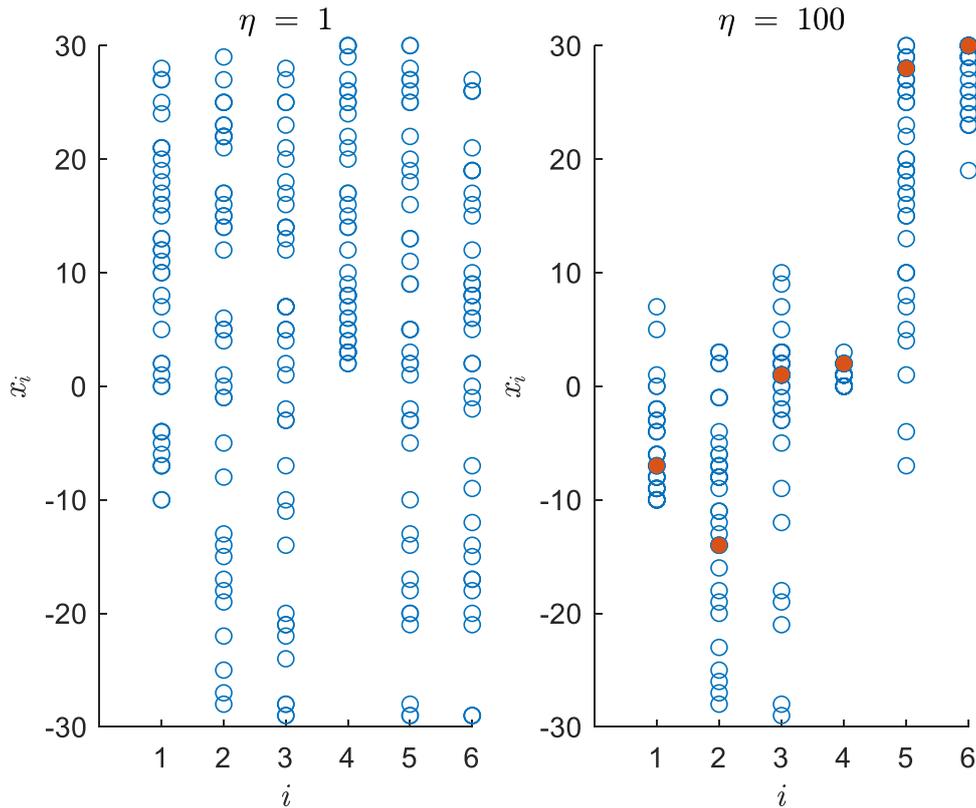


Figure 32 – The values of each optimisation variable at initialisation ($\eta = 1$) and the end of the optimisation ($\eta = 100$), shown for forty repeats of the experiment. The optimisation variable index is shown on the x-axis. The value of the variable is shown on the y-axis. For the $\eta = 100$ plot, the markers for the values relating to the 'best' solution found are filled.

For the $\eta = 1$ plot in Figure 32, the sum of the standard deviation for all six variables is 91, whereas for the $\eta = 100$ plot it is 35. This indicates that the repeat optimisations are finding solutions that are more similar than the initialisations. Furthermore, the values of the optimisation variables inform how the timetable might be modified. For example, all the solutions have $x_4 = 0$ to 3, indicating that it is important not to modify the timetable of Train 4 by more than 3 minutes. However, there is a wider range for the value of x_3 indicating that the timetable of Train 5 is less important, but adding a time supplement of 28 minutes to its timetable is most advantageous.

To investigate the effect of the best solution upon passengers, Figure 33 displays the histograms of the individual distance-normalised passenger journey scores for the unmodified timetable and the timetable relating to the best solution. So that differences are visible across the whole range of journey

scores, the histograms have been split over two plots. The left plot shows the frequency of journey scores from 0 to 5 and the right plot shows the frequency of journey scores from 5 to 11 with differently scaled y-axis. Comparing the histograms shows that the best timetable increases the frequency of the two lowest journey score groups by 20% and reduces the modal journey score, indicating that more passengers have a ‘very good’ journey in comparison to the unmodified timetable. The histograms also show the best solution timetable changes the frequency of all journey scores. This indicates that even though only 46% of passengers change train at Derby, and hence will have their journey affected directly by changes to the timetable, many passengers have their journey improved by the best solution timetable because of the knock-on effect of reduced crowding. Furthermore, observing the right plot shows that the maximum journey score of the best solution timetable is 20% less than the unmodified timetable, indicating that ‘very bad’ journeys are also improved. Passengers who experience very bad journeys are possibly more likely to complain, therefore improving the journey of these passengers, even though there are few, might be important to network managers. Where passengers have been delayed, reducing the severity of ‘very bad’ journeys might also reduce the proportion of passengers that request ‘delay compensation’ and, as a result, reduce losses in ticket revenue for network managers.

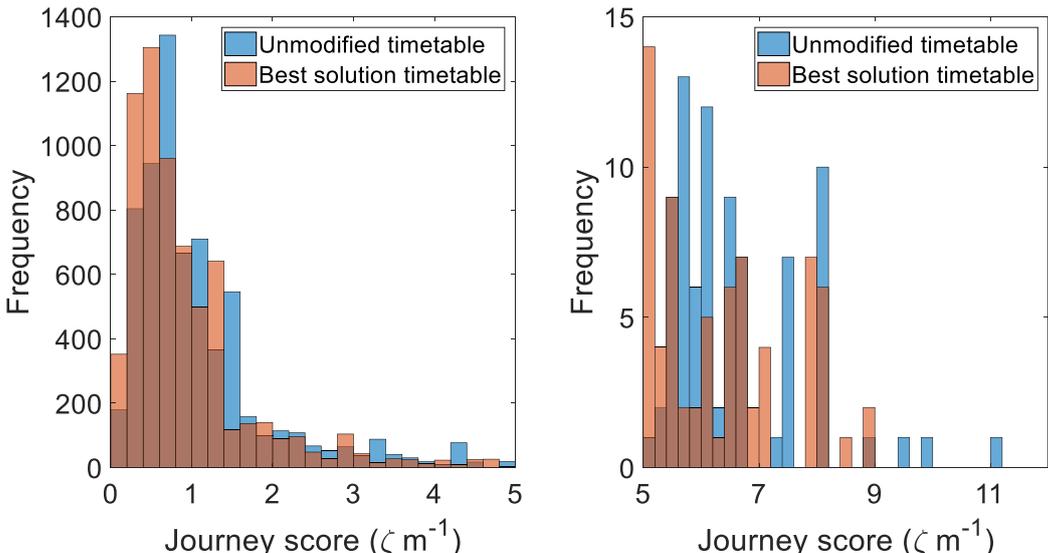


Figure 33 – Two plots showing the histogram of passenger journey scores for the unmodified and best solution timetable. So that differences are visible for

the whole range of journey scores, the x-axis is split over the two plots with differently scaled y-axis.

To investigate the effect of increasing the budget of algorithm iterations, the value of η_{max} was increased to 300 and the experiment repeated with the same initial candidates as used for the results in Figure 31. Figure 34 uses markers to show features of the distribution of 'best network score found so-far' values at different η . It can be seen that, compared to $\eta_{max} = 100$, the additional algorithm iterations improve the solution found by the worst-case optimisation so that it has performance closer to the solution found by the best-case optimisation. As a percentage of the unmodified network score, the difference between the minimum and maximum marker at $\eta = 100$ is 8.9%, whereas at $\eta = 300$ it is 1.6%. Compared to $\eta_{max} = 100$, the additional algorithm iterations also improve the solution found by the best-case optimisation, however this is difficult to see in Figure 34 because the improvement is very small. As a percentage of the unmodified network score, the difference between the minimum at $\eta = 100$ and $\eta = 300$ is 0.2%. These results indicate that, for this test task, increasing η_{max} from 100 to 300 reduces the range of network scores relating to solutions found by SUPREME, rather than substantially improving the best solution, i.e. it improves the consistency of SUPREME. Increasing η_{max} from 100 to 300 more than triples the computational cost of each optimisation because, as Chapter 5 has shown, the algorithm computation cost of the BO implementation increases super-linearly. Since timetabling is typically conducted with a planning window of months or more, in the real-world application of SUPREME this increase in computational cost may be justifiable for even a 0.2% improvement in the solution. However, for the purposes of investigating the effectiveness of SUPREME when applied to the task described in Section 6.1, the results of Figure 34 indicate that a value of $\eta_{max} = 100$ is sufficient because the best solution is of greater interest than the consistency of SUPREME. This will have little effect on other results discussed in this section because either: the worst-case optimisation is not included, e.g. Figure 33, or the full distribution of optimisations is shown but focus is on the best-case, e.g. Figure 32.

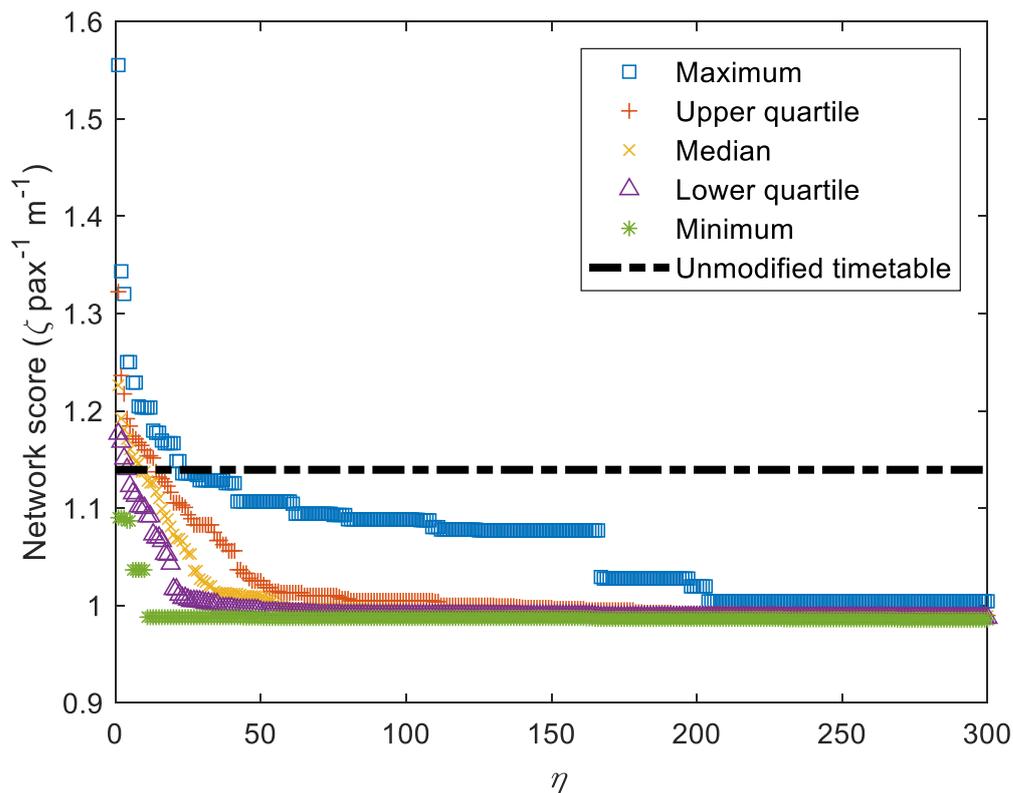


Figure 34 - Markers to show features of the distribution of 'best network score found so far' values at different numbers of objective function evaluations(η). The network score of the unmodified timetable is also shown.

The interval between trains at stations other than Derby is also a consideration for passengers. To investigate the effect of capturing this an additional passenger journey stage is introduced to PRaM which captures the 'schedule delay', i.e. the difference between the passenger's 'ideal' departure time and the actual departure time of the soonest train after this (Balcombe et al., 2004). This approach has precedent - Hendrickson and Kocur (1981) have also examined timetabling decisions including the concept of schedule delay. The term 'schedule delay' also relates to the time difference between a passenger's actual arrival time and their 'ideal' arrival time at their destination, if they catch a train which gets them to their destination 'early' (Small, 1982). Passenger agents in the newly introduced stage are modelled as outside of the network and their movements are not considered. To distinguish between the two meanings and highlight that the introduced stage occurs before departure, rather than at the arrival station, the newly introduced journey stage is referred to as the 'Spare Time' stage. Figure 35 illustrates how the 'Spare Time'

stage relates to the other stages described in Chapter 4 and waiting time described in Section 6.1.3. Passengers begin the 'Spare Time' stage at their 'ideal' departure time (assigned when determining λ and described in Section 6.1.3), and finish the 'Spare Time' stage when PB1 occurs (described in Chapter 4). Conditions are not captured during the 'Spare Time' stage. Waiting time describes the time from when passengers enter the station to catching their first train.

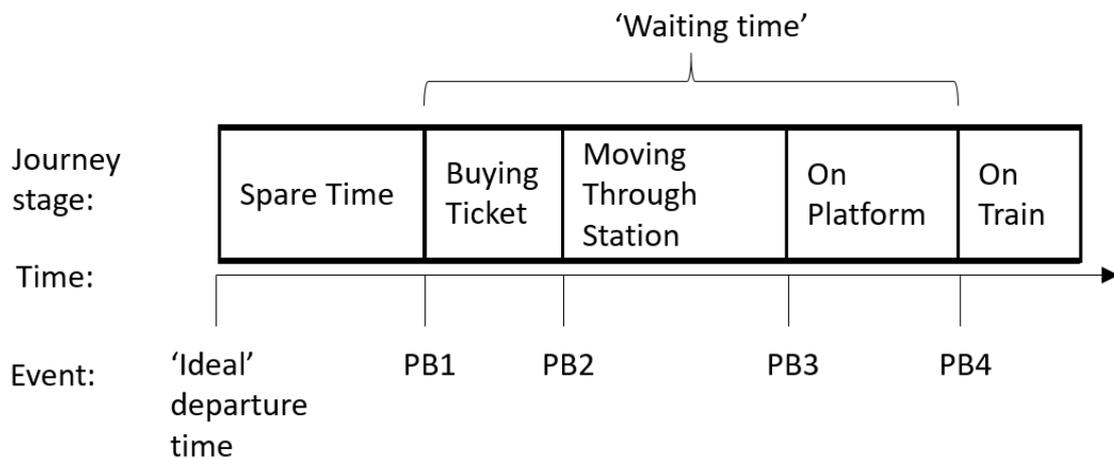


Figure 35 – A sequence of journey stages for a partial journey when the 'Spare Time' stage is introduced. The time at which different events occur, including the passenger behaviours enumerated with 'PB' and described in Chapter 4, and corresponding journey stages are shown inside boxes. The journey stages included in the 'waiting time' described in Section 6.1.3 are shown.

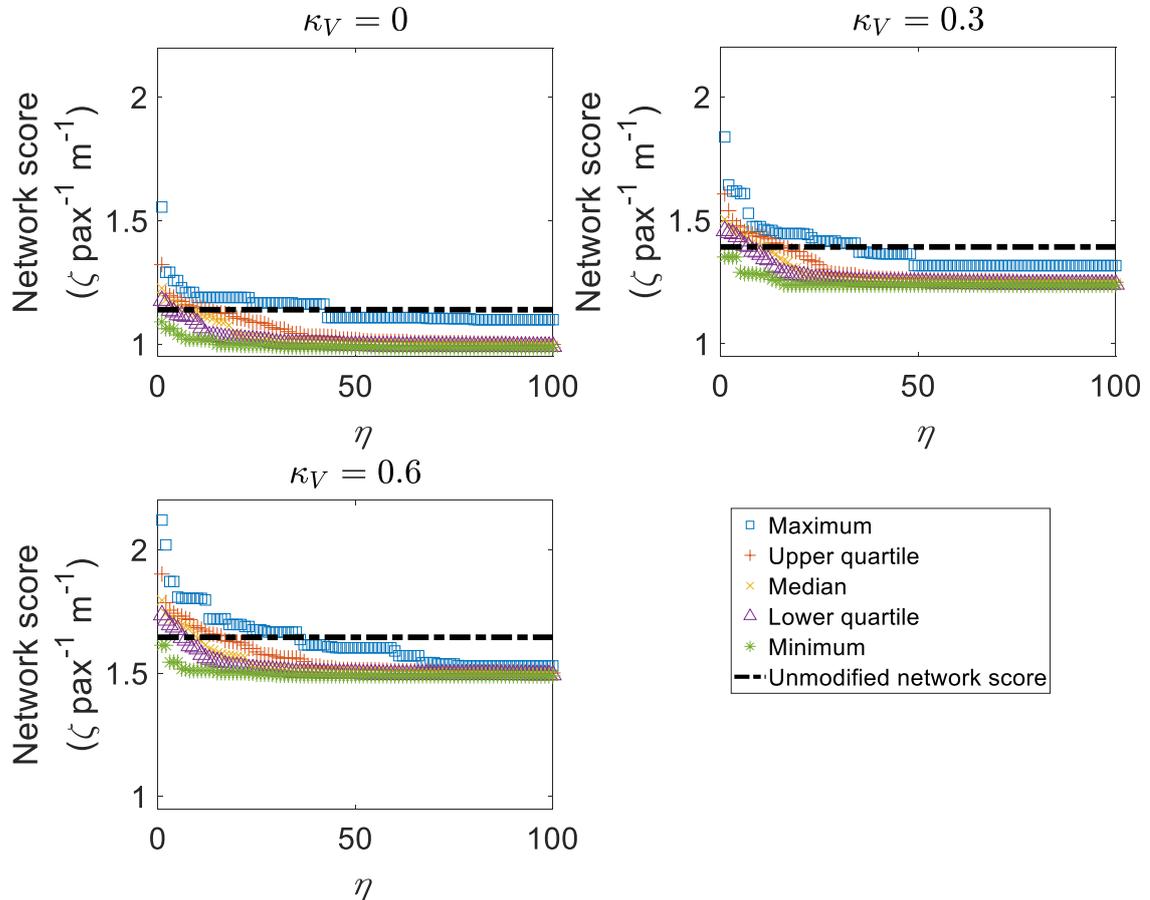
No data could be found relating to the VoT weighting (Ω) during this stage, therefore its value is set to the experimental parameter, κ_V . The value of κ_V is varied between experiments to investigate the sensitivity of the results to this parameter. A range of 0 to 0.6 was chosen for κ_V , to reflect the assumption that the disutility of schedule delay to passengers is less than travelling on an uncrowded train because the time may be used for work or leisure activities as the passenger chooses. In the case that this assumption is incorrect, the experiments can easily be repeated with an alternative value of κ_V .

Figure 36 displays plots using markers to show features of the distribution of 'minimum network score found so-far' values and the network score of the unmodified timetable. The value of κ_V used for the experiments is shown above each plot. Comparing the plots it can be seen that greater κ_V values cause an

increase in the network scores. This is because the journey score associated with the 'Spare Time' stage is proportional to κ_V , which affect the total journey score of the passenger. For the results relating to κ_V values of 0, 0.3 and 0.6, the median marker at $\eta = 100$ is respectively 0.146, 0.148 and 0.149 $\zeta \text{ pax}^{-1} \text{ m}^{-1}$ less than the network score for the unmodified timetable. This indicates that as the value of κ_V increases, i.e. passengers value their time more, there are more benefits to be gained by optimising the timetable. This is what would be intuitively expected and is evidence to suggest the introduction of the 'Spare Time' stage is valid. However, despite increasing in absolute magnitude, the improvement in network score through optimisation becomes respectively 12%, 11% and 9% of the network score for the unmodified timetable, i.e. a decreasing proportion of the network score. This indicates that, at greater values of κ_V , the network score is less sensitive to the arrival and departure times of trains than it is to the frequency of trains. Therefore, increasing the number of trains will have more substantial benefit than optimising the interval between them. Observing all the plots it can be seen that when η is greater than approximately 50, the location of the markers for the upper quartile and the minimum become indistinguishable, indicating that at least 75% of the optimisations have found a solutions 'close' to the best solution. However for the $\kappa_V = 0$ results the value of maximum marker at $\eta = 100$ is 0.11 $\text{pax}^{-1} \text{ m}^{-1}$ greater than the minimum marker, and is distinguishable in the plot, indicating a substantial difference between the performance of the solution found by the best-case and worst-case optimisation. This difference decreases to 0.08 and 0.04 $\text{pax}^{-1} \text{ m}^{-1}$ for $\kappa_V = 0.3$ and 0.6 respectively and indicates that the BO implementation becomes more consistent for greater values of κ_V . This might be because optimisation algorithm implementation is less easily 'trapped' by local minima since they are less 'deep'. Becoming trapped in areas of the

search space that do not contain the global optimum is further discussed in Chapter 2.

Figure 36 - Three plots displaying markers for the distribution, from 40 repeats of the experiment, of 'minimum network score found so-far' values and the network score of the unmodified timetable. Each plot relates to a different VoT



weighting used for the 'Spare Time' journey stage, κ_V , with value shown above the plot. The number of objective function evaluations, η , so-far is shown on the x-axis.

The results of this case study indicate that the SUPREME framework can be used to support improving the timetable. The results of Figure 31 show that the BO implementation can consistently locate 'good' solutions. Compared to the unmodified timetable, the best timetable identified gives a 13% improvement in the network score, indicating that the VoT for the mean journey was reduced by 13% and therefore passenger experience was improved. Bruzelius (1981) describes the 'generalised cost of travel' to a passenger as the sum of the ticket price and VoT for the journey, i.e. ψ in Chapter 3. Assuming this simple model, if the VoT is reduced, passengers may be willing to pay more for a ticket whilst maintaining the same 'generalised cost of travel'. To calculate the potential

extra ticket revenue: the VoT for commuter passengers provided by ARUP (2015) is reduced by 13%, multiplied by the median journey duration for the unmodified timetable, and multiplied by the number of passengers in θ - giving a potential ticket revenue increase of £10,000 per day for this network. Commuter passengers are focussed upon here for simplicity because the VoT for business travellers varies with journey distance, however passenger journey purpose has not been included in the case study so the journey VoT for passengers could not be determined. Although Peluffo (2018) reports 57% of rail passengers in 2017 were commuters, a substantial proportion were not and therefore had a different VoT. Consequently, there is uncertainty in the potential extra ticket revenue value. Nonetheless, it demonstrates that substantial extra revenue might be generated and considering that the best timetable identified requires no additional physical resources, implementing it might require little investment from network managers therefore giving it a strong business case. Furthermore, for the purposes of this case study a maximum of 100 objective function evaluations have been used but in practice, because timetabling is often an activity with a long planning window, it may be possible for users to increase the number of objective function evaluations and possibly find better solutions. The results of Figure 32 show how the output of the SUPREME framework can be used to demonstrate general trends as well as supplying the best solution. Many tasks abstract the real world challenge in some way, therefore supplying the human manager with more information will allow them to choose solutions which are sufficiently robust to transfer back to the real world. The results of Figure 33 demonstrate that network changes can have knock-on effects causing many passengers to be positively affected and improving the experience of the small number of 'very bad' journeys. This further supports the case for a methodology which models passengers and components individually, to ensure that knock-on effects are captured. This case study has also shown that, to better capture or assist with more challenges, the SUPREME framework can be extended to include other factors, i.e. in this case the effect captured by the 'Spare Time' journey stage. The results of Figure 36, show that optimising the timetable will have greater

absolute effect on the network score if the interval between trains is considered in the assessment, indicating that the SUPREME framework is more useful when more factors of the network performance are considered. However, because the results also show that optimising the timetable has less relative effect on the network score for larger values of κ_V , the results also indicate that reducing the VoT weightings for journey stages is also important for improving the network score. Network managers might consider the two approaches simultaneously since, generally, reducing VoT weightings would require investment to improve rolling stock or station comfort whereas optimising the timetable can be comparatively cheap if no additional resources are required. Although a sensitivity study has been used here to mitigate the uncertainty of unknown parameters, this case study has identified that obtaining a VoT weighting for the 'Spare Time' stage is an area for development of the SUPREME framework.

6.2 Case study two – optimising the allocation of attributes to trains in the regional network around Derby

This case study considers the same physical network as the first case study, however a different optimisation task is developed to address additional challenges identified for the network: allocating additional rolling stock, maintaining different levels of train cleanliness, and rolling stock defects. Both Furness (2017) and The Friends of Derwent Valley Line (2019) report that the overcrowding challenge described in Section 6.1 is sometimes addressed by allocating additional rolling stock to the trains. Consequently the task in this case study considers the optimum allocation of additional rolling stock between a fixed number of trains. This is operationally feasible because the network is serviced by BR Class 153 rolling stock, which can operate in formations of multiple units or with other rolling stock types such as BR Class 156 (Haigh, 2018). The cleanliness of trains is a challenge for many GB networks, for example it is identified by the Department for Transport (2017a) for the East Midlands franchise and Network Rail (2016) for the GB network, and so is considered for this case-study network also. In the UK passenger trains are

typically cleaned every night (Watchdog, 2017) however, events such as industrial action might mean that not all trains are cleaned (n.a, 2016, RMT, 2017) and cleanliness deteriorates during use throughout the day. In an interview conducted on 15/6/2017, the Customer Experience Strategy Manager for EMT stated that determining which train to allocate less recently cleaned rolling stock to, for the minimum effect to passengers, was a challenge (Emma Davies, 2017)⁸. Consequently, the task in this case study also considers the optimum allocation of cleaned and less-recently-cleaned rolling stock. For many networks, component defects in rolling stock is a challenge which can cause can cause substantial delay. For example, Railnews (n.a, 2018) report on door failures causing a delay of up to 45 minutes to a train. Ideally rolling stock with component defects would be removed from service and replaced without affecting the service. However, owing to operational constraints, this is not always feasible and the UK's Rail Safety and Standards Board (2015) state that, depending on the nature and severity of the defect as well as other situational considerations, removing a train from service may be more dangerous to passengers than keeping it in service with the component defect. Consequently, for minor defects, the rolling stock may remain in service until repair or replacement can be scheduled. Dinmohammadi et al. (2016) find that for a fleet of 38 BR Class 380 trains there were 205 door defects, ranging in severity, which caused 518 minutes of train delay. Although Dinmohammadi et al. do not specify the time period over which this data was collected, from other information they report it can be inferred to be approximately 100 days which would mean a defect rate of approximately one defect per train every twenty days. Clearly the majority of these do not cause the train to be removed from service, but it is reasonable to assume that they might cause the train to have longer door opening and closing times. If defects are detected prior to a train entering service, it might be operationally feasible to allocate rolling stock with defects to trains such that the effect to passengers is minimised. Therefore the task discussed here also optimises the allocation of rolling stock

⁸ For more information, see the list of personal communications in the reference list.

with door defects to the trains. The task considers all the challenges concurrently and optimises which trains are allocated additional rolling stock, which trains are allocated less-recently-cleaned rolling stock and which train is allocated the defective rolling stock.

6.2.1 Formal definition

The task described in the previous section can be considered as one of optimising the allocation of five attributes between six trains. There are three unique attributes, but some are allocated to two trains. Table 12 displays how these attributes are enumerated with the index, i .

Attribute, i	Description
1	Train has an additional rolling stock unit
2	Train has an additional rolling stock unit
3	Train is less-recently cleaned
4	Train is less-recently cleaned
5	One rolling stock unit of the train has door failure

Table 12 – The five attributes to be allocated to the trains.

The vector x captures which train each attribute is allocated to with form shown by:

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5] \quad (6.6)$$

where the element x_i denotes the train that attribute i is allocated to. The formal definition of the task can therefore be written as (6.7) subject to (6.8) and (6.9) which captures the fact that there are two separate less-recently-cleaned trains.

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in X} F_2(\mathbf{x}; \lambda, \theta) \quad (6.7)$$

$$1 \leq x_i \leq 6 \quad \text{for } 1 \leq i \leq 5 \quad (6.8)$$

$$x_3 \neq x_4 \quad (6.9)$$

6.2.2 Data input and experimental method

The model network has the same topography and permissible line speeds described in Section 6.1.2. The timetable has hour intervals between the trains

and uses interstation run times provided by East Midlands Trains (2019). Trains 1 and 4 are the first trains to depart on their route, followed by Trains 2 and 5 and finally Trains 3 and 6. Trains that are not allocated extra rolling stock units have the performance characteristics of a BR Class 153 unit, described in Section 6.1.2. Trains that are allocated extra rolling stock have scaled parameters relating to their tare mass, passenger standing and seating capacity, number of doors, and for simplicity, the available tractive effort and Davis formula coefficients. These parameters are scaled directly proportional to the number of rolling stock allocated to the train. Trains that are allocated to have faulty doors have the value of the door parameter (C_0) in formula (4.7) multiplied by the experimental factor, κ_D . The value of the VoT weighting function (Ω) for passengers in the 'On Train' stage is multiplied by the experimental factor, κ_Ω , when riding on a less-recently-cleaned train. The passenger load described in Section 6.1.3 is also used for this case study. The same experimental method as described in Section 6.1.4 is used, however, to reflect that in practice this task might need to be carried out with a short time window, e.g. on the morning of operation, the budget of objective function evaluations (η_{max}) is reduced to 30. Because the task is sensitive to the value of κ_D and κ_Ω these are respectively varied from 2 to 32 and 1.5 to 2.5.

6.2.3 Results

Figure 37 uses markers to display features of the distribution, from 40 repeats of the optimisation, of the 'minimum network score found so-far' values at every objective function evaluation (η). For these optimisations, κ_D and κ_Ω were respectively 8 and 2, i.e. the median of the values investigated. The network scores in the distribution at $\eta = 1$ relate to 40 random samples of the search space. Therefore, comparing the features of the distribution at larger values of η against their value at $\eta = 1$ allows comparison of the BO implementation against random sampling of the search space. Observing Figure 37 it can be seen that the maximum marker at $\eta = 6$, is less than the median marker at $\eta = 1$. This indicates that the 'worst case scenario' of using the BO implementation for six objective function evaluations, found a solution better

than the median of 40 random selections. Similarly the results indicate that using 23 objective function evaluations, 75% of the optimisations have found a solution at least as good as the best one found in 40 random selections. Furthermore, the upper quartile and minimum marker at $\eta = 25$ are within 8% of each other, this indicates that 75% of the optimisations find a solution ‘close’ to the best solution. Since all the optimisations found a better network score than the one relating to their initial candidate and the median improvement in network score is 32% with standard deviation of 14, these results indicate that the BO implementation is consistently finding ‘good’ solutions with an efficient use of objective function evaluations.

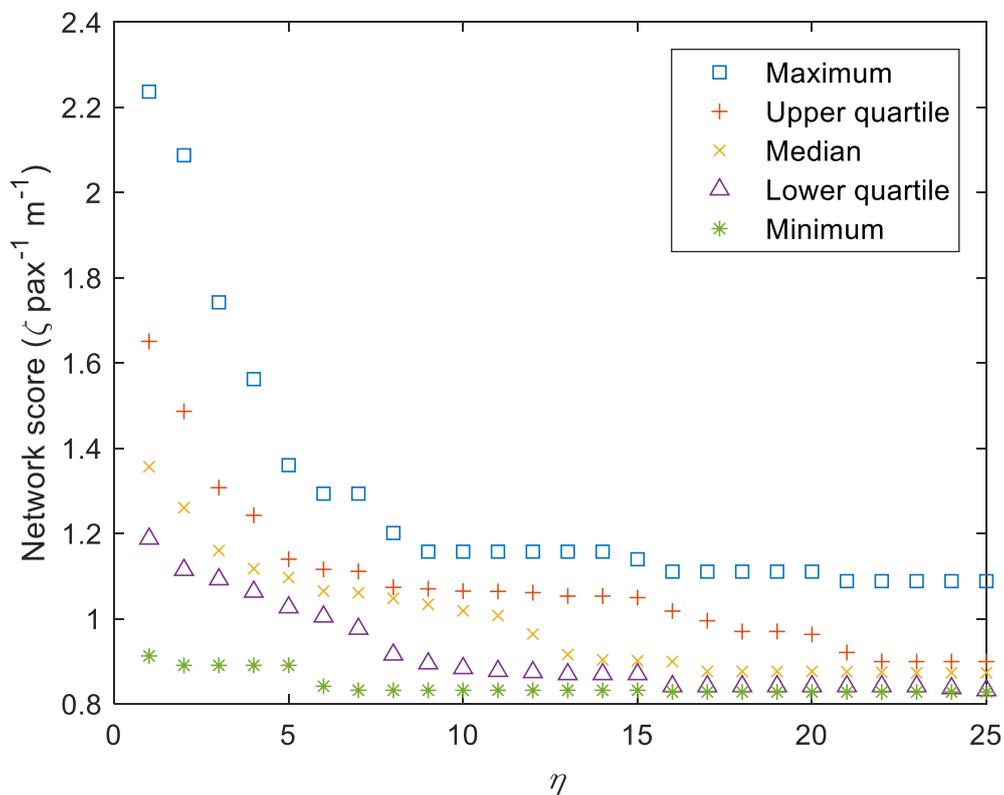


Figure 37 - Markers to show the value of features of the distribution, from the distribution of 40 optimisations, of ‘minimum network score found so-far’ values at different numbers of objective function evaluations, η . The values of κ_D and κ_Ω were respectively 8 and 2.

Figure 38 plots the frequency that each attribute was allocated to a certain train in the distribution of x relating to ‘minimum network score found so-far’ values at $\eta = 1$ and $\eta = 25$. Each column represents a train and attribute pairing, with its height representing the frequency of that allocation in the distribution.

Trains with ID numbers one to three operated on the Crewe-Derby route and ID numbers four to five operated on the Nottingham-Matlock route. To further investigate whether the solution found is dependent on the initial candidate, for all the optimisations, no attributes were initially allocated to Train 6 and consequently the columns relating to Train 6 are obscured from view in the $\eta = 1$ plot. However the results in the $\eta = 25$ plot show that the BO implementation did find solutions where Train 6 was allocated attributes, indicating that the BO implementation was able to 'discover' this region of the search space.

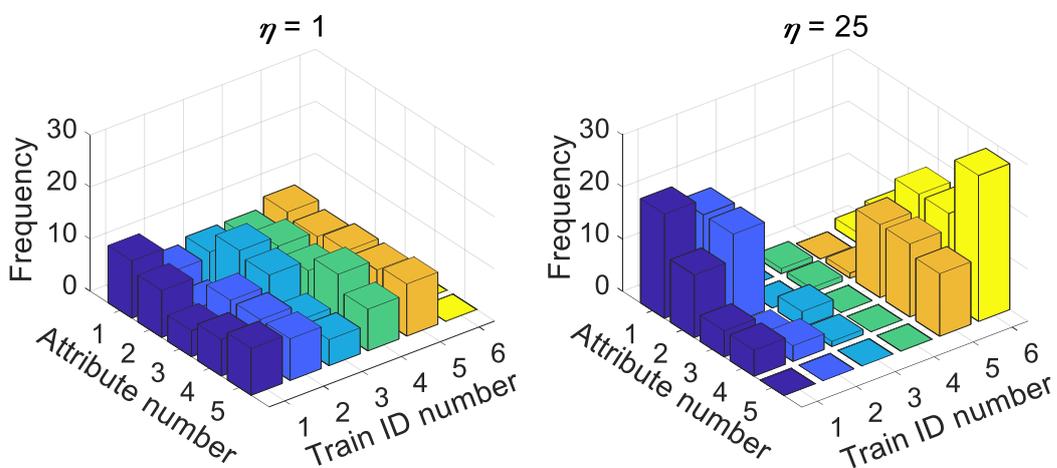


Figure 38 - From a distribution of 40 solutions, the frequency that each attribute was allocated to different trains at different numbers objective function evaluations, η .

Excluding Train 6 in Figure 38, it can be seen that the attributes are initially allocated uniformly between the trains indicating that for all other trains, the initial allocation code was working correctly. In comparison, the $\eta = 25$ plot shows that the solutions in this distribution have clear grouping of attributes and train numbers which, by observing the results in Figure 37, corresponds with better performing solutions. For example, 63% of solutions in the $\eta = 25$ distribution have Attribute 1 and 2 allocated to Train 1 and 2 indicating that additional rolling stock should be allocated to the Crewe-Derby route. Similarly, 60% of solutions allocated Attributes 3, 4 and 5 all to Trains 5 and 6, indicating that these are the least critical trains in the network. Trains 3 and 4 rarely have attributes allocated to them, this indicates that they are 'mid-critical' trains who do not necessarily need additional resources but should

have the standard of their service maintained. For a given route, trains with smaller ID numbers are more 'critical' to the network score than trains with greater ID numbers because the order in which they begin operation means they are in-service for longer and hence carry more passengers.

To investigate the sensitivity of the results to the parameters κ_D and κ_Ω , Figure 39 displays graphs of the results from experiments where these have been set to the maximum and minimum values used. Similar to Figure 37, each graph displays markers describing the distribution of 'minimum network score found so-far' values at different algorithm iterations. The values of κ_D and κ_Ω are shown above each plot. For all plots of Figure 39, the value of the distribution features marked and the spread of the distribution, reduces at greater values of η . This is consistent with the results of Figure 37 and indicates that the BO implementation is able to find 'good' solutions regardless of the parameters defining the optimisation task. Observing Figure 39 shows that in general, for a given feature of the distribution and value of η , the network score value is greater for experiments with greater κ_D and κ_Ω values. This is expected because increasing dwell time will decrease passenger disutility and penalising dirty trains more heavily will also result in a greater, i.e. worse, network score. The results also show that for low values of η , the distribution features become increasingly more sensitive to κ_D and κ_Ω in order from the distribution minimum to the maximum. For example, comparing values between the results for $\eta = 1$ for $\kappa_D = 2$ and $\kappa_\Omega = 1.5$ with $\kappa_D = 32$ and $\kappa_\Omega = 2.5$, the increase in the network score relating to the minimum marker, is 20% and the increase in the network score relating to the maximum marker is 200%. This indicates that 'good' solutions are less sensitive to κ_D and κ_Ω as would be expected since their negative effects have been minimised. This means that optimising resource allocation can significantly reduce the effect of faulty rolling stock and less-recently-cleaned trains upon network performance.

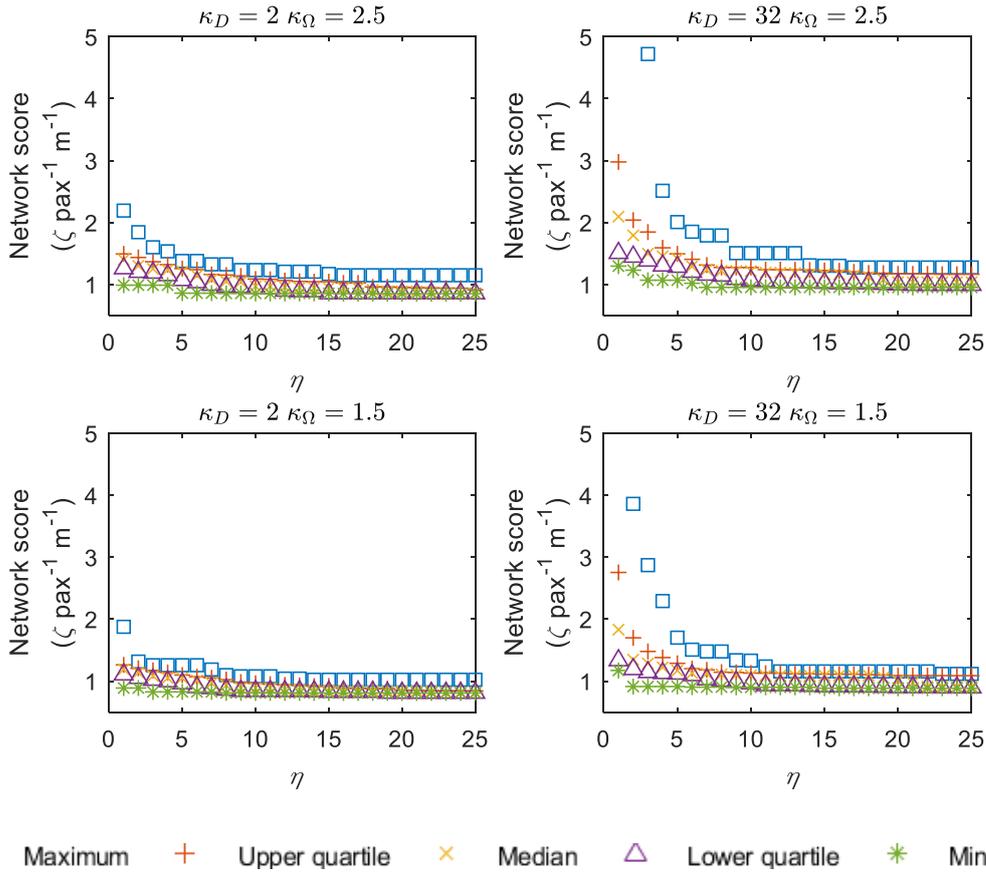


Figure 39 – Markers to features of the distribution of ‘minimum score found so far’ values at different numbers of algorithm iterations (η). The distribution is displayed four different experimental parameter sets. Each plot shows the value of the parameters determining the door opening and closing time of faulty trains, κ_D , and the Value of Time weighting for passengers on a dirty train, κ_Ω .

For the results shown in Figure 39, Figure 40 plots the number of times, in the distribution of ‘minimum network score found so far’ values at $\eta = 25$, each attribute was allocated to a certain train. The plots use the same axes as described for Figure 38 and the value of κ_D and κ_Ω are shown above each plot. Observing Figure 40 it can be seen that for all values of κ_D and κ_Ω there is trend towards allocating Attribute 1 and 2 to Trains 1 and 2 and Attributes 3, 4 and 5 to Trains 5 and 6. Comparing the plots with $\kappa_D = 2$ with $\kappa_D = 32$, it can be seen that the former results occasionally have Attribute 5 allocated to Train 1, 2 or 4 but the latter do not. This suggests that as κ_D is increased, causing longer dwell time for faulty doors, the train allocation of faulty doors, i.e. Attribute 5, becomes more critical. This is what intuitively would be expected and indicates that the task is capturing the real-world challenge well.

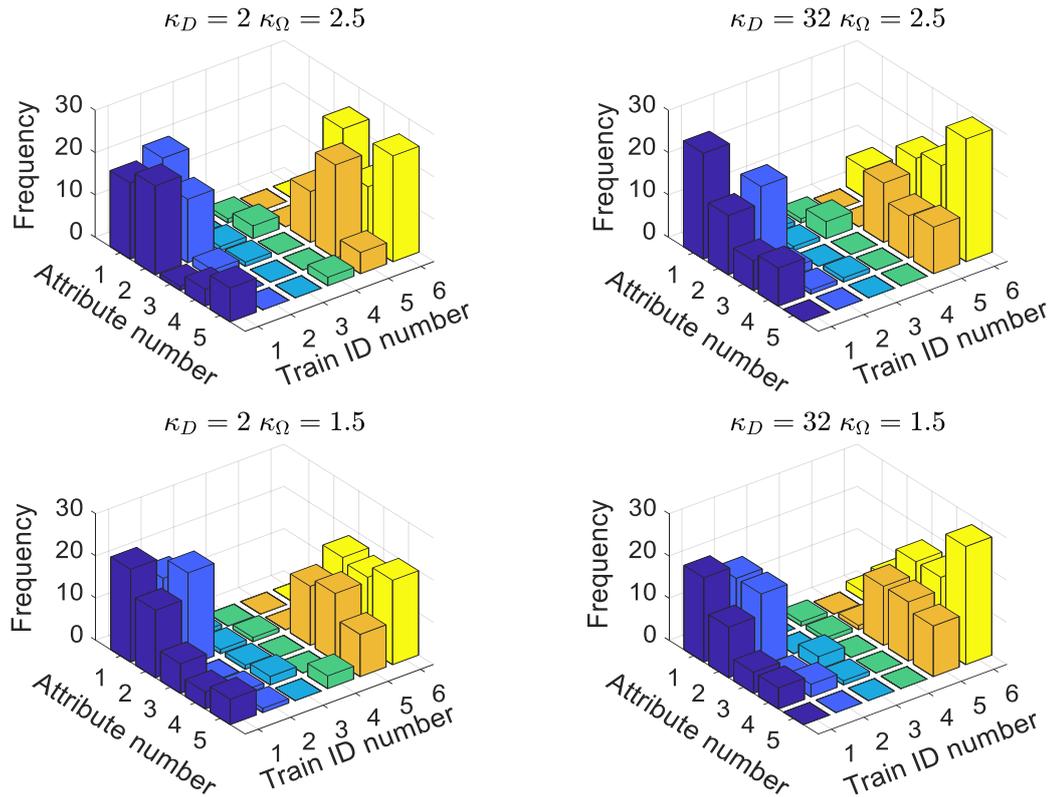


Figure 40 – Four plots indicating the attributes allocated to each train under four combinations of κ_D and κ_Ω . Each plot shows the number of times that each attribute was allocated to different trains, from a distribution of 40 solutions.

This case study has shown how the SUPREME framework can be applied to tasks where resources of different types must be allocated with a short time window, for example on the morning of operation due to unforeseen circumstances. The results of Figure 37 have shown that the SUPREME framework can find ‘good’ solutions regardless of the candidates it is initialised with, meaning that it can be applied quickly without specialist knowledge of the challenge and ‘good’ candidates. During these experiments eight optimisations were computed in parallel on one machine, for which the median time required was one hour of wall clock time, with standard deviation of one minute. To better investigate the performance distribution of the SUPREME framework, in these experiments 40 optimisations have been computed, however, in practice the best of eight would likely still find a ‘good’ solution. Machines with performance characteristics equivalent to the one used for these experiments are available to rail network managers, meaning that the

required computational power and time window might be operationally feasible for the SUPREME framework to be used in practice. If less time is available, the results of Figure 37 indicate that even if the budget of objective function evaluations is reduced to 6, relating to 14 minutes of computation time, the solution found by the SUPREME framework is better than the median of 40 random selections. The results of Figure 38 and Figure 40, could be used by a rail network manager to identify general strategies for improving network performance or maintaining it in unexpected circumstance. For example, increasing the capacity of Train 1 and 2 is of higher priority than increasing the capacity of other trains in the network. Although these experiments have used parameters whose value is estimated, i.e. κ_D and κ_Ω , the results of Figure 39 and Figure 40 indicate that they do not affect the primary findings, therefore the solutions identified are likely to be robust to the real-world value of these parameters.

6.3 Case study three – optimising a value engineering strategy for the High Speed Two network

In the UK, the High Speed Two (HS2) network is under construction to link the cities of London, Birmingham, Leeds and Manchester with high-speed, i.e. >250km/hour (Civity Management Consultants, 2014), rail travel on mostly dedicated, purpose built lines. There is concern that the project might be more expensive than predicted (Calder, 2018) and, although the budget has been increased (Smale, 2019), HS2 managers might need to consider options to reduce the initial cost of the network by reducing investment into certain components, i.e. a *value engineering strategy*. Considering that there are multiple components for which savings could be made, choosing the optimum value engineering strategy could be a challenge for managers. This case study demonstrates how the SUPREME framework might assist with this challenge by applying it to an optimisation task which maximises passenger satisfaction whilst considering a 'minimum total saving' constraint.

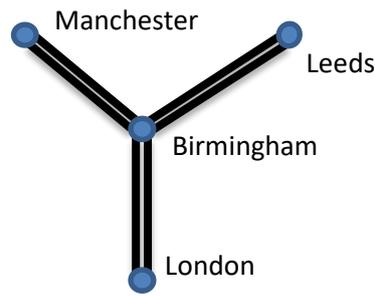


Figure 41 – The topography of the model HS2 network. Stations are represented by circles and the Birmingham station is connected to all other stations by two single direction lines, i.e. an ‘up’ and a ‘down’ line.

The optimisation task models the HS2 network using a ‘Y’ topography, shown in Figure 41 where four stations are represented by circles, and the Birmingham station is connected to all the outer stations by two single direction lines, i.e. an ‘up’ and a ‘down’ line. Components where savings are considered are the trains, lines and stations. Currently, it is planned that the network will use newly designed and built rolling stock that can travel at 360km/hour (HS2 Limited, 2019). This is 40 km/hour faster than the next fastest rolling stock operated in the UK, the BR Class 374 (Siemens Mobility, 2016). Using already designed rolling stock with lower maximum speeds, such as the Class 374, might reduce the initial cost of HS2 but the effects of longer train journey times might also affect passenger experience. The task considered here chooses between four types of rolling stock that might be used on the HS2 network and, by optimising this choice, this task captures the trade-off between train maximum speed and savings. The speed-dependent maximum tractive effort available associated with each rolling stock type is also included. Together with train maximum speed, permissible line speeds can also affect train journey times and hence passenger experience. Civity Management Consultants (2014) describe that although the construction costs of lines is highly sensitive to the terrain, e.g. mountainous or urban, for a given terrain the permissible line speed also affects the construction costs. In this task, the choice of permissible line speed is also optimised with the trade-off against savings. As well as journey times, the comfort of both trains and stations also affects passenger experience, with network managers investing in refurbishing both (Rail Technology Magazine, 2014, Network Rail, c2019) and Preston et al. (2008)

showing that station investment increases demand for travel from passengers. For overall passenger experience, increasing the comfort of stations and trains might offset any increase in journey times, or vice-versa, at a reduced cost. To capture this, the VoT weighting for passengers in the 'On Train' or 'In Station' stage is optimised whilst considering the expense of creating trains and stations with that VoT value. By modelling the network defined by the parameters associated with each value engineering strategy, the resultant network performance is assessed. The 'total savings' associated with each value engineering strategy is also calculated with a *total savings function*, which is then constrained to capture the 'minimum total savings' requirement.

6.3.1 Formal definition

The optimisation task described in the previous section considers nine saving areas, which are shown in Table 13 and enumerated by the index, *i*. All rolling stock in the model are identical therefore savings 1 and 2 apply to all trains. The permissible line speeds of lines connecting each city pair in Figure 41 are optimised together, though each pair is optimised in isolation, and the permissible line speed applies for the whole length of the line, consequently there are three saving areas associated with permissible line speed. The passenger comfort of each station is considered individually.

Saving, <i>i</i>	Description
1	Rolling stock maximum speed and tractive effort available
2	Rolling stock comfort
3	London – Birmingham, Birmingham – London permissible line speed
4	Birmingham – Manchester, Manchester – Birmingham permissible line speed
5	Birmingham – Leeds, Leeds – Birmingham permissible line speed
6	London station comfort
7	Birmingham station comfort
8	Manchester station comfort
9	Leeds station comfort

Table 13 – The nine areas for savings.

For all of the saving areas, the possible 'design options' are considered as a discrete set, where each option is enumerated. The vector *x* captures the design options of each value engineering strategy with form shown by:

$$\mathbf{x} = [x_1, x_2 \dots x_9] \quad (6.10)$$

where the discrete variable, x_i , denotes the design option of the i th saving area listed in Table 13. The formal definition of the task is written as:

$$\mathbf{x}^* = \underset{\mathbf{x} \in X}{\text{arg min}} F_2(\mathbf{x}; \lambda, \theta) \quad (6.11)$$

$$0 \leq x_i \leq 3 \quad \text{for } i = 1 \quad (6.12)$$

$$0 \leq x_i \leq 4 \quad \text{for } i = 2 \quad (6.13)$$

$$0 \leq x_i \leq 13 \quad \text{for } 3 \leq i \leq 5 \quad (6.14)$$

$$-5 \leq x_i \leq 5 \quad \text{for } 6 \leq i \leq 9 \quad (6.15)$$

$$G(\mathbf{x}) \geq \kappa_S \quad (6.16)$$

where $G(\mathbf{x})$ denotes the total saving function and κ_S the minimum total saving required. The constraints (6.12) to (6.15) capture the number of design options available in each saving area, with increasing values of x_i representing increasing savings. The constraint (6.16) captures the minimum total saving requirement with the total savings function given by:

$$G(\mathbf{x}) = G_T(x_1) + G_C(x_2) + \sum_{i=3}^{i=5} G_L(x_i) + \sum_{i=6}^{i=9} G_V(x_j) \quad (6.17)$$

where $G(\mathbf{x})$ denotes the total savings, $G_T(x)$, the saving function relating to rolling stock type, $G_C(x)$, the saving function relating to train comfort, $G_L(x)$, the saving function relating to permissible line speed, $G_V(x)$, the savings function relating to station comfort and i is an index. There are 1.5×10^9 candidates in the search space.

6.3.2 Parameters capturing the fixed network components

The line distance from Birmingham to Leeds is 198km (HS2 Limited, 2016b), London to Birmingham is 225km (Railway Technology, c2019b) and the total for the network is 531km (Railway Technology, c2019b), therefore Birmingham to Manchester is 108km. The number of platforms at each station in the model are provided by HS2 Limited (2016b, c2019), as 11 for London, 5 for Leeds and 4

for Manchester. One station with eleven platforms is modelled to represent Birmingham Curzon Street and Birmingham Parkway with seven (HS2 Limited, c2019b) and four (ARUP, n.d) platforms respectively. A timetable was not available to the author, but interstation run times and trains-per-hour predictions were provided by the Department for Transport (2016, 2017b,p.43) and HS2 Limited (HS2 Limited, c2019b). All services were modelled to stop at Birmingham. The interstation run times were combined with a two minute dwell time (HS2 Limited, 2019), to create a timetable with equal intervals between trains and constant service pattern between 6am and 11pm. 21 identical trains are modelled to fulfil this timetable.

6.3.3 Parameters capturing the variable network components

The four types of rolling stock that train agents can represent are shown in Table 14, together with the corresponding value of x_1 , the values used for each characteristic and the source of data where applicable. The tare mass for the HS2 Design Specification is unknown, however, the value for the Class 374 is used because they are a similar configuration high-speed train (Siemens Mobility, 2016, HS2 Limited, 2019). For the BR Class 395 and BR Class 373, speed dependent $F_{available}$ data was extracted from the BRaVE simulator, described by Umiliacchi (2016). No $F_{available}$ data was available for the Class 374 or the HS2 Design Specification, so these were based on the values for a Class 373. For a Class 374, the Class 373 values were used with the value of $F_{available}$ remaining constant above the maximum speed of the Class 373, i.e. 300km/hour. For the HS2 Design Specification, in order to meet the minimum performance requirements specified by HS2 Limited (2019), the values of $F_{available}$ for the Class 373 were increased by 10%. The value of $F_{available}$ also remained constant above 300km/hour. Appendix III describes how the Davis formula coefficients were determined for the Class 395 and 373 by fitting to speed dependent rolling resistance data extracted from the BraVE simulator. Similarly, no resistance data was available for the Class 374 or HS2 Design Specification so the values for a Class 373 were used. The BraVE simulator provides the value of the rotary allowance (r) as 7.8 for the Class 395 and 373,

this is also applied for the Class 374 and the HS2 Design Specification. The braking performance used in Chapter 4 was applied for all trains. To represent the case study hypothesis that, rolling stock already in manufacture might be long-formed or adapted to meet HS2 requirements, all rolling stock were modelled to have 700 pax seating capacity and 250 pax standing capacity, i.e. the high-density capacity requirement set by HS2 Limited (2019). For trains with smaller passenger capacity, e.g. the Class 395 has a seating capacity of 340 pax (Mochida et al., 2010), the extra mass of these additional passengers would cause up to a 16% change in the effective mass of the train. The sensitivity study in Chapter 4 has shown that uncertainties of this magnitude in mass are acceptable. This is also true for uncertainty introduced by estimates of r values used for the HS2 Design Specification and the Class 374, as well as the estimate of M_T for the HS2 Design Specification.

$x_1 =$	0	1	2	3
Train type	HS2 Design Specification	BR Class 374	BR Class 373	BR Class 395
Maximum speed, v_{max} , (km/hour)	360 (HS2 Limited, 2019)	320 (Siemens Mobility, 2016)	300 (Railway Museum, n.d)	225 (Mochida et al., 2010)
Tare mass, M_T (tonnes)	970	970 (Clinnick, 2018)	665 (BraVE simulator)	300 (BraVE simulator)
$F_{available}$	110% x Class 373	Class 373	Class 373 (BraVE simulator)	Class 395 (BraVE simulator)
Davis formula coefficients	Class 373	Class 373	Class 373 (BraVE simulator)	Class 395 (BraVE simulator)
Rotary allowance, r	7.8	7.8	7.8 (BraVE simulator)	7.8 (BraVE simulator)
Passenger capacity	950	950	950	950

Table 14 - The maximum speed associated with each value of the optimisation variable, x_1 . The data source is shown for each train.

To reflect the planned maximum permissible line speed (HS2 Limited, 2016b) as well as the maximum speed of the fastest train and slowest train considered, in the model the permissible line speed can be set in the range of 360km/hour to 230km/hour with discrete steps of 10km/hour. Table 15 shows the permissible line speed associated with the value of x_3 , x_4 or x_5 .

x_3, x_4, x_5	0	1	2	...	12	13
Permissible line speed, (km/hour)	360	350	340	...	240	230

Table 15 - The permissible line speed associated with the value of the optimisation variables x_3 , x_4 and x_5 .

To capture changes to the comfort of trains and stations, the VoT weighting (Ω) is modified for passengers that are in the 'On Train' or 'In Station' stage. Data presented by Wardman and Whelan (1998) shows up to 7% variation in the VoT for passengers travelling on different rolling stock, relating to a Class 442 and Mk2 carriage. Making the assumption that this variation could be increased if greater emphasis was put on designing rolling stock to be more comfortable, in this investigation Ω is varied by $\pm 5\%$ with 1% increments and 0% relating to $x = 0$.

6.3.4 Parameters capturing the passenger load

A total passenger load of 300,000 passengers a day is provided by HS2 Limited (HS2 Limited, c2019a). However, this value has been scaled to 117,000 to reflect that only 21 of the 54 trains planned for HS2 are being modelled. One reason for this discrepancy in train numbers is HS2 Limited are purchasing rolling stock for services that travel outside of the model network, e.g. via Sheffield, and these are not included in the creation of the model timetable. Predicted passenger flows for the HS2 network, and current flows between the cities it serves, were not available to the author. Consequently, the relative passenger flows were determined by scaling the train flows in the model timetable and are shown in Table 16. The origin station of passengers relates to the row of the table, the destination relates to the column. It can be seen that between each station pair, the flows in either direction are equal. As described in Section 6.1.3, for each flow, a passenger arrival time to their origin station is set to capture hourly variation in demand for travel and that passengers plan their origin station arrival time considering the departure time of trains.

	London	Birmingham	Manchester	Leeds
London	0	14584	11667	11667
Birmingham	14584	0	8750	8750
Manchester	11667	8750	0	2917
Leeds	11667	8750	2917	0

Table 16 – The modelled daily passenger flows on the HS2 network between cities. The row represents the origin stations, the column represents the destination station, the values represent number of passengers. The total passenger load is 117,000.

6.3.5 The savings functions

The values of $G_T(x)$ over the domain of x_1 are shown in Table 17 and capture the saving for the full fleet. These values are speculative because rolling stock prices represent a commercial package dependent on finance and long-term service agreements, however, the values used do allow demonstration of the SUPREME framework. Appendix III describes how these have been determined by considering the cost of each train type in 2019 prices, and then scaling it to reflect the number of carriages needed to form a 950 capacity train. The saving is then calculated by comparing the scaled cost of the train against the budget for the planned HS2 Design Specification.

x_1	0	1	2	3
Train maximum speed and tractive force equivalent to:	HS2 Design Specification (HS2 Limited, 2019)	BR Class 374	BR Class 373	BR Class 395
$G(x_1)$ (£ millions)	0	320	630	1580

Table 17 – The value of $G_T(x_1)$ over the domain of x_1 and the train type it relates to.

The savings functions for train comfort, permissible line speed and station comfort are respectively given by:

$$G_C(x_2) = m_C x_2 \quad (6.18)$$

$$G_L(x_{3,4,5}) = m_L x_{3,4,5} \quad (6.19)$$

$$G_V(x_{6,7,8,9}) = m_V x_{6,7,8,9} \quad (6.20)$$

where m_C , m_L and m_V are parameters obtained by fitting to observed data, and have values shown in Table 18. These parameters describe the relationship between the savings associated with a component performance or comfort, and the value of the integer variable, x_i , enumerating this component type. In the case of m_L , the value is shown per kilometre and must be multiplied by the length of the line being considered. To determine the value of m_L , no data relating to different permissible line speeds and construction costs, of the HS2 network, is available to the author. However, data provided by Bodman (2012) is scaled to 2019 prices then used to compare the construction costs and permissible line speeds of the High Speed One network, which is also a high-speed network operating out of London, with those planned for the first phase of HS2. To determine the value of m_C , no data was available to the author directly relating HS2 train cost and comfort. However, data reported by Railway Technology (c2019a) indicates that the Class 373 have been refurbished at £12 million per train in 2019 prices⁹. This cost is assumed to relate to changing the comfort from 'poor' to 'average' and be reflected by $x_2 = 5$ and $x_2 = 0$ respectively. Consequently, when scaled to 54 trains, a value of £108 million is determined for m_C . Owing to a lack of data available to the author which directly related HS2 station cost and comfort, a similar approach is taken for determining the value of m_V . Using data provided by the Rail Delivery Group (2017), the mean cost of refurbishment for six stations is £139 million in 2019 prices⁹. Again assuming this relates to changing the comfort from 'poor' to 'average', i.e. from $x_{6,7,8,9} = 5$ to $x_{6,7,8,9} = 0$, gives a value of $m_V = £28$ million. Although there is uncertainty for the value of m_C , m_L and m_V , their accuracy is sufficient for the purpose of demonstrating how the SUPREME framework can be used to support an optimisation challenge of this type. HS2 Limited (2016a) have conducted cost-benefit analyses under different spending options, so it is likely that data to reduce the uncertainty and possibly improve the accuracy of these values exists even if it is not in the public domain. If this data were to be made available the optimisation task could easily be repeated

⁹ Converted to 2019 prices using an inflation calculator provided by Alioth Finance (c2019).

with updated parameters. Applying the parameters listed in Table 17 and Table 18 to (6.17), and setting x_1 to x_9 to their maximum values listed in (6.12) to (6.15), the maximum possible saving is calculated to be £6 billion. This maximum total saving relates to a network with the slowest train and line speeds, and lowest train and station comforts, included in this investigation. It relates to the boundary of the search space and is defined by the upper bounds on the optimisation variables, x_i .

Parameter	m_C	m_L	m_V
Value (£ millions)	108	0.58 per km	28

Table 18 - Saving function parameter values.

6.3.6 Experimental method

Using the same method as described by Section 6.1.4, solutions to the task defined by (6.10) to (6.16) were found using a value of 100 for η_{max} . The optimisation was repeated independently 24 times with different, uniformly distributed, initial feasible candidates. The parameter κ_S was set to 5000 to represent a minimum total savings requirement of £5 billion. This value relates to a constraint on the feasible region of the search space, rather than the £6 billion value which relates to the upper bounds of the optimisation variables. The value of κ_S was chosen to demonstrate SUPREME making a substantial saving, whilst allowing the feasible region of the search space to be large enough to present a meaningful optimisation task.

6.3.7 Results

Figure 42 plots markers to display features of the distribution, from 24 repeats of the optimisation, of the 'minimum network score found so-far' values at every algorithm iteration, counted by η . It can be seen that the network scores for this network are an order of magnitude less than the scores for the network investigated in Sections 6.1 and 6.2. This is expected because of the higher speeds encountered in this network and the distance normalisation applied when calculating the network score, described by Chapter 3, as well as

the greater frequency of trains. So that the effect of the value engineering strategies can be compared against a 'baseline' case, Figure 42 also shows the network score when no savings are applied. From $\eta = 50$ all the optimisations found a solution better than when no savings are applied because they choose to improve the comfort of the trains. It can be seen that, in general, features of the distribution reduce in value at greater values of η and the range of the distribution becomes less. The network score of the best and worst-case optimisation at $\eta = 100$, is respectively 8% and 4% less, i.e. better, than the network score when no savings are applied.

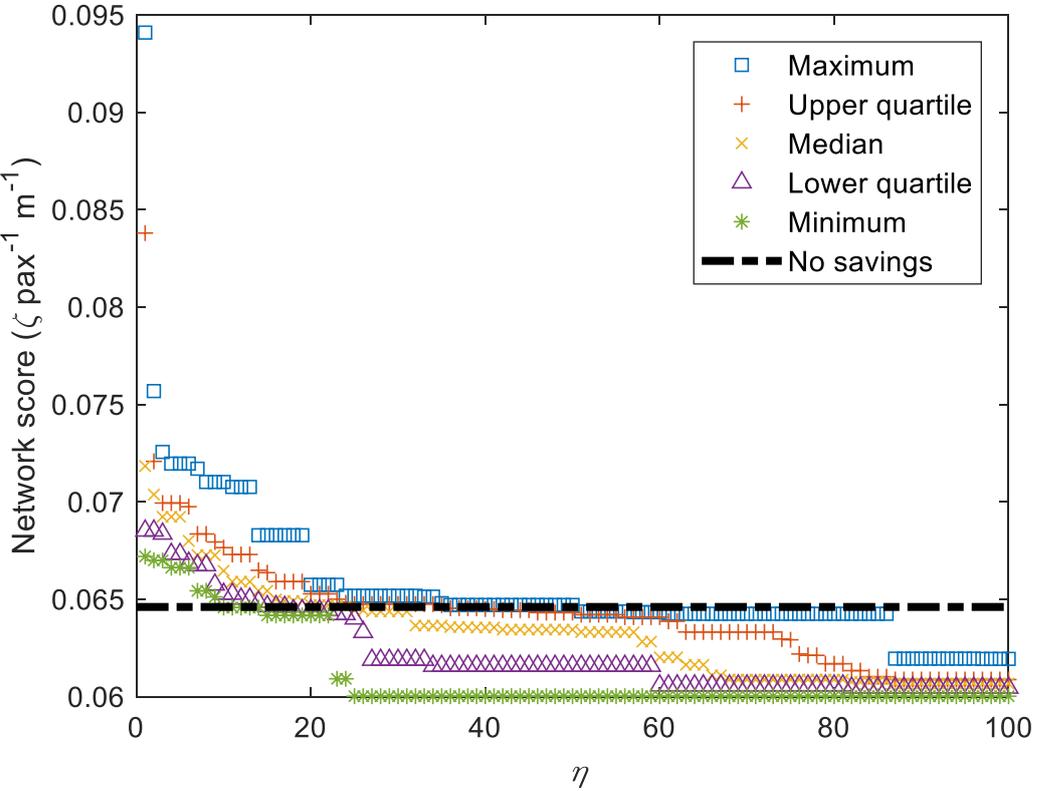


Figure 42 – Markers to show the value of features of the distribution, from 24 optimisations, of 'minimum network scores found so-far' values at different numbers of objective function evaluations (η).

For the 'minimum network scores found so-far' distribution from 24 repeat optimisations, Figure 43 plots the number of times that the vector (x) contained an optimisation variable (x_i) at a certain value. There is a plot for every optimisation variable, with the value of the index (i) shown above each plot. The value of x_i is shown on the x-axis, the number of times it occurred in

the distribution is shown by the y-axis. The results are plotted for the distribution at $\eta = 1$ and $\eta = 100$ and bars for these are plotted next to each other to allow comparison. Observing the plot for rolling stock maximum speed and tractive effort, it can be seen that most initial candidates have $x_1 = 3$, this is because the proportion of candidates that are feasible increases with the value of x_1 and the initial candidates are selected with uniform probability from the pool of feasible candidates. However, at $\eta = 100$ all the optimisations select $x_1 = 2$, as well as indicating the optimum maximum speed of the train, this suggests the BO implementation is successfully exploring all regions of the search space.

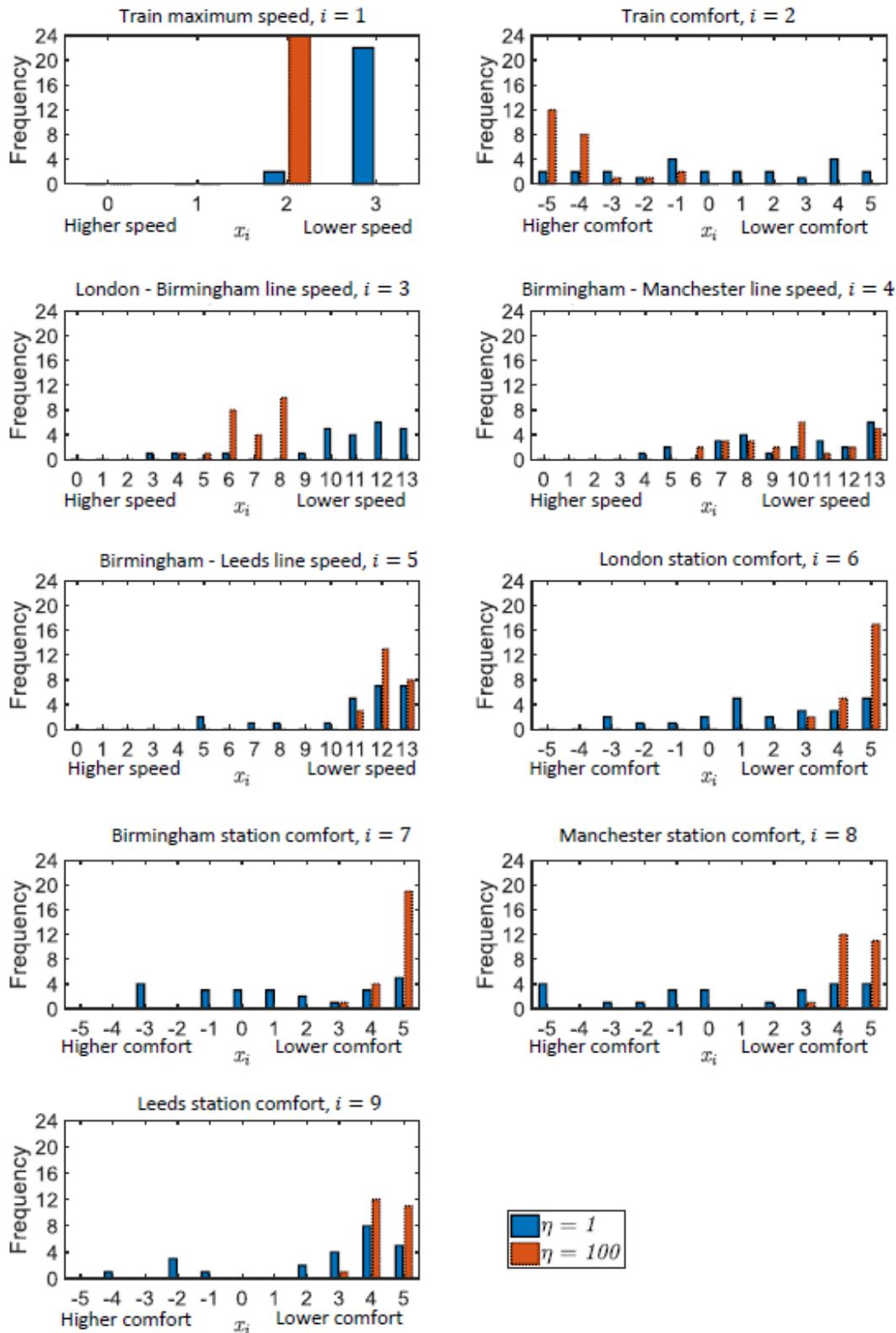


Figure 43 – The frequency that each optimisation variable took different values. The values the optimisation variable took, x_i , are shown on the x-axis. The number of times there was a vector (x) with the variable at the value shown by the x-axis, is shown by the y-axis. For every value of x_i , the values shown at the first algorithm iteration ($\eta = 1$) and last algorithm iteration ($\eta = 100$) are plotted next to each other for comparison.

For the rolling stock comfort plot in Figure 43, the variation in frequency of the initial values of x_2 is within the bounds of what would be expected from a small sample size of a uniform probability distribution. This is because the saving associated with x_2 is small compared to the other optimisation variables so there is a weak relationship between its value and the feasibility of the candidate. However the values at $\eta = 100$ show that all the solutions cluster towards increased train comfort despite the reduction in savings, with most solutions relating to increasing train comfort by close to the maximum amount. The $i = 3$ plot shows that most optimisations found a solution to have permissible line speeds of 300 to 280 km/hour between London and Birmingham, whereas the plots for $i = 4$ and $i = 5$ show that most optimisation had speeds less than this for the Birmingham-Manchester and Birmingham-Leeds lines. This is because the London-Birmingham line has the greatest number of passengers travel upon it, meaning that shortened journey times here have a large positive effect. However, because this task models a constant line capacity that is feasible at the highest permissible line speed, this task does not capture the trade-off between permissible line speed and line capacity for trains. As intuitively expected, the solution permissible line speed has been approximately 'matched' to the solution train maximum speed. The plots for $i = 6, 7, 8$ and 9 show that all the optimisation found solutions which decreased the comfort of stations in order to make savings.

The results indicate that by redistributing investment in HS2 to prioritise the comfort of the trains rather than speed and station comfort, a £5 billion saving can be made and the network score can be improved by up to 8%. Whilst an improved network score at lower investment might be counterintuitive, it is plausible - the 'base case', i.e. $X = [0,0,0,0,0,0,0,0]$, network might relate to a network defined by a larger amount of investment which is poorly allocated. Similarly, the magnitude of this improvement (8%) might be counterintuitive given that the maximum improvement in passenger VoT is 5%. A hypothesis for this difference is - the solutions have slower trains, this leads to better connections between trains, i.e. less waiting time for passengers, at Birmingham and hence better network score. There is evidence to confirm this

hypothesis. Figure 43 shows that the 'optimum network' is using slower trains. The total amount of passenger time spent in the 'On Platform' stage can be used to compare the quality of the connections at Birmingham. For the 'optimum network' this time is 12% less than it is for the 'base case' network, which indicates better connections in the optimum network. This is consistent with the real world – slower trains mean less headway is required and therefore increased frequency of train arrivals. A consequence of the dependence of connection frequency to X is that the improvement in network score from the 'base case' to the 'optimum' network is a combination of more factors than just train comfort – and hence the magnitude of the improvement is greater than the change to the VoT weighting for the 'On Train' journey stage.

The optimum train maximum speed and acceleration was found to be that represented here by a Class 373 with permissible line speeds to facilitate its maximum speed. If permissible line speeds are to be reduced further for greater savings, maintaining the permissible line speed of the London-Birmingham lines is most critical. However, the investigation does not capture the benefit of investing in greater permissible line speed capability now so that faster rolling stock might be used in the future. The comfort of stations is found to be a low priority and reducing investment in these can be considered. The results are likely sensitive to the relationship between savings and component performance. Consequently, another area for further investigation is to conduct a sensitivity analysis on the parameter values used to represent these. Nonetheless, the case study has shown how the SUPREME framework can be used to support network design and has indicated a value engineering strategy for the HS2 network. Investigating the effect of varying the minimum savings requirement (κ_S) is an area for further investigation.

6.4 Summary

The case studies in this Chapter have demonstrated the application of the SUPREME framework to challenges at the three management levels discussed in Chapter 2: strategic (case study three), tactical (case study one), and

operational (case study two). This confirms the suitability of the SUPREME framework with respect to the adaptability aims of this thesis.

The case studies have also shown that an area for future development of the SUPREME framework is to capture the time between the passengers 'ideal' departure time and the time of the soonest train after that, i.e. the 'Spare Time' stage. In this chapter PRaM has been developed to capture this, however, the VoT weighting value for the 'Spare Time' journey stage has been estimated because no data could be found for this. Improving the accuracy of this value would be beneficial.

The case studies have shown that the PRaM model can capture different networks, again confirming the adaptability of the SUPREME framework. In general, the results show that the BO implementation can find solutions to the task which typically reduce, i.e. improve, the network score by more than 10% with a budget of objective function evaluations that is representative to the planning window of the challenge. It has also been seen that 'good' solutions can be found by arbitrarily choosing the initial candidate. This indicates that the SUPREME framework could be used by non-expert practitioners without the insight to provide 'good' initial candidates. This chapter also demonstrates how the results of each optimisation can be interpreted to support decision making in rail network management.

Chapter 7

Conclusions

In this thesis the Sheffield University Passenger Rail Experience Maximiser (SUPREME) framework has been developed to assist rail network managers optimise their network for maximum passenger satisfaction. This is achieved by minimising a parameter capturing passenger disutility which is dependent on parameters describing the network. To quantify the effect of a network upon passenger satisfaction, a network assessment metric was developed in Chapter 3 which considers the distribution of normalised individual journey scores captured with a journey metric. Chapter 2 describes evidence in the literature that a journey metric which includes different journey stages and conditions, will better capture passenger satisfaction than end-to-end journey time only. Chapter 3 demonstrates how the multi-stage nature and conditions of a journey can be captured using the Value of Time (VoT) concept to weight the time spent in different passenger states. The developed metric therefore captures passenger disutility which, if minimised, is taken to be the equivalent of maximising passenger satisfaction; an assumption that is supported by evidence discussed in Chapter 2. The developed network metric based on journey scores is shown to predict observed changes in passenger satisfaction for the London Underground Limited (LUL) Victoria Line. This validation indicates that the metric can be used to determine a parameter summarising network performance from the passenger perspective, which can then be optimised.

The investigation also focusses on comparison of the journey score metrics from the UK and Japan. In what is, to the best of the author's knowledge, the first comparison of the journey score metrics used in the developed network metric, Chapter 3 has shown that the VoT for periods in similar states varies between journey metric. Because the journey metrics originate from different countries, this indicates that VoT weightings might vary and reflect the differing priorities of passengers in different regions, as would be intuitively expected. Therefore it is concluded that, for the best representation of passenger

experience when using the developed metric, network managers should determine VoT values relevant to the passengers of their network, either at the country level or even more locally. This can be achieved with the methods described by ARUP (2015). Nonetheless, the network metric does predict relative changes to the satisfaction of passengers when either journey metric is used, so could likely be used 'as is' if determining network-specific VoT values is not possible. To the best of the author's knowledge, Chapter 3 describes the first whole-network assessment metric to capture the multi-stage nature of passenger journeys and crowding, that has been quantitatively validated against observed data.

To model candidate networks the Passenger Rail Model (PRaM) simulation of passenger rail networks has been developed in Chapter 4 using Agent-Based Modelling (ABM) techniques. It is shown that train control actions are predicted to within 23 seconds and, when train control is known, the journey time required by a train for a 100km journey is predicted to within two seconds. This demonstrates that using simple models of the tractive force and resistance acting on a train, as well as driving behaviour and signalling, a model can be developed with accuracy acceptable for investigating challenges within rail network optimisation such as timetable design and rolling stock allocation. Similarly, it is shown that passenger agents can be used to model the stages of a journey required to calculate the network assessment metric developed in Chapter 3. As well as demonstrating the suitability of PRaM against the aims of this thesis to capture whole-network effects and individual passenger journeys, this indicates that ABM techniques are well-suited to capturing individual passenger journeys in the detail required for an assessment of network performance from the passenger perspective. Consequently, ABM techniques might be further applied for tools modelling rail networks. Within the PRaM simulation, a model of the train dwell times is also included, demonstrating that PRaM can be extended to include sub-models of different rail network effects. This extendibility and adaptability is a strength of the ABM approach because it has the potential to better capture the interconnected nature of rail networks and the effect upon passenger journeys. This could be achieved with

commercial train movement simulations, e.g RailSys and OpenTrak, and models of other factors which affect passenger journeys, e.g. pedestrian movement within stations and the platform-train interface. The previously mentioned industrial project (University of Sheffield, 2017) involving the author has investigated this and shown it to be a promising area for development.

Following their validation, the network model and assessment metric, respectively developed in Chapter 4 and Chapter 3, were used within the SUPREME framework as an objective function to quantify the performance of different candidate networks within the optimisation procedure. However, representing trains and passengers individually increases the computational cost of the objective function. For example, Chapter 6 has shown that computing this objective function requires approximately eight minutes of computation time to capture a day's operation of the prospective HS2 network, i.e. is 'expensive-to-compute'. If applied with many of the common optimisation algorithm methods used within rail network optimisation, e.g. the Genetic Algorithm (GA) method, a task involving an expensive-to-compute objective function might be intractable. Therefore an optimisation tool using the developed objective function with the GA method might have been restricted to challenges involving small networks and long planning windows. However, Chapter 5 has shown that using the Bayesian Optimisation (BO) method within the SUPREME framework means that 'good' solutions can be found with a mean reduction of 43 times fewer objective function evaluations than if the GA method is used, and consequently is selected for use within SUPREME. To the best of the author's knowledge, this is the first time that BO has been applied for optimising the performance of a rail network. For the expensive-to-compute objective functions encountered in this thesis, the use of BO substantially reduces the computational cost of the optimisation procedure. Consequently, this indicates that using the BO method allows the SUPREME framework to be applied to challenges involving larger networks than if the GA method was used with the same computational budget. Furthermore, it also means that the SUPREME framework can assist with challenges where there is a limited time available for computation, e.g. re-timetabling after perturbation.

The investigation in Chapter 5 identifies some drawbacks to the BO implementation used and possible directions for its improvement. Developing optimisation algorithm methods is outside the scope of this work, which concentrates on their application within the wider framework, however if improved optimisation algorithms were developed they could readily be integrated within the SUPREME framework owing to its modular structure. The benefits of BO identified in this thesis might be transferable to other rail network optimisation tools, allowing either the use of objective functions which better capture the real-world challenge, or reducing the computational cost of the tool.

Chapter 6 has shown the application of the SUPREME framework in case studies with varying networks, time windows for the optimisation and resources to be allocated. This indicates that the SUPREME framework is adaptable enough to be used for different challenges within rail networks, thus increasing its usefulness to network managers and confirming its suitability with respect to the aims of this thesis. Furthermore, Chapter 2 explains that it is likely that a solution to a task which captures multiple challenges at once will perform better in the real-world than if solutions to individual challenges are combined. Therefore a methodology adaptable enough to capture multiple challenges is likely to yield better solutions. Chapter 6 has also shown that although, in general, better solutions are found when more objective function evaluations are used, in some cases 'good' solutions can still be found with approximately ten objective function evaluations. This is unlikely to be possible with population based methods, e.g. GA, because the typical population sizes used mean that ten objective function evaluations is equivalent to a Random Search algorithm. This indicates the use of the BO method within the SUPREME framework has meant that it can be applied to tasks with a limited time window, e.g. on the morning of operation. The case studies also demonstrated how the output of SUPREME can be analysed to better understand the solutions and consequently demonstrates an advantage of choosing to distance-normalise the passenger journey scores in Chapter 3. Furthermore, for the case studies investigated, the SUPREME framework is shown to

improve the network score by the order of 10%, hence improving the experience of passengers. Section 6.1.5 describes that if passenger experience is improved, passengers may be willing to spend more on their ticket, as would be intuitively expected. Using the VoT provided by ARUP (2015) and focussing on commuter passengers because they are the national majority (Peluffo, 2018), a 10% improvement in the network score would result in an additional £1.12 of ticket revenue per passenger travel hour. Although no data was available to the author describing the number of GB rail passenger travel hours, an estimate for this value can be calculated by considering that there were approximately 1.8 billion passenger journeys in 2017 (Peluffo, 2018) and estimating that the mean journey time was half an hour based on the 38km average journey distance in 2018 (Office of Rail and Road, 2019). Consequently, a 10% improvement in GB network score could enable an extra billion pounds of ticket revenue to be collected annually. This would represent approximately a 10% increase on the current ten billion pound annual ticket revenue reported by the Office of Rail and Road (2019). Clearly, before it can be scaled to a GB sized network, the SUPREME framework requires further development and this is described in the following sub-sections.

7.1 The network assessment metric

There are two strands of further work relating to the network assessment metric, which will be considered in order: to further investigate the accuracy of the developed metric, and to further capture a greater range of passenger experiences.

Although Chapter 3 has shown some experimental evidence that the developed network metric better predicts changes to passenger satisfaction than a metric based on journey time only, there is not enough data to determine a statistically significant difference beyond the 95% confidence level. In order to determine a statistically significant difference with greater confidence level, further work would be needed to conduct an investigation where more than ten measurements of network performance can be collected and corresponding predictions computed (in the case of the experiment in

Chapter 3 each measurement equates to a year over which there is available passenger satisfaction data relating to the timetable operated that year, but any timescale over which a system change and its effect can be measured may be considered in future experiments). This might be achieved by re-investigating the Victoria Line in the future as additional years of customer satisfaction data become available.

The data available for the investigation in Chapter 3 meant that individual passenger routes were not captured in the validation experiment. Consequently, further work might also to investigate networks where a more detailed description of the passenger route is available so that the effect of train transfer on passenger experience can be captured. The network assessment metric could then be validated for journeys which include this activity and potentially also allow a statistically significant difference with end-to-end journey time to be discerned. Furthermore, if data describing passenger satisfaction for more networks is available, additional comparisons between prediction and measurement would make it possible to determine if the results in Chapter 3 are network-specific or general.

Updating the network assessment metric with new VoT weightings to capture other factors which influence passenger experience is also an area for further work. For example, the findings of Transport Focus (2016) indicate that the cleanliness of trains has an effect on passenger satisfaction. To capture cleanliness conditions, The VoT weighting for passengers in the 'On Train' stage could be modified with a 'cleanliness penalty'. The case studies of Chapter 6 have shown that SUPREME might be improved if more journey stages are included. For example, the 'Spare Time' stage is introduced but the VoT weighting is estimated. Furthermore, Transport Focus identify that the punctuality of trains is an important factor of passenger satisfaction. To capture this, additional journey stages could be introduced representing when a passenger is waiting for, or on-board, a delayed train. All these additions to the metric would require models and parameters relating to the VoT weightings and would necessitate the revalidation of the metric. The VoT weightings for different passenger travel purposes are already available, e.g.

commuter and business (ARUP, 2015). These could be implemented within the current SUPREME framework, however data describing the number of passengers for each journey type was not available so this was not demonstrated.

Consistent with the aims of this thesis, the developed network assessment metric considers passenger experience only. However, for network managers there are often competing interests that need to be balanced. To address this, some authors consider multiple-objectives simultaneously, e.g. Goodwin et al. (2016) optimise the trade-off between energy use and journey times.

Therefore, a metric to capture passenger experience could be combined with a metric which captures other considerations for network managers, e.g. energy use.

7.2 The network model

There are three strands of further work relating to PRaM, which will be considered in order: to better understand its performance and limitations as it currently is, to develop its modelling capability further, and to reduce its computational cost. The Train Control Model (TCM) validation experiment in Chapter 4 has shown that, when data describing a real-world situation is available, the TCM can predict the time of control points to within 23 seconds. However, this comparison is for one train journey only, where not all the data describing situations which may have affected the control of the real-world train is available (e.g. temporary permissible line speed reductions and the location of other trains). The validation could be improved by using the live data published by Network Rail (c2017), which was unknown to the author at the time of investigation, describing the signalling communicated to trains as well as their position and velocity. Train performance data for the input parameters of the Train Dynamics Model (TDM) would be needed for all trains investigated, as well as the line gradients of the journeys. This would allow the comparison of predictions and measurements for many more train journeys, and would better capture the variation of human driving as a result.

Similarly, quantifying the sensitivity of the TDM to different parameters has been considered for one journey only, yet it may vary for different line gradients and train types. By considering the distribution of results when investigating the sensitivity of the TDM for more journeys and train types, a value can be obtained with greater confidence of representing a 'general train and journey'. To achieve this, train performance and line gradient data would be required from the sources already described. Furthermore, the sensitivity of the TDM has only been considered for individual errors in the input parameters, however, combined interactions are likely to be significant (e.g. if the train mass is overestimated and the maximum tractive force underestimated). This investigation could be undertaken with no additional input data, but has been a low priority task compared to developing the other modules of the SUPREME framework.

Developing the movement model for train agents so that junctions can be represented would improve the modelling capability of PRaM. The BRaVE simulator and others developed by authors such as Dessouky and Leachman (1995) and Lee-Gunther et al. (1995), can represent networks with junctions and could be used to inform the development of this functionality in PRaM. Power distribution within the network is another topic of interest to rail managers where models have previously been developed, e.g. by Mott MacDonald (2006). The inclusion of more models within PRaM allows more factors relating to rail network operation to be considered simultaneously, however it also increases the computational cost.

Modelling the physical location and movement of passengers within stations and trains is another area that could improve PRaM. Currently, for example, PRaM models the amount of time passengers require to move through a station as being constant, however passenger numbers within the station or the station's layout may affect this. This variation in time might be the difference between passengers making a connection or not. Similarly, Farnsworth et al. (2017) discuss how a heterogeneous distribution of passengers along the platform and train affects dwell times and passenger comfort. Train capacity is also affected because passengers on the platform at

densely packed areas of the train erroneously believe it to be full. ABM simulations of crowd movements through buildings (Karmakharm et al., 2010) and the platform-train interface (Fletcher et al., 2018) have already been developed. The author has been involved in a project to integrate the model of train agents developed in this project with a pedestrian simulation (University of Sheffield, 2017), extending this and interfacing these models with PRaM would allow it to better capture the interactions between the network and passengers and the effect on individual passenger journeys. A challenge to this is the increase in computational cost. A further development that could be made to PRaM is to capture the variations in individual passenger preferences and behaviours. For example, passenger agents could be given individual walking speeds, or tolerances of crowding which determine whether they board a crowded train agent or wait for the next. Programmatically, these are simple adjustments to make to PRaM, however, identifying or developing models which capture the distribution of these preferences and behaviours may be more difficult. Models of passenger route choice (Ke et al., 2012) and information systems (Kattan and Bai, 2018) would better allow PRaM to capture changes to the flow of passengers when the network is altered.

The case studies of Chapter 6 have shown that PRaM could better capture some challenges if models of the interactions between different network components were included. For example, in the HS2 network investigation the trade-off between permissible line speed and the capacity for trains is not captured. All candidate networks are modelled to have the capacity for trains associated with the minimum permissible line speed investigated. The inclusion of a line planning and timetabling model, e.g. those reviewed by Scheepmaker et al. (2017), would allow this to be captured. Similarly, the network capacity at stations and junctions model could be captured using the relationship with delay described by Armstrong and Preston (2017), for example. The addition of these sub-models to PRaM would increase the computational cost of the objective function so only relevant models would be calculated at every objective function evaluation. Those that are not sensitive to the optimisation

variables would be calculated once at the start of the optimisation and their output applied for all evaluations.

Considering that simulating a small network once with PRaM requires approximately 20 seconds and that, within SUPREME, variations of the network will be simulated multiple times, there is a clear benefit to reducing the computational cost of a single simulation. Using existing computational resources more effectively is preferential to requiring a more powerful machine (compared to the Intel Xeon Dual Processor @ 2.4 GHz used for the experiments in this thesis - a 'server' machine with performance likely available to the large institutions involved in network management). Some savings may be made by using more efficient serial coding techniques. Parallel programming techniques have been shown to allow substantial speed up (Eager et al., 1989) and might be well suited to computing the updates for passenger agents because of the large number of agents involved. However, for the benefits of parallelism to be realised, many agents must be undergoing a similar process. This requires arranging passenger agents undergoing similar processes to be arranged into 'blocks' for computing. For example, all passengers in the same carriage might be arranged into one block. The arrangement process has a computational overhead in itself and may need to be repeated each time the occupancy of the carriage changes. Investigation is needed to see whether the savings from computing the number of passengers in a block in parallel outweigh the overhead. Because passenger numbers and the frequency of changes to the occupancy of a carriage vary, depending on the network (e.g. inter-city network compared to metro), the benefits of parallelism may be network dependent.

7.3 The optimisation algorithm

The experiments in this thesis are conducted on tasks with up to 16 optimisation variables. If SUPREME is to be scaled to larger networks, it is likely that the tasks would involve more optimisation variables. Investigating the effectiveness of the BO implementation for tasks with more than 16 optimisation variables is an area for further work. This could be achieved using

hypothetical tasks, similar to those used in Chapter 5. Other BO implementations, specifically those for higher-dimension tasks, could also be compared as well as the effect of using a different probabilistic model, i.e. not a Gaussian Process regression model, within the BO method. Chapter 5 also describes evidence that it might be most effective to use a combination of BO and GA methods. Developing and investigating an implementation to do this is an area for further work that could possibly have wide-spread application in other fields.

The case studies in Chapter 6 indicate that, for a given computational budget, there is a trade-off between the strategies of computing a single optimisation with a large number of objective function evaluations, or computing multiple optimisations each with less objective function evaluations. Investigating this trade-off, and whether it itself can be optimised, is an area for further investigation. This could be achieved by comparing the best solution found when using different strategies for a given total computational budget. The investigation would need to be repeated for different tasks to see whether the results are task-specific or general. If a general relationship for the best strategy can be determined this might further reduce the total cost of optimising with the BO method and have application in many scenarios where BO is applied.

References

- ABBINK, E., VAN DEN BERG, B., KROON, L. & SALOMON, M. 2004. Allocation of Railway Rolling Stock for Passenger Trains. *Transportation Science*, 38, 33-41.
- ABID, M. M., KHAN, M. B. & MUHAMMAD, I. 2015. Timetable Optimization For Single Line Railway. *European Transport/Trasporti Europei*, 17.
- ALI, A. A., ELIASSON, J. & WARG, J. 2017. Measuring the Socio-economic Benefits of Train Timetables Application to Commuter Train Services in Stockholm. *Transportation Research Procedia*, 27, 849-856.
- ALIOTH FINANCE. c2019. *UK Inflation Calculator* [Online]. Available: <http://www.in2013dollars.com/UK-inflation> [Accessed 09/08/2019].
- ANGEL TRAINS. c2019a. *Class 153 - Keolis Amey, Great Western Railway, East Midlands Trains, Northern Rail (by Arriva)* [Online]. London, UK. Available: <https://www.angeltrains.co.uk/Products-Services/Regional-Passenger-Trains/4> [Accessed 15/05/2019].
- ANGEL TRAINS. c2019b. *Mk3 HST Trailer Cars* [Online]. London, UK. Available: <https://www.angeltrains.co.uk/Products-Services/High-Speed-Trains/49> [Accessed 01/03/2019].
- ARMSTRONG, J. & PRESTON, J. 2017. Capacity utilisation and performance at railway stations. *Journal of Rail Transport Planning & Management*, 7, 187-205.
- ARUP. n.d. *Post Consultation Route Birmingham Interchange* [Online]. London, UK: Department for Transport. Available: <http://assets.dft.gov.uk/publications/hs2-maps-20120110/hs2arp00drrw05303issue2.pdf> [Accessed 07/06/2019].
- ARUP, INSTITUTE FOR TRANSPORT STUDIES LEEDS & ACCENT. 2015. *Provision of market reserch for value of travel time savings and reliability: Non-Technical Summary Report* [Online]. London, UK: Department for Transport. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470229/vtts-phase-2-report-non-technical-summary-issue-august-2015.pdf [Accessed 20/03/2019].
- ASSOCIATION OF TRAIN OPERATING COMPANIES. c2015. *Official National Rail Map*, [Online]. London, UK: National Rail,. Available: <https://www.nationalrail.co.uk/static/documents/content/OfficialNationalRailmaplarge.pdf> [Accessed 20/5/2019].
- BALCOMBE, R. MACKETT, R., PRESTON, J., TITHERIDGE, H., WARDMAN, M. & WHITE, P. 2004. *The demand for public transport: a practical guide* [Online]. London, UK: TRL Limited. Available: <https://trl.co.uk/sites/default/files/TRL593%20-%20The%20Demand%20for%20Public%20Transport.pdf> [Accessed 28/12/2019].

- BALDWINSON, T. 2015. *Manchester to Paris night trains with existing rail resources* [Online]. Manchester, UK: TBR Consulting. Available: <https://tonybaldwinson.files.wordpress.com/2015/05/2015-05-16-innovation-north-of-london-obc-for-night-trains-manchester-to-paris.pdf> [Accessed 05/06/2019].
- BALLARD, G., BENSON, A. R., DRUINSKY, A., LIPSHITZ, B. & SCHWARTZ, O. 2016. Improving the Numerical Stability of Fast Matrix Multiplication. *SIAM Journal on Matrix Analysis and Applications*, 37, 1382-1418.
- BENTLEY SYSTEMS. c2020. *LEGION Simulator* [Online]. Exton (PA), USA. Available: <https://www.bentley.com/en/products/product-line/building-design-software/legion-simulator> [Accessed 23/01/2020].
- BEŠINOVIĆ, N., QUAGLIETTA, E. & GOVERDE, R. M. P. 2013. A simulation-based optimization approach for the calibration of dynamic train speed profiles. *Journal of Rail Transport Planning & Management*, 3, 126-136.
- BODMAN, A. 2012. *Committee of Public Accounts - The completion and sale of High Speed 1* [Online]. London, UK: Parliament. Available: <https://publications.parliament.uk/pa/cm201213/cmselect/cmpublicacc/464/464we05.htm> [Accessed 05/06/2019].
- BONABEAU, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99, 7280.
- BRUZELIUS, N. A. 1981. Microeconomic theory and generalised cost. *Transportation*, 10, 233-245.
- BURDETT, R. L. 2015. Multi-objective models and techniques for analysing the absolute capacity of railway networks. *European Journal of Operational Research*, 245, 489-505.
- BURGGRAEVE, S., BULL, S. H., VANSTEENWEGEN, P. & LUSBY, R. M. 2017. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C - Emerging Technologies*, 77, 134-160.
- CACCHIANI, V., CAPRARA, A. & TOTH, P. 2010. Solving a real-world train-unit assignment problem. *Mathematical Programming*, 124, 207-231.
- CACCHIANI, V. & TOTH, P. 2012. Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219, 727-737.
- CACERES, L. P., LOPEZ-IBANEZ, M. & STUTZLE, T. 2015. Ant colony optimization on a limited budget of evaluations. *Swarm Intelligence*, 9, 103-124.
- CAIMI, G., KROON, L. & LIEBCHEN, C. 2017. Models for railway timetable optimization: Applicability and applications in practice. *Journal of Rail Transport Planning & Management*, 6, 285-312.
- CALDER, S. 2018. *Costs of HS2 'Vastly Underestimated', whistleblower says* [Online]. London, UK: Independent Digital News & Media. Available: <https://www.independent.co.uk/travel/news-and-advice/hs2-cost-high->

- [speed-2-spending-price-train-line-budget-when-a8687866.html](#)
[Accessed 06/06/2019].
- CANDELIERI, A., PEREGO, R. & ARCHETTI, F. 2018. Bayesian optimization of pump operations in water distribution systems. *Journal of Global Optimization*, 71, 213-235.
- CARNE, M. 2018. *Strategic Business Plan 2019 - 2024* [Online]. London, UK: Network Rail,. Available: <https://cdn.networkrail.co.uk/wp-content/uploads/2018/02/Strategic-business-plan-high-level-summary.pdf> [Accessed 15/11/2018].
- CHAMBERS, J., CLEVELAND, W., KLEINER, B. & TURKEY, P. 1983. Comparing Data Distributions. In: CHAMBERS, J. (ed.) *Graphical Methods for Data Analysis*. Belmont (CA), USA: Wadsworth International Group.
- CHAN, J. 2007. *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data*. Master of Science in Transportation, Massachusetts Institute of Technology.
- CHANDRASHEKARAN, A. & LANE, I. 2016. Automated Optimization of Decoder Hyper-parameters for Online LVCSR. *IEEE Workshop on Spoken Language Technology*. San Diego (CA), USA: IEEE.
- CHANG, Y. H., YEH, C. H. & SHEN, C. C. 2000. A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line. *Transportation Research Part B - Methodological*, 34, 91-106.
- CHEN, C. & CHEN, F. 2010. Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists. *Tourism Management*, 31, 29-35.
- CHEN, F., YANG, H. & YANG, Y. 2019. MAS-based Model for Evaluating Train Timetables to Minimise the Waiting Time. *KSCE Journal of Civil Engineering*, 23, 1360-1368.
- CHONG, E. K. P. 2013. *An introduction to optimization*, Hoboken (NJ), USA, Wiley.
- CIVITY MANAGEMENT CONSULTANTS. 2014. *Further Development of the European High Speed Rail Network* [Online]. Hamburg, Germany. Available: https://civity.de/asset/en/sites/2/2018/05/civity_dev_eu_hsr_network_012014.pdf [Accessed 05/06/2019].
- CLINNICK, R. 2018. *It's a knockout!* [Online]. Peterborough, UK: Rail Magazine. Available: <https://www.railmagazine.com/news/rail-features/it-s-a-knockout> [Accessed 08/06/2019].
- COMMITTEE OF PUBLIC ACCOUNTS. 2017. *Modernising the Great Western Railway* [Online]. London, UK: Parliament, House of Commons. Available: <https://publications.parliament.uk/pa/cm201617/cmselect/cmpubacc/776/776.pdf> [Accessed 15/11/2018].

- DAVIS, R. 2018. *Rail Factsheet* [Online]. London, UK: Department for Transport. Available:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/761352/rail-factsheet-2018.pdf [Accessed 13/12/2018].
- DEPARTMENT FOR TRANSPORT. 2009. *The Passenger Demand Forecasting Handbook Methodology* [Online]. London, UK. Available:
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.375.1644&rep=rep1&type=pdf> [Accessed 15/01/2020].
- DEPARTMENT FOR TRANSPORT. 2012. *Rail passenger numbers and crowding on weekdays in major cities in England and Wales: 2012* [Online]. London, UK: gov.uk. Available:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/252516/rail-passengers-crowding-2012-revised.pdf [Accessed 13/05/2019].
- DEPARTMENT FOR TRANSPORT. 2016. *High Speed Two: From Crewe to Manchester, the West Midlands to Leeds and beyond* [Online]. London, UK: gov.uk. Available:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/568208/high-speed-two-crewe-manchester-west-midlands-leeds-web-version.pdf [Accessed 07/06/2019].
- DEPARTMENT FOR TRANSPORT. 2017a. *East Midlands Rail Franchise Public Consultation* [Online]. London, UK. Available:
<http://www.englandseconomicheartland.com/Documents/East%20Midlands%20Rail%20Franchise%20Consultation.pdf> [Accessed 16/01/2019].
- DEPARTMENT FOR TRANSPORT. 2017b. *High Speed Two Phase Two Economic Case* [Online]. London, UK. Available:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/634196/high-speed-two-phase-two-economic-case.pdf [Accessed 07/06/2019].
- DEPARTMENT FOR TRANSPORT. 2018. *TAG UNIT A1.1 Cost-Benefit Analysis* [Online]. London, UK. Available:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/712699/tag-unit-a1.1-cost-benefit-analysis-may-18.pdf [Accessed 14/01/2020].
- DESSOUKY, M. M. & LEACHMAN, R. C. 1995. A Simulation Modeling Methodology for Analyzing Large Complex Rail Networks. *Simulation*, 65, 131-142.
- DORRIAN, J., HUSSEY, F. & DAWSON, D. 2007. Train driving efficiency and safety: examining the cost of fatigue. *Journal of Sleep Research*, 16, 1-11.
- EAGER, D. L., ZAHORJAN, J. & LAZOWSKA, E. D. 1989. Speedup versus efficiency in parallel systems. *IEEE Transactions on Computers*, 38, 408-423.

- EAST MIDLANDS TRAINS. 2019. *Download Train Timetables* [Online]. Derby, UK. Available: <https://www.eastmidlandstrains.co.uk/train-times/printed-timetables/> [Accessed 21/05/2019].
- ELBES, M., ALZUBI, S., KANAN, T., AL-FUQAHA, A. & HAWASHIN, B. 2019. A survey on particle swarm optimization with emphasis on engineering and network applications. *Evolutionary Intelligence*, 12, 113-129.
- EUROPEAN COMMISSION. 2019. *The European Deployment Plan and National Implementation Plans* [Online]. Brussels, Belgium. Available: https://ec.europa.eu/transport/modes/rail/ertms/ertms_deployment_en [Accessed 13/02/2019].
- FAN, B., ROBERTS, C. & WESTON, P. 2012. A comparison of algorithms for minimising delay costs in disturbed railway traffic scenarios. *Journal of Rail Transport Planning & Management*, 2, 23-33.
- FARNSWORTH, M., KIRKWOOD, L., COURT, S., SHEHAB, E. & TINWORTH, N. 2017. Optimisation Strategy for Efficient Platform Train Interface Activity. In: GAO, J., EL SOURI, M. & KEATES, S. (eds.) *Advances in Manufacturing Technology XXXXI*. Amsterdam, Netherlands: IOS Press.
- FERNANDEZ-RODRIGUEZ, A., FERNANDEZ-CARDADOR, A., CUCALA, A. P., DOMINGUEZ, M. & GONSALVES, T. 2015. Design of Robust and Energy-Efficient ATO Speed Profiles of Metropolitan Lines Considering Train Load Variations and Delays. *IEEE Transactions on Intelligent Transportation Systems*, 16, 2061-2071.
- FLETCHER, D., HARRISON, R., KARMAKHARM, T., NALLAPERUMA, S. & RICHMOND, P. 2018. RateSetter: roadmap for faster, safer, and better platform train interface design and operation using evolutionary optimisation. *Proceedings of the Genetic and Evolutionary Computation Conference*. Kyoto, Japan: ACM.
- FLORIAN, M., CHAPLEAU, R. NGUYENS, S., ACHIM, C., JAMES, L. & LEFEBVRE, J. 1977. EMME: A planning method for multi-modal urban transportation systems. In: Visser E.J. (eds) *Transport decisions in an age of uncertainty*. Dordrecht, Netherlands: Springer.
- FORD, R. 1986. The DMU Replacement Story. *Modern Railways*. Shepperton, UK: Ian Allan.
- FOUSKAKIS, D. & DRAPER, D. 2002. Stochastic optimization: a review. *International Statistical Review*, 70, 315-349.
- FURNESS, J. 2017. *Should improvements be made to the Crewe-to-Derby rail line?* [Online]. Stoke-on-Trent, UK: StokeOnTrentLive. Available: <https://www.stokesentinel.co.uk/news/local-news/should-improvements-made-crewe-derby-962825> [Accessed 21/05/2019].
- FURNESS, N., VAN HOUTEN, H., ARENAS, L. & BARTHOLOMEUS, M. 2017. ERTMS Level 3: The Game Changer. *IRSE News*. London, UK: Institution of Railway Signal Engineers.

- GARDNER, J., PLEISS, G., WEINBERGER, K. Q., BINDEL, D. & WILSON, A. W. 2018. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. *In: BENGIO, S., WALLACH, H., LAROCHELLE, H., GRAUMAN, K., CESA-BIANCHI, N. & GARNETT, R. (eds.) Neural Information Processing Systems 2018*. Montreal, Canada: Neural Information Processing Systems Foundation.
- GOLDBERG, D. E. 1989. *Genetic algorithms in search, optimization, and machine learning*. Reading (MA), USA, Addison-Wesley.
- GOODMAN, C. J. & TAKAGI, R. 2004. Dynamic re-scheduling of trains after disruption. *Computers in Railway Six*, 15, 765-774.
- GOODWIN, J. C. J., FLETCHER, D. I. & HARRISON, R. F. 2016. Multi-train trajectory optimisation to maximise rail network energy efficiency under travel-time constraints. *Proceedings of the Institution of Mechanical Engineers Part F - Journal of Rail & Rapid Transit*, 230, 1318-1335.
- GRIMM, M. & SANBLAD, B. 2000. Simulation of different railway signalling systems. *In: BREBBIA, C., ALLAN, J., HILL, R., SCIUTTO, G. & SONE, S. (eds.) Computers in Railways VII*. Southampton, UK: WIT Press.
- GRONAU, R. 1970. *The value of time in passenger transportation: the demand for air travel*, New York (NY), USA, National Bureau of Economic Research.
- GUPTA, S. D., TOBIN, J. K. & PAVEL, L. 2016. A two-step linear programming model for energy-efficient timetables in metro railway networks. *Transportation Research Part B - Methodological*, 93, 57-74.
- HAIGH, P. 2018. *ScotRail bike train plans to take shape* [Online]. Peterborough, UK: Bauer Consumer Media. Available: <https://www.railmagazine.com/news/network/scotrail-bike-train-plans-take-shape?image=27331> [Accessed 23/05/2019].
- HANSEN, H. S., NAWAZ, M. U. & OLSSON, N. 2017. Using operational data to estimate the running resistance of trains. Estimation of the resistance in a set of Norwegian tunnels. *Journal of Rail Transport Planning & Management*, 7, 62-76.
- HARRIS, N. G. & ANDERSON, R. J. 2007. An international comparison of urban rail boarding and alighting rates. *Proceedings of the Institution of Mechanical Engineers, Part F - Journal of Rail & Rapid Transit*, 221, 521-526.
- HENDRICKSON, C. & KOCUR, G. 1981. Schedule Delay and Departure Time Decisions in a Deterministic Model. *Transportation Science*. 15, 1, 62 - 77.
- HICKEY, S. 2011. *Improving the estimation of platform wait time at the London Underground*. Master of Science in Transportation, Massachusetts Institute of Technology.

- HICKISH, B., FLETCHER, D. I. & HARRISON, R. F. 2017. Maximising passenger satisfaction through optimised train movements. *The Stephenson Conference*. London, UK: The Institution of Mechanical Engineers.
- HICKISH, B., FLETCHER, D. I. & HARRISON, R. F. 2019. A rail network performance metric to capture passenger experience. *Journal of Rail Transport Planning & Management*.
- HOROWITZ, A. J. 1978. The subjective value of the time spent in travel. *Transportation Research*, 12, 385-393.
- HOWLETT, P. G. & PUDNEY, P. J. 1995. *Energy-Efficient Train Control*, London, UK, Springer.
- HS2. c2019. *London Euston* [Online]. Birmingham, UK. Available: <https://www.hs2.org.uk/stations/london-euston/> [Accessed 07/06/2019].
- HS2 LIMITED. 2016a. *High Speed Two Phase 2b Crewe to Manchester West Midlands to Leeds Economic Case* [Online]. London, UK: Department for Transport. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/587329/D14_Phase_2b_Economic_Case_TG_14-11-16_edit20170130.pdf [Accessed 10/06/2019].
- HS2 LIMITED. 2016b. *Route engineering report 2016* [Online]. London, UK: Department for Transport. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567616/West_Midlands_to_Leeds_Route_engineering_report.pdf [Accessed 1/06/2019].
- HS2 LIMITED. 2017. *HS2 rolling stock: scope and requirements presentation* [Online]. London, UK: Department for Transport. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/603774/HS2_rolling_stock_scope_and_requirements.pdf [Accessed 20/06/2019].
- HS2 LIMITED. 2019. *Train Technical Specification* [Online]. London, UK: gov.uk. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/794108/HS2-HS2-RR-SPE-000-000007_P11_TTS_Main_Body_External_.pdf [Accessed 01/06/2019].
- HS2 LIMITED. c2019a. *About us* [Online]. Birmingham, UK. Available: <https://www.hs2.org.uk/why/about-us/> [Accessed 08/06/2019].
- HS2 LIMITED. c2019b. *Birmingham Curzon Street* [Online]. Birmingham, UK. Available: <https://www.hs2.org.uk/stations/birmingham-curzon-street/> [Accessed 07/06/2019].
- INGVARDSON, J. B., NIELSEN, O. A., RAVEAU, S. & NIELSEN, B. F. 2018. Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. *Transportation Research Part C - Emerging Technologies*, 90, 292-306.

- INRO. c2020. *Emme* [Online]. Montreal, Canada. Available: <https://www.inrosoftware.com/en/products/emme/> [Accessed 21/01/2020]
- JOHNSON, D. SHIRES, J. NASH, C. & TYLER, J. 2006. Forecasting and appraising the impact of a regular interval timetable. *Transport Policy*, 13, 5, 349 - 366.
- JORDAN, M. I. & MITCHELL, T. M. 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349, 255-260.
- KANAI, S., SHIINA, K., HARADA, S. & TOMII, N. 2011. An optimal delay management algorithm from passengers' viewpoints considering the whole railway network. *Journal of Rail Transport Planning & Management*, 1, 25-37.
- KANDASAMY, K., SCHNEIDER, J. & PICZOS, B. 2015. High dimensional Bayesian optimisation and bandits via additive models. *Proceedings of the 32nd International Conference on Machine Learning*. Lille, France: JMLR.
- KARMAKHARM, T., RICHMOND, P. & ROMAN, D. 2010. Agent-based Large Scale Simulation of Pedestrians With Adaptive Realistic Navigation Vector Fields. In: COLLOMOSSE, J. & GRIMSTEAD, I. (eds.) *Theory and Practice of Computer Graphics*. Sheffield, UK: The Eurographics Association.
- KATTAN, L. & BAI, Y. 2018. LRT passengers' responses to advanced passenger information system (APIS) in case of information inconsistency and train crowding. *Canadian Journal of Civil Engineering*, 45, 583-593.
- KE, Q., PENG, Z. & ZHI-PENG, Q. Model of Passenger Route Choice in the Urban Rail Transit Network. In: ZHANG, Z., ZHANG, R. & ZHANG, J., eds. *Logistics, Information and Service Science, 2012 Berlin, Germany*. Springer, 463-468.
- KENNEDY, J. & MENDES, R. Population structure and particle swarm performance. In: EBERHART, R., ed. *Congress on Evolutionary Computation, 2002 Honolulu (HI), USA*. IEEE, 1671-1676.
- KOHMOTO, K., KATAYAMA, K. & NARIHISA, H. 2003. Performance of a genetic algorithm for the graph partitioning problem. *Mathematical and Computer Modelling*, 38, 1325-1332.
- KOKUBO, T. & FUKUYAMA, Y. 2018. Practical Train Crew Scheduling Problems Using Parallel Tabu Search. In: [N.K.] (ed.) *57th Annual Conference of the Society of Instrument and Control Engineers of Japan*. Nara, Japan: IEEE.
- KONING, J. 2002. Comparing The Performance Of ERTMS Level 2 Fixed Block And ERTMS Level 3 Moving Block Signalling Systems Using Simulation Techniques. In: ALLAN, J., HILL, R., BREBBIA, C., SCIUTTO, G. & SONE, S. (eds.) *Computers in Railways VIII*. Southampton, UK: WIT Press.
- KUNIMATSU, T., HIRAI, C. & TOMII, N. 2012. Train timetable evaluation from the viewpoint of passengers by microsimulation of train operation and passenger flow. *Electrical Engineering in Japan*, 181, 51-62.

- KUNIMATSU, T., HIRAI, C., TOMII, N. & TAKABA, M. 2009. Evaluation of timetables by estimating passengers' personal disutility using micro-simulation. In: HANSEN, I. (ed.) *RailZurich2009 - Third International Seminar on Railway Operations Modelling and Analysis*. Zürich, Switzerland: ETH Zürich.
- LEE-GUNTHER, J., BOLDUC, M. & BUTLER, S. Vista rail network simulation. IEEE/ASME Joint Railroad Conference, 1995 Baltimore (MD), USA. IEEE, 93-98.
- OXFORD UNIVERSITY PRESS. c2019. *Dictionary definition of 'methodology'* [Online]. London, UK: Lexico. Available: <https://www.lexico.com/en/definition/methodology> [Accessed 20/12/2018].
- LI, C., GUPTA, S., RANA, S., NGUYEN, T. V., VENKATESH, S. & SHILTON, A. 2017. High dimensional bayesian optimization using dropout. *International Joint Conference on Artificial Intelligence*. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization.
- LI, D., DAAMEN, W. & GOVERDE, R. M. P. 2016. Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station. *Journal of Advanced Transportation*, 50, 877-896.
- LI, J. J., RANKA, S. & SAHNI, S. 2011a. Strassen's Matrix Multiplication on GPUs. *17th IEEE International Conference on Parallel and Distributed Systems*. Tainan, Taiwan: IEEE.
- LI, L., ZHANG, H., WANG, X., LU, W. & MU, Z. 2011b. Urban Transit Coordination Using an Artificial Transportation System. *IEEE Transactions on Intelligent Transportation Systems*, 12, 374-383.
- LIN, D. Y. & TSAI, M. R. 2019. Integrated Crew Scheduling and Roster Problem for Trainmasters of Passenger Railway Transportation. *IEEE Access*, 7, 27362-27375.
- LISICKI, M., LUBITZ, W. & TAYLOR, G. W. 2016. Optimal design and operation of Archimedes screw turbines using Bayesian optimization. *Applied Energy*, 183, 1404-1417.
- LONDON UNDERGROUND LIMITED 2007. Victoria Line Working Timetable No. 31. London.
- LONDON UNDERGROUND LIMITED 2009. Victoria Line Working Timetable No. 32. London.
- LONDON UNDERGROUND LIMITED 2011. Victoria Line Working Timetable No. 33. London.
- LONDON UNDERGROUND LIMITED 2012a. Victoria Line Working Timetable No. 34. London.
- LONDON UNDERGROUND LIMITED 2012b. Victoria Line Working Timetable No. 35. London.

- LONDON UNDERGROUND LIMITED 2014. Victoria Line Working Timetable No. 36. London.
- LONDON UNDERGROUND LIMITED 2015a. Victoria Line Working Timetable No. 37. London.
- LONDON UNDERGROUND LIMITED 2015b. Victoria Line Working Timetable No. 38. London.
- LONDON UNDERGROUND LIMITED. 2016a. *FOI 0121 1617.xlsx* [Online]. London, UK: What Do They Know. Available: <https://www.whatdotheyknow.com/request/i-need-to-understand-the-definit> [Accessed 20/12/2018].
- LONDON UNDERGROUND LIMITED 2016b. Victoria Line Working Timetable No. 39. London.
- LONDON UNDERGROUND LIMITED 2016c. Victoria Line Working Timetable No. 40. London.
- LONDON UNDERGROUND LIMITED 2017. Victoria Line Working Timetable No. 41. London.
- LONDON UNDERGROUND LIMITED. 2018a. *Performance Data Almanac* [Online]. Transport for London. Available: <https://tfl.gov.uk/corporate/publications-and-reports/underground-services-performance> [Accessed 29/11/2018].
- LONDON UNDERGROUND LIMITED. 2018b. *Rolling Origin Destination Survey* [Online]. London, UK: Transport for London. Available: <https://tfl.gov.uk/info-for/open-data-users/our-open-data> [Accessed 20/12/2018].
- MACAL, C. M. & NORTH, M. J. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4, 151-162.
- MARINOV, M., SAHIN, I., RICCI, S. & VASIC-FRANKLIN, G. 2013. Railway operations, time-tabling and control. *Research in Transportation Economics*, 41, 59-75.
- MATHESON, J. 2010. 'Average' Briton highlighted on UN World Statistics Day [Online]. London, UK: Office for National Statistics. Available: <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/theaveragebriton> [Accessed 28/02/2019].
- MATHWORKS. 2019. *Vectorization* [Online]. Natick (MA), USA. Available: https://uk.mathworks.com/help/matlab/matlab_prog/vectorization.html [Accessed 14/05/2019].
- MATOUŠEK, J. & GÄRTNER, B. 2007. Integer Programming and LP Relaxation. In: MATOUŠEK, J. & GÄRTNER, B. (eds.) *Understanding and Using Linear Programming*. Berlin, Germany: Springer.
- MCLEOD, M., OSBORNE, M. & ROBERTS, S. 2018. Optimization, fast and slow: optimally switching between local and Bayesian Optimization. *35th*

- International Conference on Machine Learning*. Stockholm, Sweden: PMLR.
- MENG, L. Y. & ZHOU, X. S. 2014. Simultaneous train rerouting and rescheduling on an N-track network: A model reformulation with network-based cumulative flow variables. *Transportation Research Part B - Methodological*, 67, 208-234.
- MITCHELL, M. 1996. *An introduction to genetic algorithms*, Cambridge (MA), USA, MIT Press.
- MOCHIDA, T., YAMAMOTO, N., GODA, K., MATSUSHITA, T. & KAMEI, T. 2010. Development and Maintenance of Class 395 High-speed Train for UK High Speed 1. *Hitachi Review*, 59, 8.
- MOTT MACDONALD. 2006. *Validation of the TRAIN Program. DC Validation* [Online]. Croydon, UK. Available: www.mottmac.com/download/file/127/24175/trainvalidationreportrevapdf [Accessed 12/04/2019].
- N.A. 2002. *Eurostar link to Yorkshire* [Online]. London, UK: British Broadcasting Corporation. Available: <http://news.bbc.co.uk/1/hi/england/2014332.stm> [Accessed 05/06/2019].
- N.A. 2003. *Document "RSC" - Rolling Stock Capacities* [Online]. London (UK): UK Government. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/4153/tpecapacities.pdf [Accessed 11/04/2019].
- N.A. 2014. *Eurostar's 20th anniversary reached* [Online]. Courbevoie, France: Atos IT Services UK. Available: <http://www.rail.co.uk/rail-news/2014/eurostar-20th-anniversary/> [Accessed 05/06/2019].
- N.A. 2016. *Great Western Railway cleaners to strike* [Online]. London, UK: British Broadcasting Corporation. Available: <https://www.bbc.co.uk/news/uk-england-38288042> [Accessed 23/05/2019].
- N.A. 2018. *Intercity Express door failures frustrate passengers* [Online]. Wellinborough, UK: Railnews. Available: <https://www.railnews.co.uk/news/2018/02/15-intercity-express-door-failures-frustrate.html> [Accessed 23/05/2019].
- NETWORK RAIL. 2013. *London North Eastern Route Sectional Appendix* [Online]. San Francisco (CA), USA: Internet Archive. Available: <https://web.archive.org/web/20131205133159/http://www.networkrail.co.uk/browse%20documents/sectional%20appendix/london%20north%20eastern%20sectional%20appendix.pdf> [Accessed 10/03/2019].
- NETWORK RAIL. 2015. *London North Eastern Route Sectional Appendix* [Online]. San Francisco (CA), USA: Internet Archive. Available: <https://web.archive.org/web/20150709045213/http://www.networkrail.co.uk/browse%20documents/sectional%20appendix/london%20north%20eastern%20sectional%20appendix.pdf> [Accessed 10/03/2019].

- NETWORK RAIL. 2016. *East Midlands Route Study* [Online]. London, UK. Available: <https://cdn.networkrail.co.uk/wp-content/uploads/2016/11/East-Midlands-Route-Study.pdf> [Accessed 20/05/2019].
- NETWORK RAIL. 2018. *Railway Upgrade Plan 2017/2018* [Online]. London, UK. Available: <https://cdn.networkrail.co.uk/wp-content/uploads/2017/08/Railway-Upgrade-Plan-Update-2017-2018.pdf> [Accessed 7/11/2018].
- NETWORK RAIL. 2019. *National Electronic Sectional Appendix* [Online]. London, UK: Network Rail,. Available: <https://www.networkrail.co.uk/industry-commercial-partners/information-operating-companies/national-electronic-sectional-appendix/> [Accessed 13/05/2019].
- NETWORK RAIL. c2017. *Open data feeds* [Online]. London, UK. Available: <https://www.networkrail.co.uk/who-we-are/transparency-and-ethics/transparency/open-data-feeds/> [Accessed 13/04/2019].
- NETWORK RAIL. c2019. *Passenger service success continues for Team Victoria* [Online]. London, UK. Available: <https://www.networkrail.co.uk/passenger-service-success-continues-for-team-victoria/> [Accessed 07/06/2019].
- NIPPON. 2018. *Japan's Most Congested Train Lines* [Online]. Tokyo, Japan: The Nippon Communications Foundation. Available: <https://www.nippon.com/en/features/h00255> [Accessed 23/10/2018].
- OFFICE OF RAIL AND ROAD. 2019. *UK rail industry financial information 2017-18* [Online]. London, UK. Available: https://orr.gov.uk/_data/assets/pdf_file/0013/40351/uk-rail-industry-financial-information-2017-18.pdf [Accessed 17/07/2019].
- OPENTRACK RAILWAY TECHNOLOGY LTD. n.d. *OpenTrack Home Page* [Online]. Zurich, Switzerland. Available: http://www.opentrack.ch/opentrack/opentrack_e/opentrack_e.html [Accessed 28/06/2016].
- PARAMITA, P., Zheng, Z., Haque, M.M., Washington, W. & Hayland, P. 2018. User satisfaction with train fares: A comparative analysis in five Australian cities. *PLoS ONE*, 13, 6, 1-26.
- PELUFFO, N. 2018. *Rail passenger numbers and crowding on weekdays in major cities in England and Wales: 2017* [Online]. London, UK: Department for Transport. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/728526/rail-passengers-crowding-2017.pdf [Accessed 29/12/2018].
- PETTIT, G. 2017. *Network Energy Strategy: Rolling Stock Gap Analysis*. Masters of Science in Railway Systems Engineering and Intergration, The University of Birmingham.

- PIU, F. & SPERANZA, M. G. 2014. The locomotive assignment problem: a survey on optimization models. *International Transactions in Operational Research*, 21, 327-352.
- POWELL, J. P. & PALACÍN, R. 2015. A comparison of modelled and real-life driving profiles for the simulation of railway vehicle operation. *Transportation Planning and Technology*, 38, 78-93.
- PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. 1992. *Numerical recipes in FORTRAN: the art of scientific computing*, Cambridge University Press.
- PRESTON, J., PRITCHARD, J. & WATERSON, B. 2017. Train Overcrowding Investigation of the Provision of Better Information to Mitigate the Issues. *Transportation Research Record*, 2649, 1-8.
- PRESTON, J. M., BLAINEY, S. P., WARDMAN, M., CHINTAKAYALA, P. K., HEYWOOD, C., SHELDON, R. & WALL, G. T. 2008. The effect of station enhancements on rail demand. *European Transport Conference*. Leeuwenhorst, Netherlands: Association for European Transport.
- QI, J. G., YANG, L. X., GAO, Y., LI, S. K. & GAO, Z. Y. 2016. Integrated multi-track station layout design and train scheduling models on railway corridors. *Transportation Research Part C - Emerging Technologies*, 69, 91-119.
- QIN, F. 2014. Investigating the In-Vehicle Crowding Cost Functions for Public Transit Modes. *Mathematical Problems in Engineering*, 2014, 1-13.
- RAIL DELIVERY GROUP. 2017. *£5bn regeneration of Britain's rail stations making local economies stronger* [Online]. London, UK. Available: <https://www.raildeliverygroup.com/media-centre/press-releases/2017/469772907-2017-06-22.html> [Accessed 10/06/2019].
- RAIL DELIVERY GROUP, EVERSOLT RAIL GROUP, ANGEL TRAINS, PORTERBROOK LEASING & SMBC LEASING. 2018. *Long Term Passenger Rolling Stock Strategy for the Rail Industry - Foreword* [Online]. London, UK: Rail Delivery Group. Available: https://www.raildeliverygroup.com/files/Publications/2018-03_long_term_passenger_rolling_stock_strategy_6th_ed.pdf [Accessed 15/11/2018].
- RAIL MANAGEMENT CONSULTANTS GMBH. c2019. *RailSys Suite* [Online]. Berlin, Germany. Available: <https://www.rmcon.de/railsys-en/railsys-suite/> [Accessed 28/06/2016].
- RAIL SAFETY AND STANDARDS BOARD. 2015. *GO/GN3637 Guidance on Defective On-Train Equipment Issue 2* [Online]. London, UK. Available: <https://www.rssb.co.uk/rgs/standards/GOGN3637%20Iss%202.pdf> [Accessed 23/05/2019].
- RAIL SAFETY AND STANDARDS BOARD. 2017. *Station duties and train dispatch - Issue 5* [Online]. London, UK. Available: <https://www.rssb.co.uk/rgs/rulebooks/GERT8000-SS1%20Iss%205.pdf> [Accessed 01/03/2019].

- RAIL SAFETY AND STANDARDS BOARD. c2019. *Rail industry readiness levels* [Online]. London, UK. Available: <https://www.rssb.co.uk/rail-technical-strategy/rail-industry-readiness-levels> [Accessed 26/06/2019].
- RAIL TECHNOLOGY MAGAZINE. 2014. *Eurostar unveils first new 200mph e320 train* [Online]. Manchester, UK: Cognitive Publishing. Available: <http://www.railtechnologymagazine.com/Rail-News/eurostar-unveils-first-new-200mph-e320-train> [Accessed 05/06/2019].
- RAILWAY GAZETTE INTERNATIONAL. 2005. *News in Brief 1st July 2015* [Online]. San Francisco (CA), USA: Internet Archive. Available: <https://web.archive.org/web/20121104135058/http://www.railwaygazette.com/news/single-view/view/news-in-brief-20.html> [Accessed 05/06/2019].
- RAILWAY GAZETTE INTERNATIONAL. 2018. *Joint venture to bid for HS2 rolling stock contract* [Online]. Sutton, UK: DVV Media International. Available: <https://www.railwaygazette.com/news/single-view/view/joint-venture-to-bid-for-hs2-rolling-stock-contract.html> [Accessed 05/06/2019].
- RAILWAY MUSEUM. n.d. *Eurostar: the UK's fastest train* [Online]. York, UK: National Railway Museum. Available: <https://www.railwaymuseum.org.uk/whats-on/eurostar-uks-fastest-train> [Accessed 07/06/2019].
- RAILWAY TECHNOLOGY. c2019a. *Eurostar e320 High-Speed Train* [Online]. London, UK: Verdict Media. Available: <https://www.railway-technology.com/projects/eurostar-e320-high-speed-train/> [Accessed 10/06/2019].
- RAILWAY TECHNOLOGY. c2019b. *High Speed 2 (HS2) Railway* [Online]. London, UK: Verdict Media. Available: <https://www.railway-technology.com/projects/high-speed-2-hs2/> [Accessed 1/06/2019].
- RASMUSSEN, C.E. & WILLIAMS, C.K.I. 2006. *Gaussian Processes for Machine Learning*. Cambridge (MA), USA, MIT Press.
- RMT. 2017. *RMT Cleaners to Strike on Southeastern* [Online]. London, UK: The National Union of Rail, Maritime and Transport Workers. Available: <https://www.rmt.org.uk/news/rmt-cleaners-to-strike-on-southeastern/> [Accessed 23/05/2019].
- SAMA, M., PELLEGRINI, P., D'ARIANO, A., RODRIGUEZ, J. & PACCIARELLI, D. 2016. Ant colony optimization for the real-time train routing selection problem. *Transportation Research Part B - Methodological*, 85, 89-108.
- SCHEEPMAKER, G. M., GOVERDE, R. M. P. & KROON, L. G. 2017. Review of energy-efficient train control and timetabling. *European Journal of Operational Research*, 257, 355-376.
- SCHULTZ, L. & SOKOLOV, V. 2018. Bayesian Optimization for Transportation Simulators. *Procedia Computer Science*, 130, 973-978.
- SEEGER, M. 2004. Gaussian Processes for machine learning. *International Journal of Neural Systems*. 14, 2, 69-106.

- SELS, P., DEWILDE, T., CATTRYSSE, D. & VANSTEENWEGEN, P. 2016. Reducing the passenger travel time in practice by the automated construction of a robust railway timetable. *Transportation Research Part B - Methodological*, 84, 124-156.
- SHAHRIARI, B., SWERSKY, K., WANG, Z., ADAMS, R. P. & DE FREITAS, N. 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104, 148-175.
- SIEMENS MOBILITY. 2016. *Eurostar e320 high-speed trains for Eurostar International Limited* [Online]. Munich, Germany: Siemens AG. Available: <https://www.mobility.siemens.com/mobility/global/SiteCollectionDocuments/en/rail-solutions/high-speed-and-intercity-trains/velaro/velaro-e320-en.pdf> [Accessed 07/06/2019].
- SMALL, K. 1982. The Scheduling of Consumer Activities: Work Trips. *The American Economic Review*. 72, 3, 467-469.
- SMALE, K. 2019. *Grayling awaits new HS2 delivery plan* [Online]. Brentford, UK: New Civil Engineer. Available: <https://www.newcivilengineer.com/latest/grayling-awaits-new-hs2-delivery-plan/10042879.article> [Accessed 06/06/2019].
- SNOEK, J., LAROCHELLE, H. & ADAMS, R. P. 2012. Practical Bayesian optimization of machine learning algorithms. *25th International Conference on Neural Information Processing Systems*. Lake Tahoe (CA), USA: Curran Associates Inc.
- STEER DAVIES GLEAVE. 2017a. *Estimates of Station Usage 2016 - 2017* [Online]. London, UK: Office of Rail and Road. Available: http://orr.gov.uk/_data/assets/pdf_file/0020/26129/estimates-of-station-usage-2016-17-key-facts.pdf [Accessed 7/11/2018].
- STEER DAVIES GLEAVE. 2017b. *Origin-Destination Matrix 2016/17 Summary Report* [Online]. London, UK: Office of Rail and Road. Available: https://orr.gov.uk/_data/assets/pdf_file/0014/26600/regional-rail-usage-odm-methodological-report-2017.pdf [Accessed 13/15/2019].
- SUMPAVAKUP, C. & KULWORAWANICHPONG, T. 2015. Mult-Train Movement Simulation Using MATLAB Object-Oriented Programming. *Applied Mechanics and Materials*, 763, 153-158.
- SUSILO, Y. O. & CATS, O. 2014. Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. *Transportation Research Part A - Policy and Practice*, 67, 366-380.
- TAO, T. & KE-PING, L. 2007. Traffic Modelling for Moving-Block Train Control System. *Communications in Theoretical Physics*, 47, 601-606.
- TECHNICAL LEADERSHIP GROUP. 2017. *Rail Technical Strategy Capability Delivery Plan* [Online]. London, UK: Rail Safety and Standards Board. Available: <https://www.rssb.co.uk/rts/Documents/2017-01-27-rail-technical-strategy-capability-delivery-plan-brochure.pdf> [Accessed 7/11/2018].

- THE FRIENDS OF DERWENT VALLEY LINE. 2019. *Achievements* [Online]. Matlock, UK. Available: <http://www.friendsdvl.btck.co.uk/Achievements> [Accessed 21/05/2019].
- TNS POLITICAL & SOCIAL. 2013. *Europeans' satisfaction with rail services* [Online]. Brussels, Belgium: The European Commission. Available: http://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_382a_en.pdf [Accessed 08/02/2019].
- TRANSPORT FOCUS. 2016. *National Rail Passenger Survey - Spring 2016 Main Report* [Online]. London, UK. Available: <http://d3cez36w5wymxj.cloudfront.net/wp-content/uploads/2016/06/29162541/TF-NPRS-Spr16-pages-ALL-WEB-v4.pdf> [Accessed 08/07/2019].
- TRANSPORT FOR LONDON. 2012. *Join in the celebrations across the capital this summer with a limited edition Summer Oyster card* [Online]. London, UK: Transport for London. Available: <https://tfl.gov.uk/info-for/media/press-releases/2012/june/join-in-the-celebrations-across-the-capital-this-summer-with-a-limited-edition-summer-oyster-card> [Accessed 4/2/2019].
- TROTTER, M., LIU, G. Y. & WOOD, T. 2017. Into the Storm: Describing Optimal Configurations using Genetic Algorithms and Bayesian Optimization. *IEEE 2nd International Workshops on Foundations and Applications of Self* Systems*. Trento, Italy: IEEE.
- UMILIACCHI, S. Improving railway operations through the intergration of macroscopic and microscopic modelling with optimisation. School of Electronic, Electrical and Systems Engineering, 2016 Birmingham, UK. The University of Birmingham, 128.
- UNIVERSITY OF SHEFFIELD. 2017. *Sheffield and Siemens next generation transport modelling* [Online]. Sheffield, UK. Available: https://www.sheffield.ac.uk/dcs/latest-news/transport_modelling-1.685272 [Accessed 19/11/2018].
- VAN WEE, B. 2007. Rail Infrastructure: Challenges for Cost-Benefit Analyses and Other ex ante Evaluations. *Transportation Planning and Technology*, 30, 1, 31-48.
- VANSTEENWEGEN, P. & VAN OUDHEUSDEN, D. 2007. Decreasing the passenger waiting time for an intercity rail network. *Transportation Research Part B - Methodological*, 41, 478-492.
- VISENTINI, M. S., BORENSTEIN, D., LI, J.-Q. & MIRCHANDANI, P. B. 2014. Review of real-time vehicle schedule recovery methods in transportation services. *Journal of Scheduling*, 17, 541-567.
- VUCHIC, V. R. & NEWELL, G. F. 1968. Rapid Transit Interstation Spacings for Minimum Travel Time. *Transportation Science*, 2, 303-339.

- WALKER, D. A. 2016. JMASM38: Confidence Intervals for Kendall's Tau with Small Samples (SPSS). *Journal of Modern Applied Statistical Methods*, 15, 868-883.
- WANG, Y., TANG, T., NING, B., VAN DEN BOOM, T. J. J. & DE SCHUTTER, B. 2015. Passenger-demands-oriented train scheduling for an urban rail transit network. *Transportation Research Part C - Emerging Technologies*, 60, 1-23.
- WANG, Y., ZHOU, Y. & YAN, X. D. 2019. Optimizing Train-Set Circulation Plan in High-Speed Railway Networks Using Genetic Algorithm. *Journal of Advanced Transportation*, 12.
- WARDMAN, M. 2004. Public transport values of time. *Transport Policy*, 11, 363-377.
- WARDMAN, M. & MURPHY, P. 2015. Passengers' valuations of train seating layout, position and occupancy. *Transportation Research Part A - Policy and Practice*, 74, 222-238.
- WARDMAN, M. & WHELAN, G. 1998. Rolling stock quality - improvements and user willingness to pay. Institute of Transport Studies, University of Leeds.
- WASSERSTEIN, R. L. & LAZAR, N. A. 2016. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70, 129-133.
- WATCHDOG. 2017. *Train Hygiene* [Online]. London, UK: British Broadcasting Corporation. Available: <https://www.bbc.co.uk/programmes/articles/17TDg18K1hMCdTHv66n3mLZ/train-hygiene> [Accessed 23/05/2019].
- WATSON, P. L. 1973. *The value of time : behavioral models of modal choice*, Lexington (MA), USA, Lexington Books.
- WEI, L. & YUAN, Z. Z. 2017. A Robust Timetabling Model for a Metro Line with Passenger Activity Information. *Information*, 8, 20.
- WORSLEY, T. 2012. *Rail Demand Forecasting Using the Passenger Demand Forecasting Handbook* [Online]. London, UK: RAC Foundation. Available: <https://www.racfoundation.org/wp-content/uploads/2017/11/pdfh-worsley-dec2012.pdf> [Accessed 14/01/2020].
- XU, X. M., LI, K. P. & LI, X. 2016. A multi-objective subway timetable optimization approach with minimum passenger time and energy consumption. *Journal of Advanced Transportation*, 50, 69-95.
- YANG, X.-S. 2008. *Introduction to mathematical optimization: from linear programming to metaheuristics*, Cambridge, UK, Cambridge International Science Publishing.
- YANG, X.-S. 2010. *Engineering optimization: an introduction with metaheuristic applications*, Hoboken (NJ), USA, Wiley.

- YAO, X., ZHAO, P. & QIAO, K. 2013. Simulation and evaluation of urban rail transit network based on multi-agent approach. *Journal of Industrial Engineering and Management*, 6, 367-379.
- YIN, J., TANG, T., YANG, L., XUN, J., HUANG, Y. & GAO, Z. 2017. Research and development of automatic train operation for railway transportation systems: A survey. *Transportation Research Part C - Emerging Technologies*, 85, 548-572.
- ZHAO, N. 2013. *Railway traffic flow optimisation with different control systems*. Doctor of Philosophy, The University of Birmingham.
- ZHU, Y. T., MAO, B. H., BAI, Y. & CHEN, S. K. 2017. A bi-level model for single-line rail timetable design with consideration of demand and capacity. *Transportation Research Part C -Emerging Technologies*, 85, 211-233.

List of personal communications

David Fletcher. Professor - Railway Engineering, The University of Sheffield.
'*Measured train movement data*', File delivered to: Hickish, B. 2015.

David Gill. Railway Signalling and Train Control Systems Engineer – Siemens
Rail Automation Holdings Limited. '*RE: Train Data for S2AM project*', Email to:
Hickish, B. 10th Oct, 2016.

Emma Davies. Customer Experience Strategy Manager - East Midlands Trains.
'*Possible collaboration between The University of Sheffield and East Midlands
Trains*', Interview conducted with: Hickish, B., Fletcher, D., Harrison, R. 15th June,
2017.

Gemma Nicholson. Research Fellow – Railway Operations Analysis, The
University of Birmingham. '*RE: Davis formula Coefficients*', Email to: Hickish, B.
5th September, 2016.

Nigel Kelt. Transport Planner – Strategy and Service Development, London
Underground Limited. '*RE: LU Crowding and Left Behind*', Email to: Hickish, B.
3rd December, 2015.

Appendix I

This thesis has resulted in the following three academic papers:

Paper A 'A rail network performance metric to capture passenger experience', published in the Journal of Transport Planning and Management (Hickish et al., 2019).

Paper B 'Maximising passenger satisfaction through optimised train movements', presented at the 2017 Stephenson Conference (Hickish et al., 2017).

Paper C 'Investigating Bayesian Optimization for rail network optimization', accepted for publication in the International Journal of Rail Transportation.

For paper A, the author of this thesis created the testable hypothesis, designed the experiments, created the experimental code, conducted the experiments and drafted the manuscript. The author of this thesis revised later versions of the paper manuscript with support from the co-authors of the paper, David Fletcher and Robert Harrison.

For paper B, the author of this thesis designed the experiments, created the experimental code, conducted the experiments and drafted the manuscript. The author of this thesis revised later versions of the paper manuscript with support from the co-authors of the paper, David Fletcher and Robert Harrison.

For paper C, with the co-authors of the paper, David Fletcher and Robert Harrison, the author of this thesis was involved in the creation of the testable hypothesis. The author of this thesis created the experimental design, created the experimental code, conducted the experiments and drafted the manuscript. The author of this thesis revised later versions of the paper manuscript with support from the co-authors of the paper, David Fletcher and Robert Harrison.

Appendix II

Demonstration of expected power fit in Section 4.5.1

In the idealised case where acceleration is constant, the standard kinematic equation applies.

$$0 = \frac{1}{2}at^2 + ut - s \quad (8.1)$$

Where a denotes the acceleration, t , time, u , initial velocity and s displacement. Substituting Newton's second law into (8.1), then comparing to the general form of a quadratic equation and the quadratic formula, shown by (8.2) and (8.3) respectively, gives (8.4) which describes the time required for an object to travel distance s from stationary when a constant force F is applied.

$$bx^2 + cx + d = 0 \quad (8.2)$$

$$x = \frac{-c \pm \sqrt{c^2 - 4bd}}{2b} \quad (8.3)$$

$$t = \sqrt{\frac{2sm}{F}} \quad (8.4)$$

where m is mass and b , c and d are coefficients. If two forces are related by (8.5) where ζ denotes a scaling factor, then plotting the change in time from t_1 to t_2 on the y -axis (as a decimal percentage) gives (8.6):

$$F_2 = \zeta F_1 \quad (8.5)$$

$$y = \frac{t_2}{t_1} - 1 \quad (8.6)$$

which by substituting (8.4) and (8.5) into (8.6), gives:

$$y = \sqrt{\zeta}(F_1 - \sqrt{F_1}) \quad (8.7)$$

If ζ is plotted on the x-axis, there is the equation of the form $y = ex^h$ where $e = (F_1 - \sqrt{F_1})$ and $h = 1/2$.

Parameter values for different trains referred to in Section

4.5.3

TDM parameters relating to the exact train are not always available, however parameters for a similar train can be used. Table 19 shows the TDM parameters for two pairs of similar trains, i.e. High Speed Trains and Commuter Trains. The InterCity 125 data provided by Nicholson (personal communication, 2016)¹⁰, can be compared against data for a Class 390 Pendolino, provided by Zhao (2013). Data provided by a Railway Signalling and Train Control Systems Engineer at Siemens (Gill, 2016)¹⁰ in a personal communication can be compared for two formations of a Desiro commuter train. In both cases of comparison between similar trains, it can be seen that all parameters are within a factor of two of each other.

Parameter	High Speed Trains		Commuter Trains	
	InterCity 125	BR Class 390 Pendolino	Desiro 12 unit formation	Desiro 8 unit formation
Effective mass, M_E , (tonnes)	489	510	492	334
Boundary speed, v_0 (km/hour)	55	88	5	10
Maximum tractive force, F_{max} , (kN)	170	203	631	421
Davis formula A (kN)	3.22	5.42	5.41	3.67
Davis formula B (kNs/m)	0.113	0.069	0.0083	0.0056
Davis formula C (kNs ² /m ²)	0.0078	0.0142	1.40×10^{-4}	1.05×10^{-4}

Table 19 – Input parameters for the TDM for four different trains

Discussion of signalling systems referred to in Section 4.2

In simulation experiments comparing the ability of networks to recover from delay under different signalling systems, Koning (2002) found a 'small' difference between a fixed block and moving block system for a range of sub-

¹⁰ For more information, see the list of personal communications in the reference list.

networks of the Dutch railway. Using the results published by Koning to compare the percentage of trains that recover from delay under the two signalling systems, it can be estimated that the mean difference is 4% with a standard deviation of 3%. Because Koning's experiment ignores the influence of traffic controllers who allocate resources (e.g. junctions) when there is competition between trains, the main method captured for networks to recover from delay is to reduce the interstation run time of trains. Therefore, the result of Koning indicates that the interstation run time is similar between moving block and fixed block signalling systems. Whilst researchers such as Grimm and Sandblad (2000) have shown that a moving block system increases the train carrying capacity of the network, this does not necessarily change the interstation run times. In simulation experiments of a 35.9 km section of the West Coast Mainline, the results of Zhao (2013) indicate that a moving block signalling system would reduce journey times by up to 6% compared with a 4-aspect fixed block system. For the accuracy required by PRaM at this stage in the development of SUPREME, this magnitude in error of predicted journey times is considered acceptable, therefore PRaM can be used to model real-world networks with fixed block and moving block signalling except in cases where maximising the train carrying capacity of the network is the objective.

Appendix III

Determining Davis formula coefficients used in Chapter 6

To obtain Davis formula coefficient values for the Class 153, 395 and 373 trains, speed dependent rolling resistance data relating to each train type was extracted from the BraVE simulator, which is described by Umiliacchi (2016), and fitted to the Davis formula described in Chapter 4. The data and fit is plotted in Figure 44 for the Class 153 and Figure 45 for the Class 395 and 373. No measure of uncertainty in the speed or resistance data was available, so error bars are not plotted. Observing Figure 44, it can be seen that the data point for the lowest speed does not fit well to a quadratic relationship. This is because it relates to the static starting resistance of the train and for simplicity it data point has been removed from the fit and the first-order term in the Davis formula (B) has been forced to be greater than or equal to 0. Observing Figure 45 shows that the Davis formula fits well to all the data points for both the Class 395 and 373 data. However, it can be seen that there is no data for either train operating above 75% of their maximum speed. This may be because of the difficulty in finding suitable locations or methods to collect this data. Nonetheless, it is assumed that the relationship seen at lower speeds can be extrapolated to higher speeds. Although errors may be introduced to the Davis formula parameters by the lack of data over the whole speed range, Chapter 4 has shown that the accuracy of the PRaM model is relatively insensitive to changes in the Davis formula coefficients compared to inaccuracies introduced by not representing all traffic on the line, for example. Consequently, the Davis formula coefficients derived by converting the fit parameters to SI units are considered acceptable for use in this thesis and are shown in Table 20.

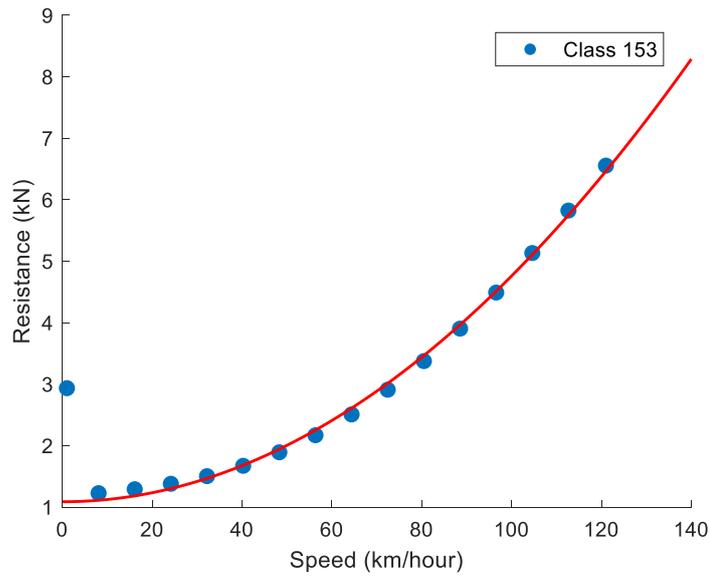


Figure 44 – Resistance data for a Class 153 train, with fits. The data has been extracted from the BraVE simulator described by Umiliacchi (2016). The data point for the lowest speed has been excluded from the fit and the first order fit parameter has been forced to be greater than or equal to 0.

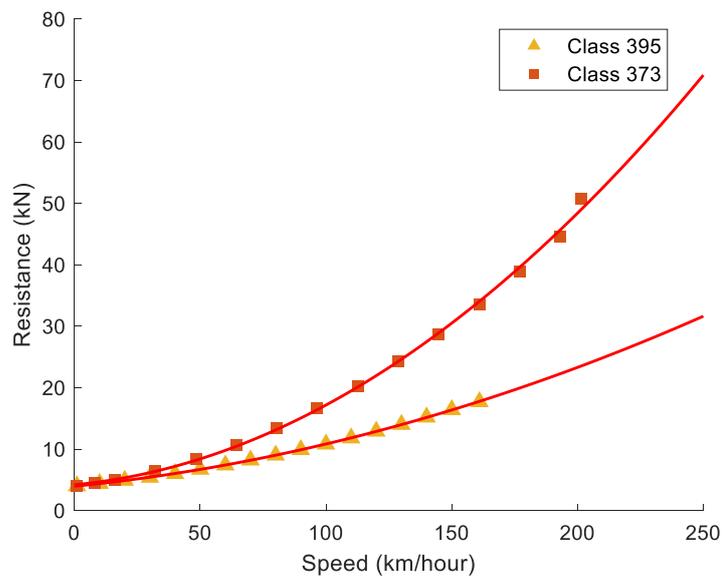


Figure 45 – Resistance data for a Class 395 and Class 373 train, with fits. The data has been extracted from the BraVE simulator described by Umiliacchi (2016).

	\mathcal{A} (kN)	\mathcal{B} (Nsm ⁻¹)	\mathcal{C} (Ns ² m ⁻²)
Class 153	1.09	1.56×10^{-10}	4.76
Class 395	3.95	1.47×10^2	33.62
Class 373	4.20	1.38×10^2	11.8

Table 20 - The Davis formula coefficient values derived from the fits shown in Figure 44 and Figure 45.

High Speed Two savings functions used in Section 6.3

The case study in Section 6.3 investigates the savings-benefit trade-off of using rolling stock with performance parameters equal to trains already in production. Consequently, to ensure comparison is like-for like, the cost of in-production trains must be scaled to capture the number of additional rolling stock units necessary to meet the capacity requirement specified by HS2 Limited. To do so the cost of all the trains in the network, \mathbb{C}_A , is calculated with the formula:

$$\mathbb{C}_A = N\mathbb{C}_S \quad (9.1)$$

where N denotes the number of trains and \mathbb{C}_S the scaled cost of a single train. In the case study the value of N is provided by HS2 Limited (2017) as 54. The value of \mathbb{C}_S is given by the formula:

$$\mathbb{C}_S = \mathbb{C}_P \left(\frac{\mathbb{P}_R \mathfrak{m}}{\mathbb{P}} + \mathbb{L}\mathbb{M} \right) \quad (9.2)$$

where \mathbb{C}_P denotes the cost for a single passenger carriage, \mathbb{P}_R , the required passenger capacity of a single train, \mathfrak{m} , the number carriages in the train, \mathbb{L} , the number of power cars in the train and \mathbb{M} a scaling factor between the cost of a single passenger carriage and a single power car. In the case study, \mathbb{P}_R has a value of 950 pax (HS2 Limited, 2019) and \mathbb{L} and \mathbb{M} both have a value of 2. The value of \mathbb{C}_P is given by the formula:

$$\mathbb{C}_P = \frac{\mathbb{C}_T}{\mathfrak{m} - \mathbb{L}(1 + \mathbb{M})} \quad (9.3)$$

where \mathbb{C}_T denotes the cost for one real-world train set. For each train Table 21 displays the values used to calculate \mathbb{C}_T , the values of \mathfrak{m} and \mathbb{P} , data sources and the value of \mathbb{C}_A . For the case study in Section 6.3, the savings associated with each train type are calculated by subtracting the value of \mathbb{C}_A relating to a certain train type from the value of \mathbb{C}_A relating to the HS2 Design Specification trains.

	Cost in 2019 ¹¹ (million £)	Number of sets	Passenger carriages per set, m	Passenger capacity per set, p	Cost for 54 scaled trains, C_A , (billion £)
Class 395	370 (Railway Gazette International, 2005)	28 (Railway Gazette International, 2005)	4 (Mochida et al., 2010)	340 (Mochida et al., 2010)	1.29
Class 373 (regional set)	198 (Baldwinson, 2015)	7 (Baldwinson, 2015)	14 (n.a, 2002)	560 (n.a, 2014)	2.24
Class 374	333 (Rail Technology Magazine, 2014)	7 (Rail Technology Magazine, 2014)	14 (Rail Technology Magazine, 2014)	900 (Rail Technology Magazine, 2014)	2.55
HS2 Design Specification	2870 (Railway Gazette International, 2018)	54 (HS2 Limited, 2017)	n/a	950 (HS2 Limited, 2019)	2.87

Table 21 – Parameter values used with formulas (9.1), (9.2) and (9.3) to calculate the cost associated with using different train types for the service modelled in Section 6.3. The right-hand column shows the cost for 54 trains.

¹¹ Converted to 2019 prices using an inflation calculator provided by Alioth Finance (c2019).