Railway Track Capacity: Measuring and Managing

Khadem Sameni, Melody

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Transportation Research Group

Railway Track Capacity: Measuring and Managing

by

Melody Khadem Sameni

Thesis for the degree of Doctor of Philosophy

October 2012

Abstract

This thesis adopts a holistic approach towards railway track capacity to develop methodologies for different aspects of defining, measuring, analysing, improving and controlling track capacity utilisation. Chapter 1 presents an overview of the concept of capacity and the railway capacity challenge is explained. Chapter 2 focuses on past approaches to defining and analysing the concept of railway capacity. Existing methods for estimating capacity utilisation are studied in four categories: analytical methods, parametric models, optimisation and simulation.

Chapter 3 examines various factors affecting capacity utilisation. Chapter 4 develops the systems engineering foundation toward railway capacity. From process improvement methods, Six Sigma and its Define, Measure, Analyse, Improve and Control (DMAIC) cycle is chosen as the underlying framework of the thesis.

Chapter 5 defines lean, micro and macro capacity utilisation based on the discrete nature of railway capacity. Data Envelopment Analysis (DEA) is used to develop two novel methodologies to analyse lean capacity utilisation. A DEA model analyses relative efficiency of train operating companies based on their efficiency to transform allocated train paths (timetabled train kilometres) and franchise payments to passenger-kilometres while avoiding delays. A case study demonstrates its application to 16 train operating companies in the UK. The operational efficiency of stations is benchmarked from similar studies for ports and airports. Two models are developed for analysing technical efficiency and service effectiveness. 96 busiest stations in Great Britain are analysed by this method.

For analysing capacity utilisation in the freight sector, the concept of 'profit-generating capacity' is introduced in chapter 6. It is applied in an American freight case study to choose between bulk and intermodal trains in a heterogeneous traffic. DEA is also used in another case study for identifying the most profitable commodities.

Chapter 7 suggests using variation reduction and failure mode and effect analysis (FMEA) to control capacity utilisation. For improving railway capacity utilisation it is suggested to find and improve the weakest line section, the weakest trains and the weakest station. A real world case study of the South West Main Line in Great Britain, demonstrates applying these aspects. For finding the weakest line section two existing methods of the UIC 406 and the CUI method are compared with each other. For finding the weakest trains a meso index is suggested. It can identify which trains can be removed to free up some capacity in the busiest section of the line. Simulating delays and removing the highest delay causing trains is another method suggested. The weakest stations are identified by applying the DEA methodology developed in chapter 5. Target values for train stops at each station are suggested to be fed to the tactical timetabling.

It is concluded that developing methodologies to analyse, improve and control railway capacity utilisation is needed and the methodologies proposed in this thesis can be a stepping stone towards them.

To

My parents, God's angels in my life

and

to

Professor Preston, my parent in the academia

Acknowledgements

The acknowledgments page is my favourite page of all in books and theses. It provides an opportunity to be grateful to all the people who have played great roles in our lives and remember that we are and can do nothing without social interactions.

The most difficult part of this thesis is finding an appropriate way of thanking my parents. What a futile ambition… They totally devoted their lives to the wellbeing and education of their children (surprisingly two out of three are railway engineers!). They provided us an ocean of unconditional love and caring to be able to swim towards growth. They taught us numerous life lessons to be better human beings in this world. If finishing this thesis puts a single smile on their face, I will be the happiest daughter. I am also very thankful to Mohammad and Mona for being the best brother and sister in the world.

Professor Preston… oh, where to begin! I did a month's full-time research to decide where to do my PhD studies. The conclusion was so strong that I made up my mind either to do a PhD under the supervision of Professor Preston or never to do a PhD. I explained my research passion to him through email in November 2007 and received a very kind reply. There were some months of anxious waiting but my dream finally came true when I gained not only admission but also a full PhD scholarship. He suggested railway capacity as a rising and interesting research topic which cleverly put me on the right track in railway research that better matched my capabilities compared to my previous experiences in pure operations research. This was the beginning of an adventurous train journey that I have been on since October 2008. He is a supervisor sent directly from heaven: he is very organised, provides very quick and meticulous feedback, replies to all emails that need an answer concisely but efficiently within a few hours and is in good temper 99.9 per cent of the times. He is internationally celebrated, profoundly knowledgeable and passionate about railways as well as being one of the greatest human beings I have met in my life.

He provided various opportunities for me to attend conferences and go on research visits to other universities. He gave me a delicate balance of freedom to develop my intellectual curiosity as well as disciplined supervision to keep me on track. In the past four years, the total number of his supporting letters for me exceeds 20. He has made countless contributions to my life. Most importantly he taught me how to face challenges, learn the lesson and he kindly and firmly took my hand whenever needed.

In the beginning of my undergraduate studies in railway engineering, I was not sure whether I would like this major and wanted to change it. The reason I am now writing my PhD thesis is solely due to the brilliant professors of School of Railway Engineering at Iran University of Science and Technology who ignited my passion for railways. My heart-felt sincere thanks to my undergraduate supervisor, Dr. Ghoseiri and my master thesis co-supervisor, Dr. Aghaei for making me interested in railway operations and for all I have learnt from them. Dr. Amin Naseri, my master's supervisor, kindly helped me to bridge between industrial and railway engineering. Dr.Yaghini and Mr. Ahmadi provided some profound insights for my bachelor and master's theses on train scheduling respectively.

My research on railway capacity started just as Dr. Landex finished his PhD thesis on this topic. When I went in search of one of his papers, he generously shared all his research and experiences with me. It led to a joint sponsored research visit to the Technical University of Denmark in October 2009 for me to learn the Railsys software and the UIC 406 capacity method. This visit laid the building blocks of my thesis and resulted in further co-authored papers and cooperation between the two universities. I owe a deep debt of gratitude to him for supporting me just like a co-supervisor. Due to the very limited number of researchers working on railway capacity, he was always there whenever I needed help or was stuck. He kindly spared a lot of time in spite of his extremely busy schedule. Without his help and the synergy of discussing research ideas with him, it would have been utterly impossible to conduct this research. I would like to acknowledge Bernd Schittenhelm as well for being a complement to Dr.Landex whenever he was away.

I am grateful to Dr. Lai from the Taiwan National University who initiated a research visit to the Railroad Engineering Program at the University of Illinois at Urbana Champaign where he did his PhD. Professor Barkan, the director, in every way supported this visit and enabled us to write a mutual proposal to win an Individual Enhancement Award from the University of Southampton to fund this. Professor Barkan, who is the pioneer of railway research and education in the North America, is very down to earth, a hallmark of great people. This is to the extent that he kindly accompanied me to buy a pair of safety shoes at 10 p.m to enable me to attend a railway site visit the following day. He managed this research visit excellently and put me in touch with one of his best master's students working on the topic of railway capacity and train delays. The collaboration with Mark Dingler, currently Manager of Operations Planning at CSX Transportation, was very fruitful and we developed the concept of revenue-generating capacity for freight trains and did a case study in an American context. Getting acquainted with Professor Tsung-Chung Kao, the world's best high speed expert, was an added bonus of this visit. I would also like to thank Riley Edwards, who invited me to a wonderful two day visit to railway infrastructure and facilities in Chicago with his class. Bryan Schlake, Seyed Mohammad Nourbakhsh and Tristan Rickett were wonderful railway companions during the stay.

Dr. Armstrong, my co-supervisor, gave very helpful support in industrial aspects and knowledge of British Railways. He is extremely efficient in email communication and whatever the problem you can be sure to have a reply within an hour. Dr. Blainey is an absolute rising star in railway research. It was a blessing to be with him in the same research group. He kindly provided me with some of his research results on station catchment area population and demand analysis and extracted some necessary data from ArcGIS software.

Jill Taylor, my landlady, was an acting mom as well as a friend through all my ups and downs of life. She provided a very cosy and relaxing accommodation and environment for more than three years. I was away from my family and she made me feel I was not a lodger but part of her family. I was invited to all family occasions, birthdays, and Christmas parties and enjoyed the time being with her warm and welcoming extended family. She kindly proofread all the papers and the whole PhD thesis. I also need to apologise to her as at a certain point there was a lot of the burden of my problems on her.

Data Envelopment Analysis was a pillar of my thesis and I owe learning this method to Dr. Emrouznejad at Aston University. I just had a serendipitous encounter to it and thought it would be useful for my research. In search of learning it, I came across him as one of the pioneers of Data Envelopment Analysis education and research. We have never met but he kindly answered all my questions via email in a time span of more than a year.

Elmira, the queen of the kingdom of my friends, has been my best friend over the past twelve years. No matter whether we were living in the same city or continents apart, she was always there to share tears and laughter or to provide lifesaving advices. I would like to extend my thanks to all my friends who have been pieces of my life puzzles at different stages. Mentioning their names here is the least I can do: Alan, Albert, Amith, Asma, Banafsheh, Faezeh, Fatemeh, Gil, Golnaz , Kate, Keira, Ksenija, Louise, Mahnaz, Mahta, Maria, Marjan, Maryam, Mehri, Mohammad Javad, Mojtaba, Nafiseh, Neda, Samira, Sanaz, Sareh, Sedigheh, Shahab, Sharareh, Tim,Yan and Yasser.

I like to thank the external examiner, Professor Pachl, for his invaluable time and comments to improve the thesis as well as his ego boosting encouragements. Dr.Waterson, the internal examiner, is thanked for his comments especially for Transfer viva which raised the bar for the thesis.

During this PhD journey I had the best and worst times of my life ranging from crying from the bottom of my heart to laughing from the bottom of my heart. It was a great medium towards my personal, social, academic and spiritual development. There is no comparison between me now and the person I was four years ago. I hope this piece of research is also of some benefit to the railway industry and other researchers.

Melody Khadem Sameni

Southampton, September 2012

Table of Contents

List of Figures

List of Tables

Declaration of authorship

I, Melody Khadem Sameni declare that the thesis entitled "Railway Track Capacity: Measuring and Managing" and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as referenced in the text

Signed: ………………………………………………………………………..

Date:………………………

1 Introduction

This chapter presents a general overview of the thesis, its aims and scope. The worldwide railway capacity challenge and its origins are briefly discussed and the approach of the thesis toward it is explained. Finally, the aims and scope of the thesis and the research questions are identified.

1.1 An overview of transportation and the concept of used capacity

The main concept of transportation is moving passengers and goods from one place to another. Transportation affects different aspects of human lives from daily individual level to long-term socio-economic welfare and sustainability of societies. Figure 1-1 is a schematic representation of transportation[1] which later on in the thesis will be applied for the concept of used capacity. The transportation of passengers and goods should be performed in an efficient, safe, secure and reliable manner with the lowest external and internal costs possible.

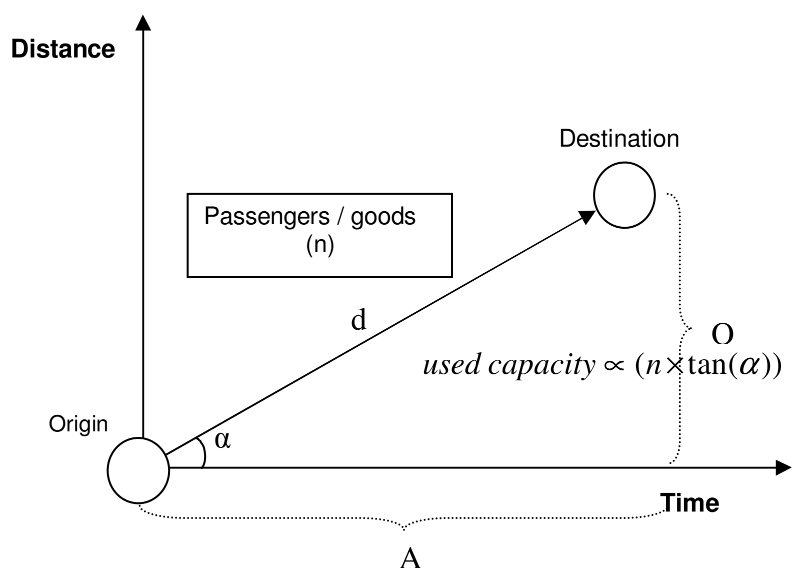


Figure 1-1- A schematic representation of transportation and the concept of capacity- Source: Author's own illustration

As seen in Figure 1-1, $\tan(\alpha) = \frac{O}{A}$ equals speed. Therefore, used capacity is a function of passenger kilometre per hour.

---

[1] This is a very simplified representation of transportation and it uses average speed between origin and destination instead of variable speed. The average speed for long travel distances may be higher than short distance travels.

Early transportation evolved from walking, domesticated animals and river boats to ox-driven carts and horse-drawn stagecoaches after the revolutionary invention of the wheel (Herbst, 2006). Steam power, diesel power, electric power and engines brought about modern transportation. Today, passengers and goods are transported via air, water, roads and rails by various modes of transportation at different socio-economic costs. However, there are some common elements in the different modes of transportation which are vehicles, guideways (feasible paths of travel), terminals and control policy (Hall, 2003).

For each mode of transportation there are vehicles moving on guideways. The practical capacity of these guideways for accommodating vehicles might be limited or abundant. Movements of vehicles on the guideways are controlled to ensure safety. Table 1-1 provides an overview of different modes of transportation.

Table 1-1 - An overview of practical capacity for different modes of transportation (Khadem Sameni et al., 2010a)

	Guideways	**Degrees of freedom for movement on guideways**	**Practical capacity of guideways**	**Bottlenecks**	**Control policy**
Air	- (Air)	3	Abundant	Airports	Air Traffic Control (ATC)
Marine	- (Water)	2	Abundant	Ports/ Locks	Automatic Identification System
Rail	Rails	1	Limited	Stations/ Junctions	Signalling
Road	Treated roads	2	Limited	Junctions	Traffic lights and signs

All modes of transportation, through movement of passengers and goods, affect the socio-economic welfare of societies and are affected by it. As Manheim (1979) puts it, the role of transportation research and analysis "is to intervene, delicately and deliberately, in the complex fabric of a society to use transport effectively, in coordination with other public and private actions, to achieve the goals of that society."

1.2 An overview of railway transportation

Railway transportation is a public mode of transportation which is mainly characterised by steel wheels that run on steel rails[2]. Early uses of this concept were found in mines

---

[2] Trains that use state-of-the-art magnetic levitation technology (Maglev) don't have wheels but still run on rails.

which later on evolved for the transportation of passenger and freight. Table 1-2 summarises the main strengths and weaknesses of railway transportation.

In essence, railways can be defined as “cheap, low-friction guideways” that have combined three important characteristics: reduced friction, reduced cost of low-friction and providing guided way (Armstrong, 1998).

Major turning points in the early development of railways were the invention of the edge-rail by William Jessop in 1789 and the building of the first steam locomotive by Richard Trevithick in 1804 (Westwood, 2009). The first public steam-operated railway opened for traffic on 27 September 1825 between Darlington and Stockton (Rangwala, 1998). Since then, railways have spread all around the world as an energy-efficient and sustainable mode of transportation

Table 1-2 - Railway transportation: Strengths and weaknesses

	Item	**Main reasons**
Strengths	Safety	Restriction in movement to just one axis - Very controlled
	Energy efficiency	Low friction due to steel wheels running on steel rail - Higher passenger/freight km per kilo equivalent of petrol
	Environment friendly	Low CO_2 emissions due to higher passenger/freight km per kilo equivalent of petrol – Low friction and adhesion
	Load handling	Enormous traction power and energy efficiency
	Less land use	High carrying capacity per square metre of infrastructure
Weaknesses	Not door-to-door	Not practical
	Capital intensive industry	Rolling stock, signalling and infrastructure are expensive.
	Extreme dependency on the infrastructure	Trains can only go where the rails go and only have one degree of freedom for movement along the rails.
	Long braking distance	Low friction due to steel wheels running on steel rails
	Noise	Steel wheels running on steel rails

1.3 Railway capacity challenge

Many railways have witnessed huge growth in passenger and freight demand over the last decade. Road congestion, higher fuel costs, rising incomes, privatisation of railways along with concern for sustainability of transport have resulted in enormous growth in rail passenger and freight in the last decade. Growth in railway demand is one side of the coin; the other side of the coin is limited railway infrastructure. Many railways around the world are facing a challenge to accommodate necessary train services on their infrastructure (Association of Train Operating Companies, 2007, Cambridge Systematics, 2007). Although in some cases double, triple or quadruple tracks have been built[3], the capacity of infrastructure in total has not increased proportionately to keep up with the pace of demand. Therefore, a so-called "railway capacity challenge" has emerged which is schematically depicted in Figure 1-2.

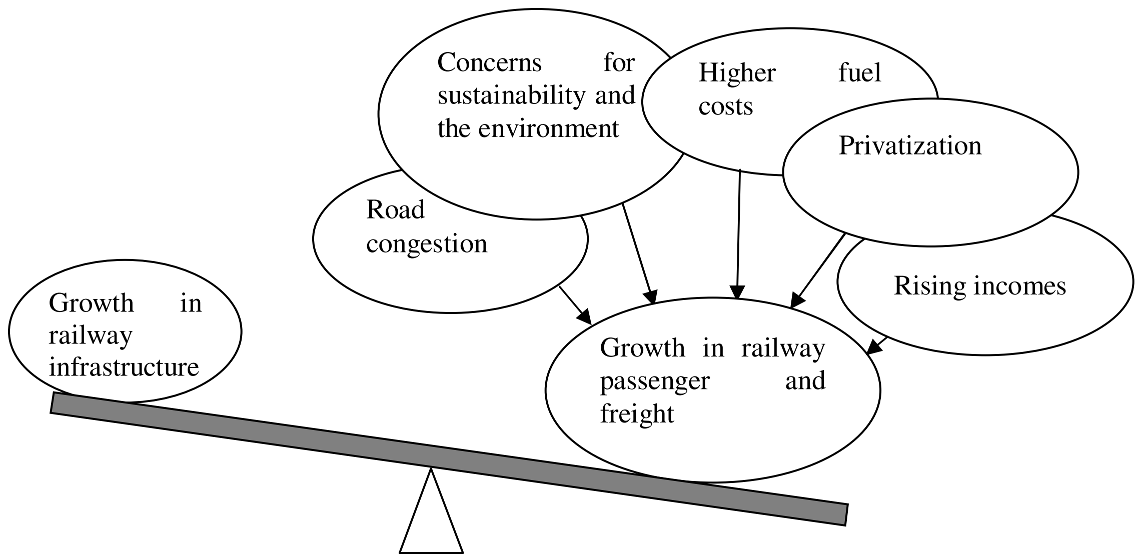


Figure 1-2 - The scale of railway capacity challenge[4]

More specifically, as Table 1-3 suggests, statistics in Europe show a total increase of 32% in tonne-km of freight transported by rail and an 9% increase in rail passenger-km between 2001 and 2010 (UIC, 2011c) . During this time, railway infrastructure has increased by just 5%.

Table 1-3 Growth in rail passenger, freight and infrastructure across Europe. Data source:(UIC, 2011c)

Year	**2001**	**2010**	**2001**	**2010**	**2001**	**2010**
Item	Passenger-km (billion)		Tonne-km (billion)		Line-km	
Value	575.3	626.2	1861.0	2,454.4	353,170	370,387.9
Growth (2010/2001)	+8.86%		+31.88%		+4.88%	

[3] Double track line often quadraples capacity NASH, C. 1982. *Economics of public transport,* London, Longman..

[4] Freight trains are usually profitable whereas passenger trains might not be economically viable without subsidies from governments. Depending on how privatisation is implemented, it can have different effects on the passenger and freight markets.

Similar studies in the continental US show a huge increase in railway demand (a projected increase of 88% in tonnage by 2035 from the 2005 level) hence a capacity challenge as in Figure 1-3 (Cambridge Systematics, 2007). Colour codes and level of service (LOS) are explained in Table 1-4.

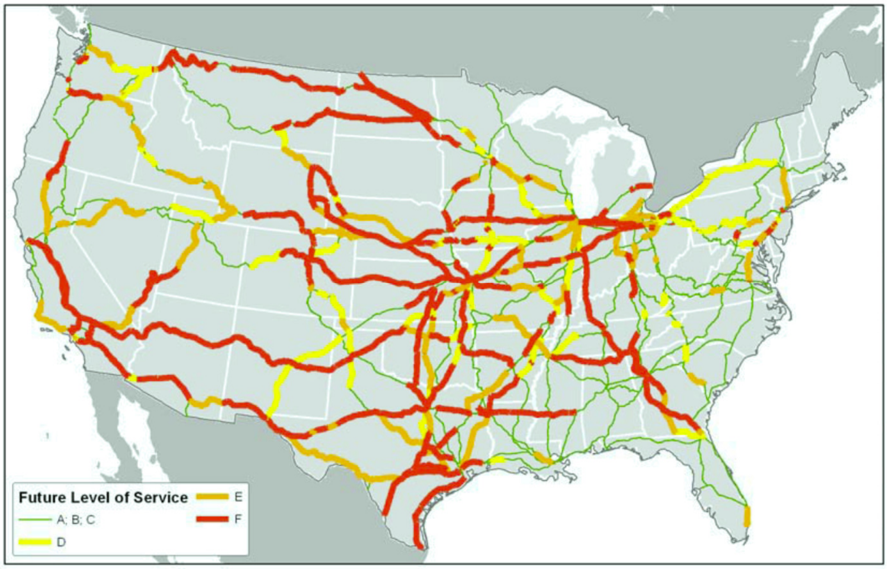


Figure 1-3 - Projected train volumes for continental US in 2035[5] (Cambridge Systematics, 2007)

Table 1-4- Level of service grades (LOS) (Cambridge Systematics, 2007)

	Level of service (LOS)	**Description**		**Volume/capacity ratio**
	A	Below Practical Capacity	Low to moderate train flows with capacity to accommodate maintenance and recover from incidents	0.0 to 0.2
	B			0.2 to 0.4
	C			0.4 to 0.7
	D	Near Practical Capacity	Heavy train flow with moderate capacity to accommodate maintenance and recover from incidents	0.7 to 0.8
	E	At Practical Capacity	Very heavy train flow with very limited capacity to accommodate maintenance and recover from incidents	0.8 to 1.0
	F	Above Practical Capacity	Unstable flows; service breakdown conditions	> 1.00

[5] Red lines indicate 'level F' of service or 'above practical capacity' with unstable traffic flow.

Efficient management and planning for measuring used capacity and taking necessary enhancement measures are needed.

1.4 Need for the study

Transporting 2.7 billion passenger-kilometres and 9.5 billion tonne-kilometres in 2010 (UIC, 2011c), railways worldwide play a major role in the socio-economic welfare of societies. The fact that they offer vital services, with low impact on the environment, macro-economic advantages for society, sustainable integration of different transportation modes and mobility as well as being the safest mode of transportation (UIC, 2011b) make them a priority for many governments.

In order to tackle railway capacity challenge, efficient utilisation of railway infrastructure is critical as building new railway lines is extremely costly and time-consuming. Compared to road transportation which also has limited practical capacity on its main infrastructure, the concept of railway capacity is not well explored. Table 1-5 compares the status of the capacity manuals for these two modes of transportation. As expressed by the Rail Capacity Joint Subcommittee of the Transportation Research Board (TRB), the need for a railway capacity manual is felt (Lindner and Pachl, 2010). Contrary to road transportation, many aspects of railway capacity have not been systematically explored, hence a comprehensive overview of capacity is very much needed to tackle the railway capacity challenge.

Table 1-5 - Capacity manual for road and railway transportation (Khadem Sameni et al., 2011b).

	Road	**Railway**
Name	Highway Capacity Manual	Capacity leaflet
Published by	Transportation Research Board (TRB)	International Union of Railways (UIC)
First edition	1950	2004
Latest edition	2010	2004
Number of pages	1650	24

The concept of railway capacity is more complicated that road capacity. Table 1-6 summarises the issues for railway capacity in comparison with road.

Table 1-6 - Issues for railway capacity in comparison with road capacity (Parts of the first two columns are extracted from the work by Rangwala (1998))

Item	**Railways Vs. Roadways**	**Issue for railway capacity**
Construction costs and maintenance	Very high	Infrastructure expansion plans are extremely expensive and time – consuming.
Cost of transportation	Usually cheaper especially for long distances	A huge potential for railway demand (passenger and freight).
Load handling capacity	With the same amount of fuel can handle more freight	Attractive for carrying freight.
Maintenance	Constant maintenance is needed for railways	Maintenance can interfere with rail operations and decrease the railway capacity.
Tractive resistance	Nearly one-fifth to one-sixth of pneumatic tyre on roads	Much longer braking distances and therefore need for a longer safety distance between two consecutive vehicles.
Operational controls	Signalling in railways and traffic lights in roads	Complex and expensive systems.
Speed	Usually higher	In roads there are different lanes for different speed ranges but in railways heterogeneity of speed is a problem.
Flexibility in case of accidents and delays	Less flexible	Due to having a single degree of freedom for movement, there is a serious domino effect in railways which causes propagation of delays and disturbances.
Terminal operations	More complicated	Railway transportation cannot offer door-to-door transportation and thus many operations must be carried out at terminals. Train formation, locomotive assignment and shunting affect capacity utilisation at stations.
Uniformity of infrastructure and technologies	Less uniform	Track gauge, load gauge, power supply and signalling technologies vary across regions or countries.

The complexity of railway capacity, the lack of a holistic manual, the importance of railways for the socio-economic welfare of the societies and the existing railway capacity challenge underline the need for a comprehensive study of different aspects of railway capacity.

1.5 Aims and scope of the thesis

The main aim of the present thesis is to conduct a comprehensive study of the concept of capacity in railways and to produce a railway capacity manual that can provide guidelines for efficient capacity utilisation. It adopts a holistic and systems approach toward measuring and managing the railway capacity challenge. The main scope of study is railway line planning (Figure 1-4). However, for feasible and realistic line planning, there are interactions with market demand and timetabling that are considered as well.

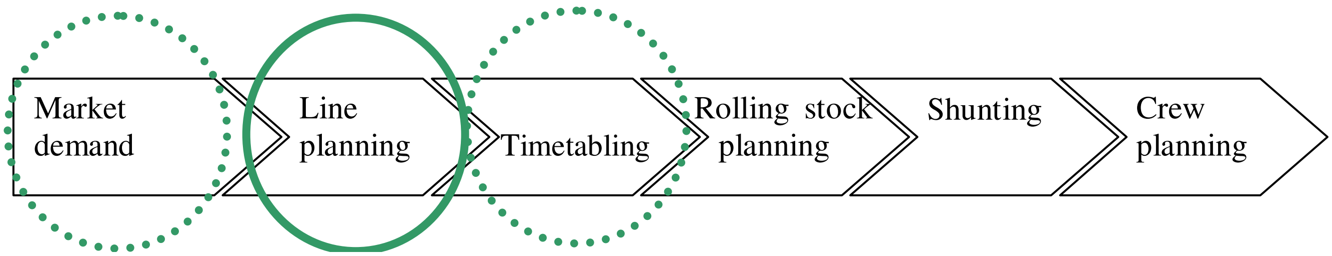


Figure 1-4 - Scope of the thesis in the sequence of railway planning based on Vromans (2004)

As discussed by Van de Velde (1999), there are three main levels in public transportation planning: Strategic covers long term (5 year) goals and deals with what needs to be achieved. These are mainly in the category of transport policy, market share and profitability. Tactical planning is for medium term (1-2 years) and identifies which services can deliver the aims. Detailed service characteristics including timetable, fares and vehicles are addressed at this level. Operational planning is short term (1-6 months) and deals with delivering these services. The approach of the thesis toward the concept of capacity is mainly intended for tactical and medium-term planning of railways while it provides some insights for operational and strategic planning as well.

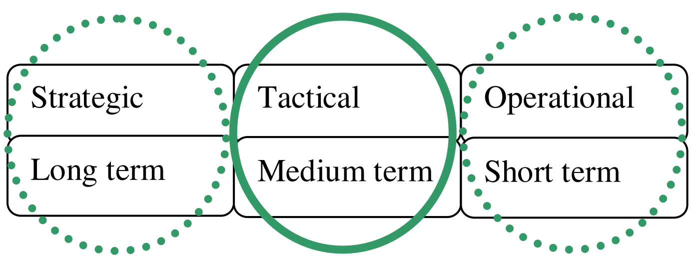


Figure 1-5 - Scope of the thesis in terms of decision level and timeline

1.6 Structure of the thesis

To provide a comprehensive study of different aspects of the railway capacity, a DMAIC improvement cycle used in six-sigma studies (Wisner et al., 2009) is developed as shown in Figure 1-6. After a literature review of previous studies in the field of railway capacity, the following chapters, based on the DMAIC cycle for railway capacity aim to provide methodologies to define, measure, analyse, improve and control capacity utilisation. Several case studies illustrate these methodologies and apply them to real world or simulated examples throughout the thesis. Chapter 5 presents two case studies in the context of Great Britain for capacity utilisation by passenger train operators and at busiest train stations. Two case studies on capacity utilisation analysis for freight trains in American context are done in chapter 6. Various methods for improving capacity utilisation at South West Main Line are discussed in chapter 7.

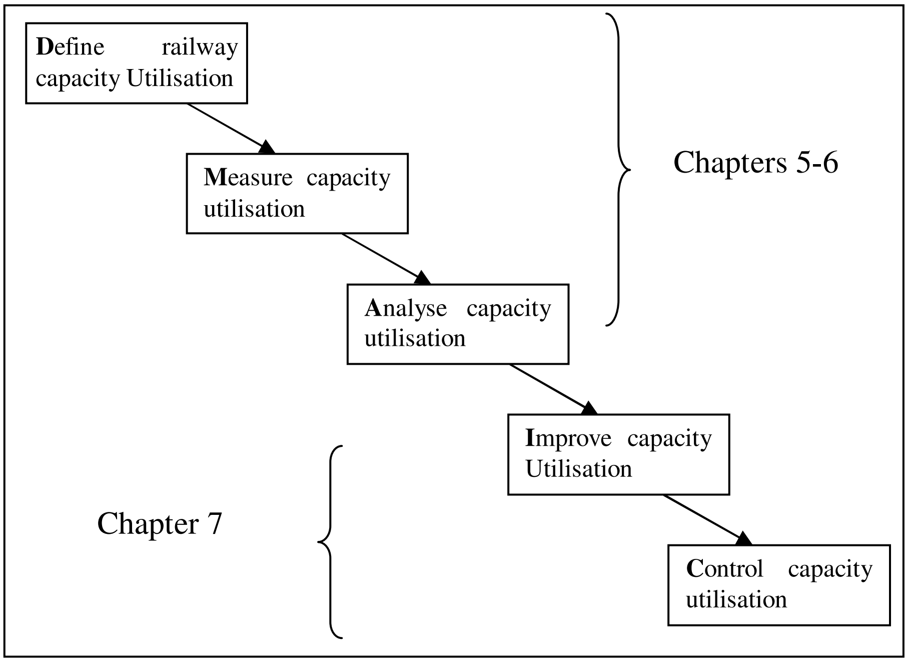


Figure 1-6 DMAIC improvement cycle for capacity utilisation (Khadem Sameni et al., 2011b)

Table 1-7 presents written papers and how the structure of the thesis is based on them.

Table 1-7- Written papers and the structure of the thesis

Chapter	Papers
Chapter 2- Past approaches to used railway capacity	(Khadem Sameni and Preston, 2012a)
Chapter 3- Factors Affecting Capacity Utilisation	(Khadem Sameni et al., 2010a), (Khadem Sameni et al., 2010b) and (Khadem Sameni et al., 2010c)
Chapter 5-Measuring and Analysing Capacity Utilisation by Passenger Operators	(Khadem Sameni and Preston, 2012a), (Khadem Sameni and Preston, 2012b) and (Khadem Sameni and Preston, 2012c)
Chapter 6-Defining, Measuring and Analysing Capacity Utilisation in the Freight Sector	(Khadem Sameni et al., 2011a)
Chapter 7 Improving and Controlling Capacity Utilisation	(Khadem Sameni et al., 2011b) and (Khadem Sameni et al., 2013)

They are written based on the research conducted during the PhD studies at the University of Southampton, one month research visit to the Technical University of Denmark- Department of Transport in 2009 and one month research visit to the University of Illinois at Urbana-Champaign- Railroad Engineering Program in 2010.

1.7 Research questions

The overall research question of the present thesis is "How to manage the railway capacity challenge?". In order to facilitate answering this question, it is broken down into the following:

- What is the most suitable way of defining railway capacity from a systems engineering point of view[6]? (Define)
- How to measure capacity utilisation from a systems point of view? (Measure and Analyse)
- How to decrease congestion delays? (Measure, Analyse and Improve)
- How best to utilise the line infrastructure with the aid of the chosen capacity measures? (Analyse, Improve and Control)
- How to choose appropriate capacity enhancement measures? (Improve and Control)
- How to ensure that capacity measures are used appropriately? (Control)

It will also try to more specifically answer the following generic questions for the key issues of railways in Great Britain:

- "How best to improve efficiency and reduce costs to taxpayers and customers?" (Office of Rail Regulation, 2011a)
- "How to get the best out of the rail network?" (Office of Rail Regulation, 2011a)

[6] From a systems engineering point of view, the description of a system and its objectives are mainly determined by its goals REIGELUTH, C. M. 1983. Meaningfulness and instruction: Relating what is being learned to what a student knows. *Instructional Science,* 12**,** 197-218..

2 Past approaches to used railway capacity

In this chapter past approaches to used railway capacity are comprehensively studied. Various definitions of railway capacity, existing methods to analyse capacity utilisation, research trends, major comprehensive works and existing software packages are presented and discussed.

2.1 Definition of railway capacity

The first step toward analysing and improving the utilisation of the railway infrastructure is to define capacity. Railway capacity is a seemingly easy but rather inaccessible concept. As expressed by Burkolter (2005), it has been "a vague expression in railway systems". Some of the major definitions of railway capacity are:

- "Capacity is the level of traffic (i.e. number of trains per day) that a rail line can accept without exceeding a specified limit of queuing time." (Peat Marwick and Partners, 1977)
- "The ability of the carrier to supply as required the necessary services within acceptable service levels and costs to meet the present and projected demand." (Kahan, 1979)
- "Capacity is the highest volume (trains per day) that can be moved over a subdivision under a specified schedule and operating plan while not exceeding a defined threshold." (Krueger, 1999)
- Capacity of the track can be identified by "The number of trains that will cause the system to lock up". "Track capacity is not constant but variable with prevailing conditions" (Kieran, 2001)
- "Line capacity is the maximum number of trains that can be operated over a section of track in a given period of time, typically one hour." (Transportation Research Board, 2003)
- "Capacity as such does not exist. Railway infrastructure capacity depends on the way it is utilised. On a given infrastructure, capacity is based on the interdependencies existing between the number of trains, the average speed, the stability and the heterogeneity." (UIC, 2004)
- "The maximum number of trains that may be operated using a defined part of the infrastructure at the same time as a theoretical limiting value is not reached in practice." (Hansen and Pachl, 2008b)
- "Capacity is measured as the count of valid train paths over a fixed time horizon within an optimal master schedule". (Harrod, 2009)
- "Capacity is the ability of infrastructure to generate value by moving passengers (or freight) toward their destination. The value generated is a function of 'macro capacity utilisation' which is the quantity of discrete steps to use railway capacity (e.g. the number of trains) and 'micro capacity utilisation' which is the quality of discrete steps to use railway capacity (e.g. load factor)". (Khadem Sameni et al., 2011b)

UIC (2004) has concluded that : "A unique, true definition of capacity is impossible." It is discussed that "Railway infrastructure capacity depends on the way it is utilised" which is a trade-off between the number of trains, heterogeneity and average speed (Figure 2-1).

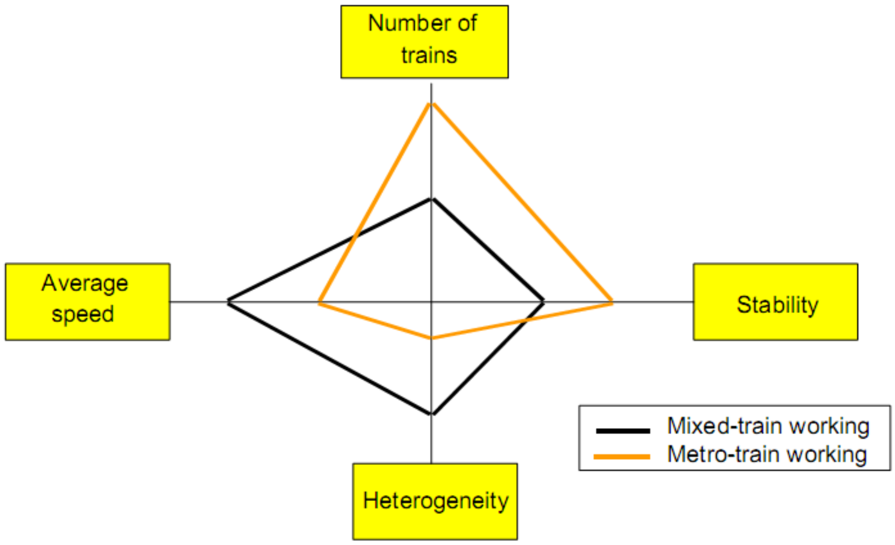


Figure 2-1 – Capacity [utilisation] balance (UIC, 2004)

UIC (2004) defines the four parameters shown in Figure 2-1 as the following: Number of trains refers to the total number of trains that use the railway infrastructure per time interval (e.g. trains per hour). Stability is considered as the impact of one minute delay of one train to other trains. Heterogeneity (discussed in more detail below in section 3.1.1) is a measure of difference between running time of various trains and is identified by the number of train types. Average speed shows the mean speed of trains that use the infrastructure. The parameters will be studied in full in chapter 3, Factors Affecting Capacity Utilisation.

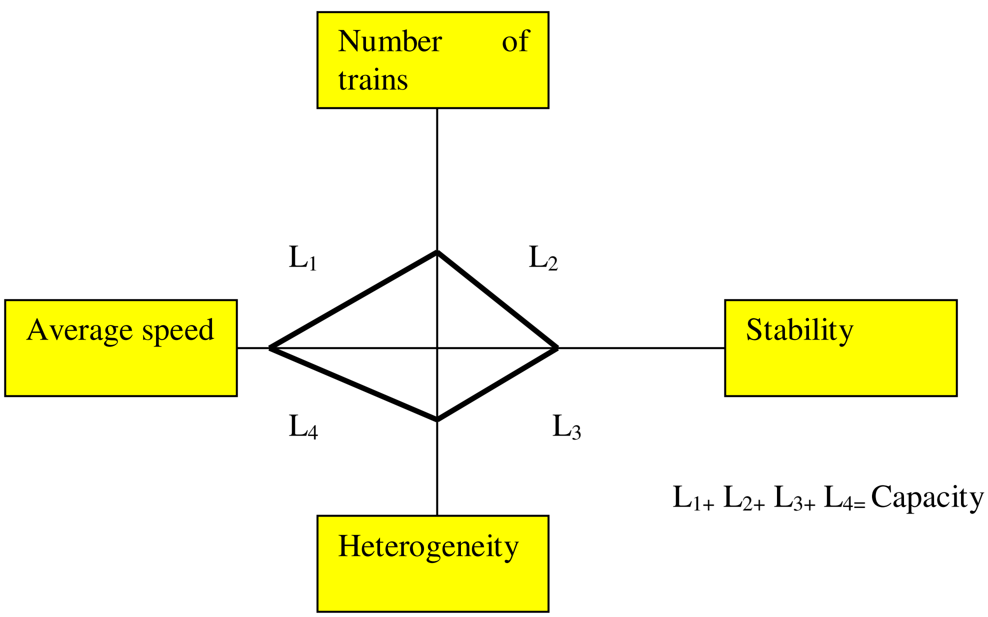


Figure 2-2 - 2D representation of capacity [utilisation]

The length of the chord that links these parameters together represents railway capacity utilisation (Figure 2-2). Alternatively, by adding an extra dimension, capacity utilisation can be regarded as the apex of the pyramid connecting these four parameters (Figure 2-3). The apex can be measured by analytic methods which would be discussed later in section 2.3.

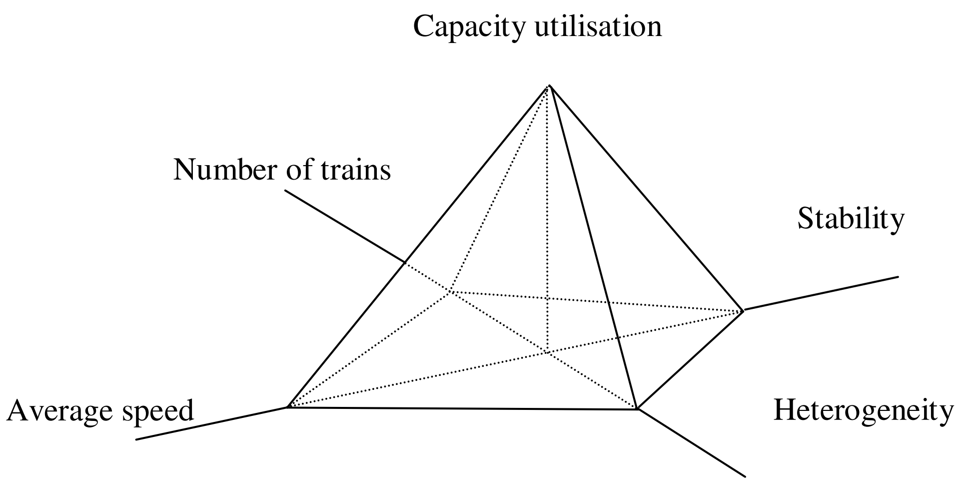


Figure 2-3 - 3D representation of capacity Utilisation (Landex, 2008)

Changes to any of these parameters will affect the others to produce a trade-off. For instance, increasing the number of trains results in reduced stability or decreased average speed. By decreasing heterogeneity, higher average speed or a more stable timetable can be achieved (Figure 2-4).

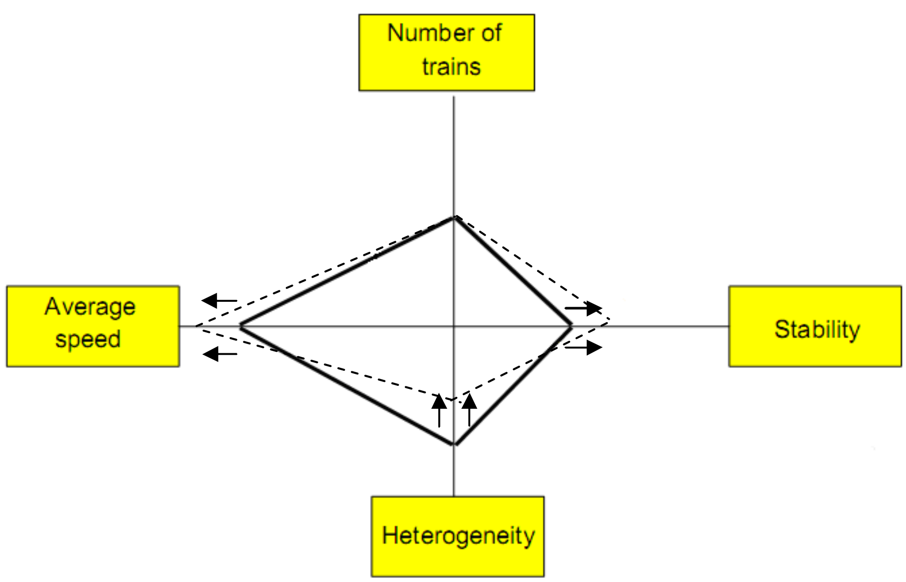


Figure 2-4 - Changing the capacity utilisation balance - based on (UIC, 2004)

As UIC (2004) points out : "Increasing capacity [utilisation] means increasing the length of the chord." If capacity is increased, all or some of the parameters can increase simultaneously as Figure 2-5 illustrates. This can be achieved by taking measures like improving infrastructure, building flyovers, sidings, choosing a better timetable, etc.

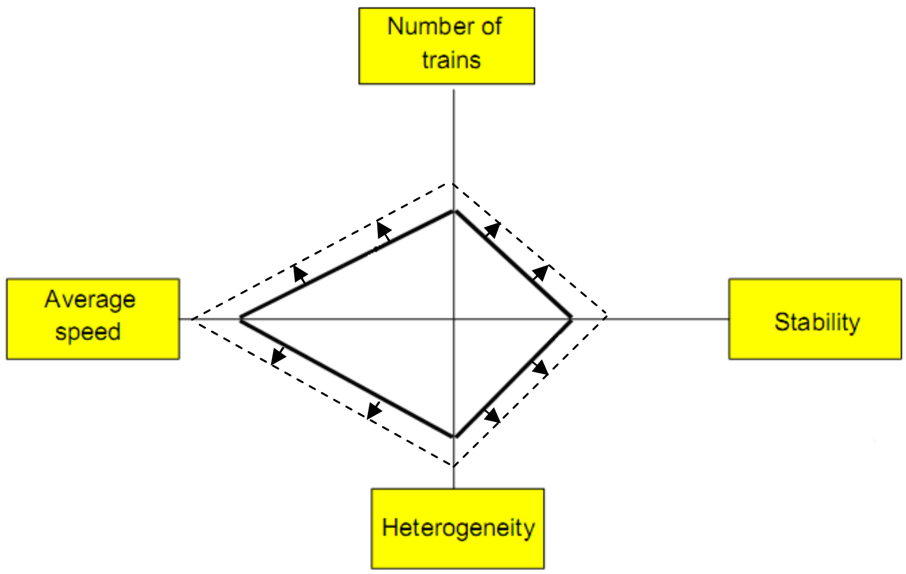


Figure 2-5 – Increasing practical capacity - based on (UIC, 2004)

Although the above-mentioned approach to railway capacity is very helpful for comprehending the concept of capacity utilisation, the length of the chord does not convey any meaning as these four parameters have different units of measurement. It is primarily useful for comparing relative capacity utilisation between railways. Figure 2-6 shows the balance of capacity for some high speed railways from a recent survey conducted by UIC (2009). It can be seen that stability of the network is very much dependant on the number of trains that operate on it. The more congested the network, the higher the effect of one minute delay on other trains would be.

The author believes that none of the above-mentioned definitions can reflect how efficiently the capacity of infrastructure is utilised. The real 'value' generated by using the capacity of infrastructure is not reflected in these definitions as they have a 'macro' approach toward used capacity and consider trains as black boxes[1]. A genuine definition for used railway capacity should consider the 'micro' aspects of it such as load factor. (Macro and micro approaches are discussed later in the thesis) It should also be noted that an appropriate definition of used capacity depends on the stakeholder. In the light of these comments, section 5.1 develops a more holistic definition of used railway capacity.

[1] Black box is a term in systems engineering indicating a system that just its inputs and outputs are considered and what actually happens in the system is ignored.

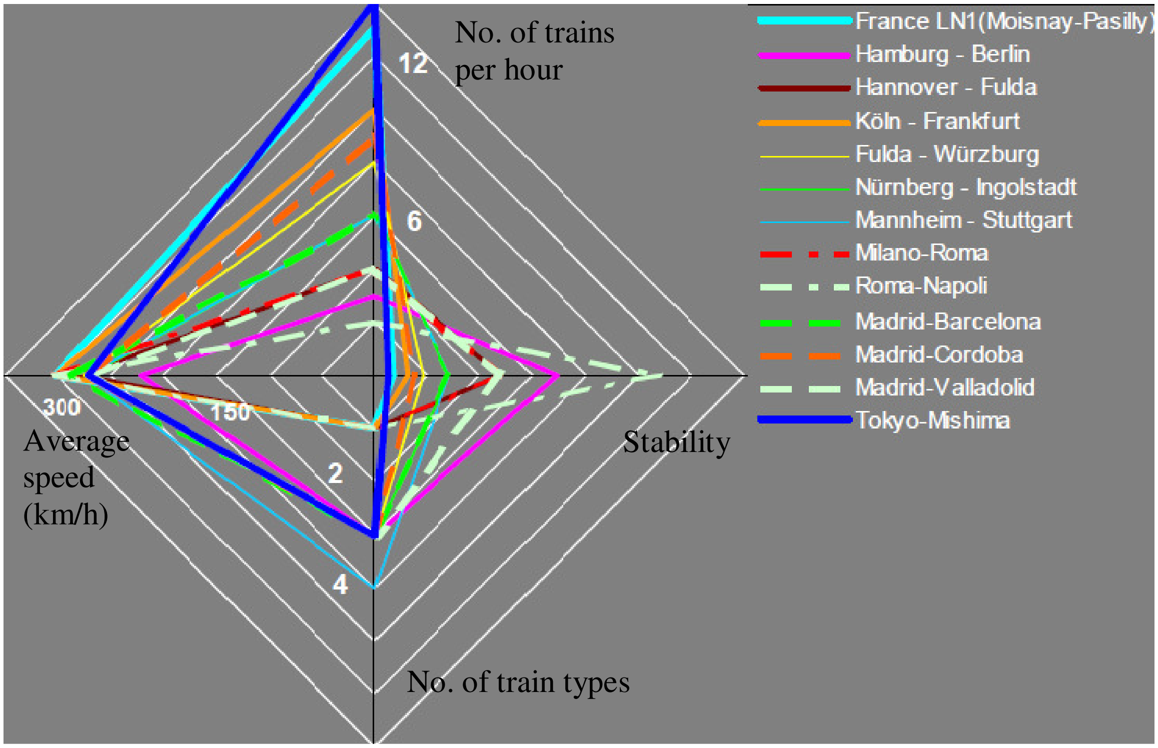


Figure 2-6 - Capacity balance for some high speed railway lines (UIC, 2009)

2.2 Types of railway capacity

Krueger (1999) categorises railway capacity into different types and stresses that practical capacity is the most important one:

- Theoretical capacity (Upper bound of capacity)
- Practical capacity (Practical limit of traffic for a defined performance level)
- Used capacity (Actual traffic volume and its variations on the line)
- Available capacity (The difference between used and practical capacity)

UIC (2004) defines four types of capacity:

- Used/consumed capacity
- Unused capacity (The difference between capacity consumption and chosen time window)
- Usable capacity (unused capacity that can be used for accommodating new train paths)
- Lost capacity (unused capacity that can not be used for accommodating new train paths)

The same concepts are rephrased by Landex (2008) in the following types:

- Maximum capacity (ideal and analytical capacity)
- Fundamental capacity (the capacity that can be used for operating trains while taking the reliability of infrastructure, rolling stock, etc. into account)
- Available capacity (daily and short-term capacity that might be less than fundamental capacity due to unfavourable weather conditions, shortage of crew, etc.)

The hierarchy of these types of capacity is depicted in Figure 2-7.

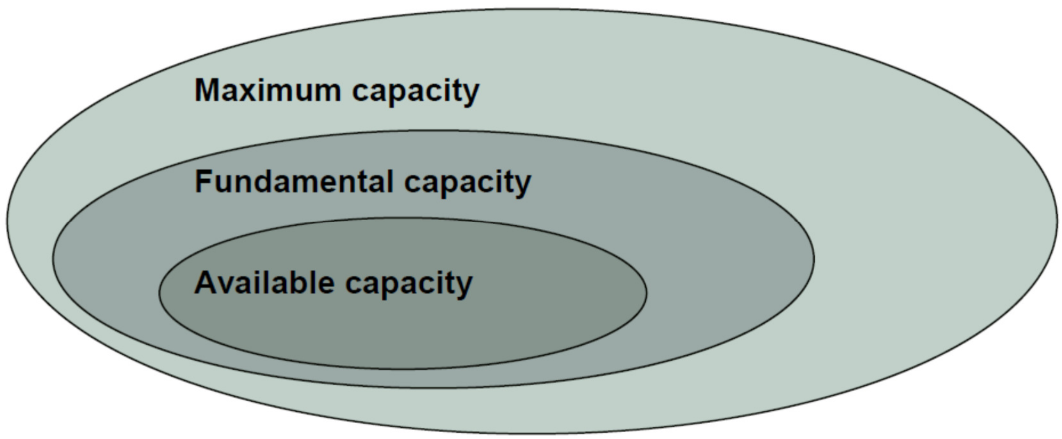


Figure 2-7 - Reduction of railway capacity (Landex, 2007)

2.3 Methods of estimating railway capacity utilisation

Capacity utilisation is defined as "the amount of capacity used for a given timetable on a given infrastructure"[1] (Landex, 2008) and there are various methods of estimating it. As described by Krueger (1999) and Abril et al. (2008), they can generally be placed into four categories:

- Analytical methods such as graphical compression methods
- Parametric models such as the works by by Krueger (1999) and Lai (2008)
- Optimisation such as the works reviewed by Lusby et al. (2009)
- Simulation such as RailSys software

In the following sections, all these methods will be reviewed in detail.

[1] In the original reference this has been defined as "capacity consumption" but to be consistent with the existing norms in the Great Britain, "capacity utilisation" is used instead.

2.3.1 Analytical methods

Some simple theoretical formulae can estimate the maximum number of trains for a line without timetable. The most basic and earliest (practical) railway capacity formula was developed by Poole (1962) for the ideal capacity where traffic is totally homogenous:

$$C = \frac{1440}{t}$$

C = [Practical] Capacity in trains per day
t = running time in minutes between two siding centres

He later on developed his formula further. The Poole equation is:

$$C = \frac{1440}{2t + \frac{t}{2} + m} \times 2$$

Where:

C = [Practical] Capacity in trains per day
1440 = Minutes in a day
t = Travel time in minutes between two sidings
$\frac{t}{2}$ = Average dwell time waiting for opposing train to arrive
m = Delay for each meet due to braking, entering the siding, running the length of the siding, leaving the siding and accelerating to full speed
2 = Number of trains per pair

It should be noted that this measure has very limited practicality as not all the minutes of a day can be used for running trains. However, it can provide some clues about theoretically maximum possible trains.

2.3.1.1 CUI method

A very simple theoretical formula for capacity analysis is the capacity utilisation index (CUI) which is defined as the time taken to operate a ‘squeezed’ or minimum technically possible timetable compared to the time taken to operate the actual timetable as in Figure 2-8 (Gibson et al., 2002). The CUI method is the measure used in the UK for capacity analysis and it is based on the minimum headways derived from Network Rail’s “ Rules of the Plan” (which have recently been renamed as “Timetable Planning Rules”. It also has less details compared to the UIC 406 method (Armstrong et al., 2009) described in the next section.

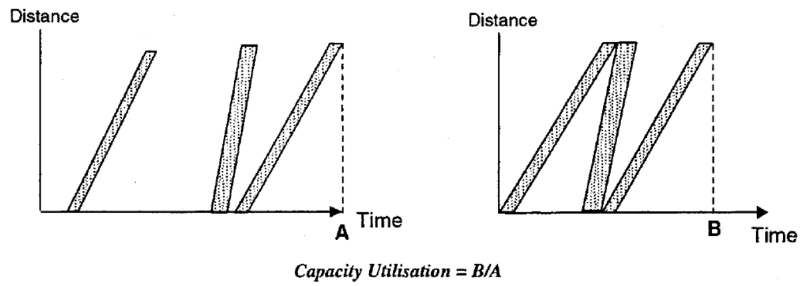


Figure 2-8 - Capacity utilisation index (Gibson et al., 2002)

The example below by Faber Maunsell (2007) illustrates how the CUI is calculated.

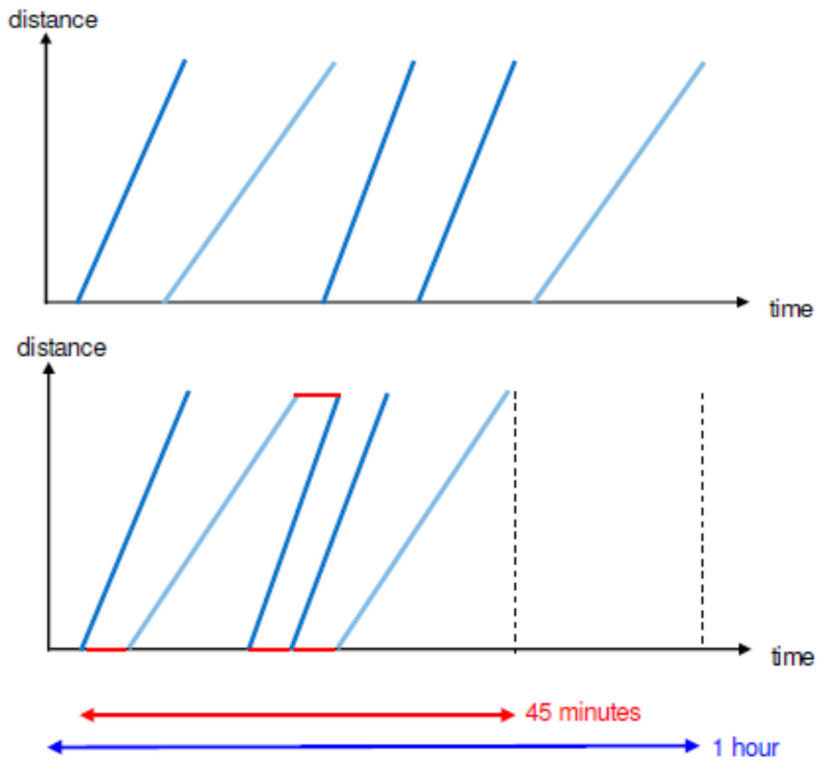


Figure 2-9 - Calculating CUI for an example (Faber Maunsell, 2007)

The capacity utilisation index in this example would be 45 min/ 60 min = 75%

The drawbacks of the CUI method are that: 1) it is a broad estimation and sensitive to the way the timetable is compressed and 2) it can not be used for nodal capacity constraints (e.g. stations) (Network Rail, 2009c).

2.3.1.2 UIC 406 method

The most famous theoretical formula for capacity analysis is the UIC 406 capacity method developed by the UIC (2004) which has been adopted in many European railways. The UIC 406 method provides a straightforward method of timetable analysis and used capacity by compressing the timetable so that the buffer time is zero (UIC, 2004). As recommended by this standard, "ideally the line section used for compression should be reduced to the line section between two neighbouring stations (without overtaking or crossing possibilities)". Firstly the timetable is produced. Then the railway network is divided into sections at:

- Junctions
- Change of train order
- Change in number of trains
- Change in number of tracks
- Overtakings and crossings stations

However, the results of the study by Landex (2008) shows it is better not to divide the network into sections at overtaking and crossing stations as it might result in very low used capacity by segmenting lines too much. For the next phase, the timetable is compressed. All train paths are "pushed together to the minimum headway" without any changes in the running times, running time supplements, dwell time at stations and block occupation time. (UIC, 2004) (Landex, 2008)

The general workflow of the UIC method is illustrated in Figure 2-10.

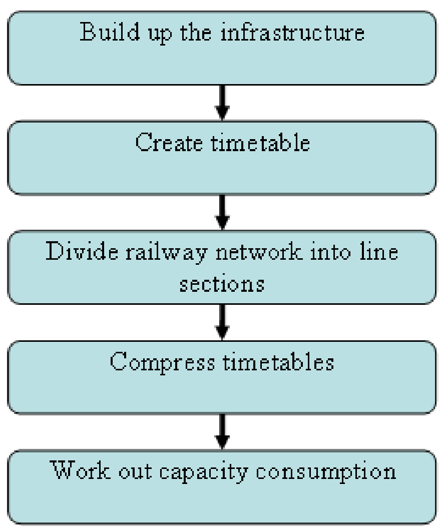


Figure 2-10 -General workflow of the UIC 406 method Source: (Landex et al., 2006)

The UIC formula for determining used capacity is:

k=A+B+C+D

k: Total used time (min)
A: Infrastructure occupation (min)

B: Buffer time (min)
C: Supplement for single-tracks (or crossing time) (min)
D: Supplements for maintenance (min)

K= k$\times$100 / U

K : Used capacity (%)
U: Chosen time window (min)
(UIC, 2004)

Buffer time is the time added to decrease the risks of delay propagation. The supplement that is added for single track operation does the same but can be added at crossing stations (UIC, 2004, Landex, 2008). (Kaas, 1998) suggest the following formulae to calculate buffer time according to capacity utilisation ratio and headway:

$$K_{\max} = \frac{\Delta T}{t_{h\min}}$$

$$K_f = u \times k_{\max}$$

$$K_f = u \times \frac{\Delta T}{t_h}$$

K_f : Usable capacity (theoretical capacity) (number of trains)
$K_{\max}$: Maximum capacity (practical capacity) (number of trains)
ΔT : Observation period (min)
u : Percentage of utilisation of the maximum capacity (used capacity)
$t_{h\min}$: Minimum headway (min)
t_b: Buffer time (min)

$$K_f = \frac{\Delta T}{t_{h\min} + t_b}$$

By inserting the second formula into the first one, the buffer time (min/train) would be:

$$K_f = u \times \frac{\Delta T}{t_h} = \frac{\Delta T}{t_{h\min} + t_b} \Rightarrow t_b = \frac{\Delta T}{K_f} - t_{h\min}$$

The simple example below illustrate the use of these formulae further. If the time window which is considered is 60 minutes (ΔT =10) and the minimum headway for the line is 5 minutes then:

$$K_{\max} = \frac{60}{6} = 10$$

U is extracted according to UIC suggestions as presented in Table 2-1. For the daily period of dedicated high speed lines the guideline suggests 60% therefore:

$$K_f = 0.6 \times 10 = 6$$

$$t_b = \frac{60}{6} - 6 = 4$$

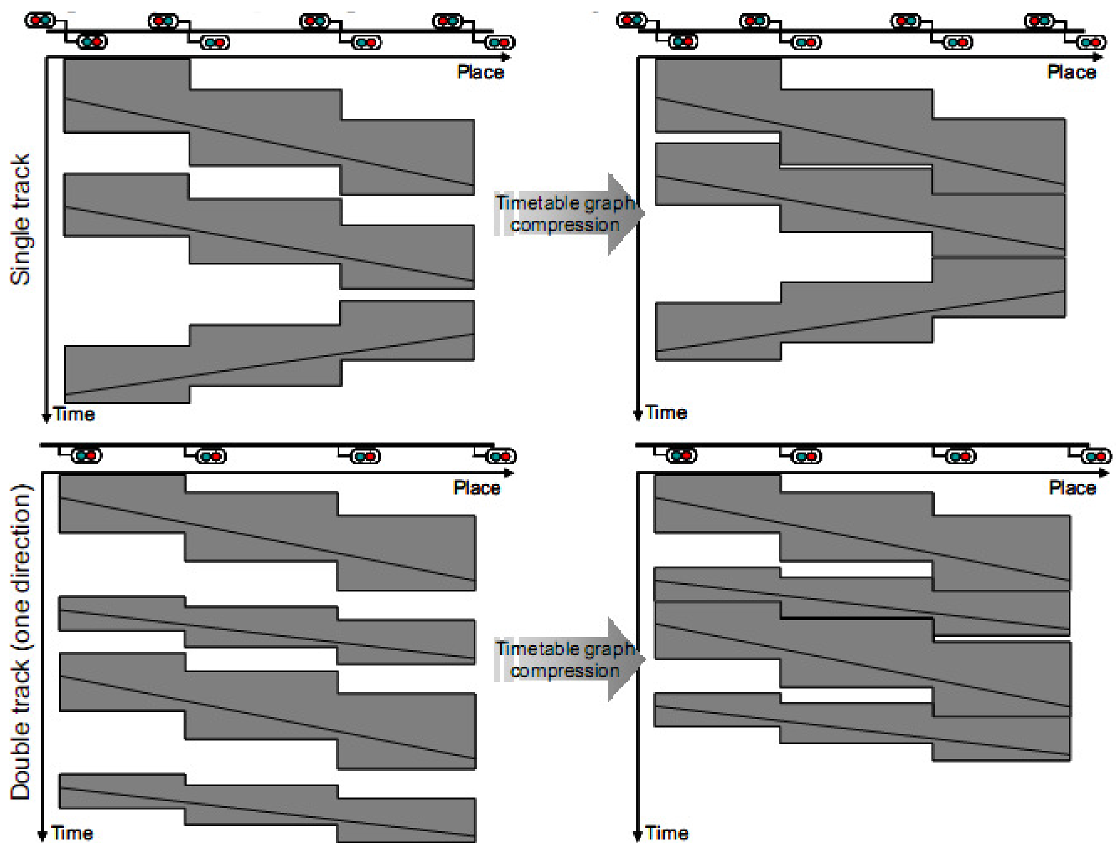


Figure 2-11- Timetable Compression according to UIC 406 method (Landex et al., 2008)

Based on the European experience, the UIC suggests some empirical limits for capacity utilisation which are presented in Table 2-1.

Table 2-1 - Guidelines for capacity utilisation (UIC, 2004)

Type of line	Peak hour	Daily period
Dedicated suburban passenger traffic	85%	70%
Dedicated high speed lines	75%	60%
Mixed traffic lines	75%	60%

The UIC 406 method has successfully been applied in several European railways. Höllmüller and Klahn (2005), Wahlborg (2005) and Landex (2008) apply the UIC 406 method to Austrian (ÖBB), Swedish (Bahnverket) and Danish railway (Banedanmark) networks respectively.

Confessore et al. (2009) combine a discrete event simulation approach with the UIC 406 compression method to calculate the practical capacity of a line in Italy (measured in number of trains). The general workflow is shown in Figure 2-12. In the simulation phase they cover factors that are not accommodated in the optimisation phase, mainly stochastic traffic perturbation.

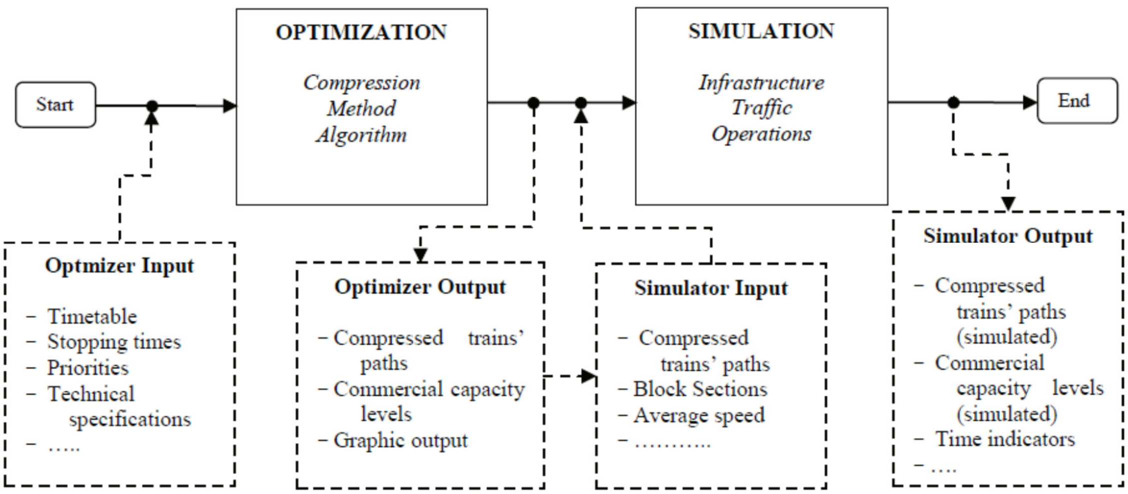


Figure 2-12 - Simulation-based approach to capacity assessment (Confessore et al., 2009)

However, Landex (2008) identifies two paradoxes in the practical use of the UIC 406 method:

- Paradox of overtaking

In the case of overtaking, as the order of trains change, line sections must be divided into smaller ones (before and after overtaking). The results may show reduced capacity utilisation. For resolving this, he suggests dividing the line only if there are many overtakings. In the case of few overtakings, he recommends maintaining the order of trains but changing dwell time into the minimum dwell time required.

- Paradox of extra train

Calculating capacity utilisation after adding an extra train may show that capacity utilisation has decreased. For single track lines, he suggests a "dummy train" method to better decide where to divide the railway line.

2.3.1.3 Overview of analytical methods

Analytical methods use simple mathematical formulae or timetable compression methods to quantify railway capacity utilisation. These methods are quick and straightforward to give a good overview of a line or network but cannot encompass the complex nature of railway capacity. As Farrell (1957) has put it "The more complex the process, the less accurate is the theoretical function"

The UIC 406 method enables the evaluation of capacity utilisation for train path management but not infrastructure planning. However, the clues it provides for railway planners are limited, for the following reasons:

- Stability and reliability aspects of the timetable are ignored whereas they greatly affect capacity utilisation. Stations may also have non-scheduled operation (e.g. train formation).
- All the trains are considered the same although the traffic they carry has different values and priorities.
- The capacity of complex stations cannot be assessed due to lack of knowledge about exact train routing and platform operations (Landex, 2008).
- Timetables that have different train combinations but nearly the same capacity utilisation cannot be compared with each other as used capacity is calculated by a non-weighted summation of all trains.
- Network effects are not examined as only short sections of the network are included in the analyses.
- Scheduled waiting time is not considered as it is in the basis of the analysis. (Khadem Sameni et al., 2010a)

2.3.1.4 Research trends

Recent research trends in the field of analytical methods are:

- Extending the applicability
  - to nodes, as in the studies undertaken by Landex (2011) and Lindner (2011);
  - to the whole network (Armstrong et al., 2011a)
  - in new contexts for freight dominated railways (Lindner and Pachl, 2010)
  - to the enrichment process for adding extra trains (Lindner, 2011)
- Automating analytical methods
  - the UIC 406 method in new versions of RailSys (RMCon, 2009)
  - the UIC 406 method for huge networks (Kuckelberg et al., 2011)
  - the CUI method by Armstrong et al.(2009)
- Developing the UIC 406 methodology
  - suggesting meso indices to add or remove trains by Khadem Sameni et al. (2011b)
  - occupation time estimation by Gasparik and Zitricky (2011)

2.3.2 Parametric models

Parametric models use some parameters of railway infrastructure and operation to describe and analyse capacity utilisation. Prokopy and Rubin (1975) developed the first parametric model that calculates used capacity by means of train delay and a function of physical, operations and control parameters (Lai and Barkan, 2009b). Another parametric model was developed by Krueger (1999) for the Canadian National Railway. He used the following parameters in his model:

- Plant parameters
  - length of subdivision (block length)
  - meet-pass point spacing and uniformity
  - signal spacing (signal type)
  - percentage of double track
- Traffic parameters
  - Traffic peaking
  - Priority of trains
  - Speed ratio
  - Running times
- Operating parameters
  - Track maintenance

Lai (2008) and Lai and Barkan (2009a) developed an enhanced parametric model based on Krueger's work (1999). The model is part of a decision support system called RCET (Railway Capacity Evaluation Tool) which can optimise investment in different capacity expansion schemes. As shown in Figure 2-13, it consists of 3 modules:

- Alternative generator (all possible expansion)
- Investment selection model (selecting appropriate parts of the network)
- Impact analysis model (trade-off between investment and costs of delay)

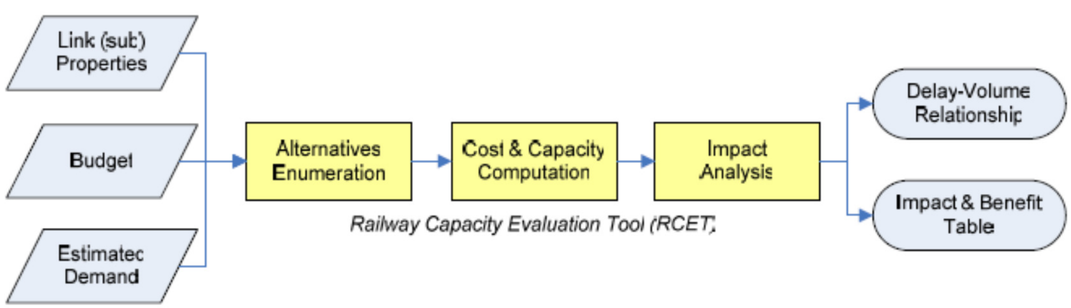


Figure 2-13 - Railway Capacity Evaluation Tool as a decision support system

Available studies in the category of parametric models are limited. It seems more appropriate for the railways that the operation manager and the infrastructure owner are the same entity to be able to include these different parameters in one model. Therefore, all the parametric model studies have been carried out in North America where the railway company operating trains own the infrastructure as well.

2.3.3 Optimisation

Due to the very complex and multidisciplinary nature of railway capacity, mathematical programming and operations research are not directly used for modelling and optimising capacity utilisation (i.e. maximising efficient capacity utilisation, subject to demand/infrastructure/ signalling/ operational/ rolling stock/ fare/ access charge constraints). Therefore, optimisation techniques are extensively used only for sub-problems of capacity utilisation especially train scheduling, rescheduling and routing as well as track and platform allocation. Assad (1980) surveyed different mathematical models for optimising railway operations. Cordeau et al. (1998) review train scheduling and rescheduling in their comprehensive survey. Tornquist (2005) provides an overview of research in the field of railway scheduling and dispatching. Hansen et al. (2008) present chapters about state-of-the-art techniques on timetable design principles, infrastructure modelling, timetable stability analysis, optimisation models for railway timetabling, simulation, rescheduling and performance evaluation. Lusby et al. (2009) provide a recent survey on track allocation models and methods. A summary of major optimisation works for train timetabling problem is presented in Appendix 2.

Improving train timetables, scheduling and rescheduling have always been in the spotlight of railway operations research for more than 40 years. For instance, some of the most popular current trends in the field of operations research are:

- Managing delays and developing fast algorithms for rescheduling (numerous studies such as the works by Schobel (2001), Mattson (2004) Yuan and Hansen (2007))
- Analysing robustness of timetables and increasing reliability (numerous studies such as the work by Vromans et al.(2006), Goverde (2007) and Törnquist (2007))
- Using Petri Net for nodal capacity constraints as in the studies by Milinkovic et al. (2011), Jia et al. (2009) and the PhD thesis by Burkolter (2005)

2.3.4 Simulation

The use of simulation for analysing railway capacity utilisation is twofold: it can be used as a tool along with other approaches like improving timetables through simulating train scheduling and rescheduling, etc. or it can be in the form of a software package that has some direct or indirect features for used capacity analysis. Simulation methods try to estimate total waiting time of all trains through simulation of the timetable. Pachl (2009) categorised the use of simulation for used capacity analysis into synchronous and asynchronous simulation. Asynchronous simulation separately simulates the running operations from scheduling, hence stochastic delays are artificially generated and solved according to dispatching rules. Synchronous simulation, in which all railway operation are simulated in real time, is more sophisticated but yields more realistic results.

Recent research trends in the field of simulation are:

- Combining simulation with optimisation for used capacity analysis like the studies by Cofessore et al. (2009), Armstrong et al. (2011b) and Schlechte (2011)
- Using simulation for assessing the interactions between capacity utilisation and other railway operations by Lindfeldt (2010), Gille and Siefer (2011) and Dicembre and Ricci (2011)

A comprehensive survey of railway simulation packages is presented by Barber et al. (2007) and Kontaxi and Ricci (2011). Table 2-2 provides a summary of key software railway software packages, their general and capacity-related features. By this comprehensive study it is concluded that the RailSys software is the most widely used one, providing the necessary features for used capacity analysis in the European context. This software would be used in the case study of section 7.2.

2.3.5 Queuing Models

Queueing models are based on operations research and they analyse systems where a service is being offered to customers through one channel or several channels and where variability in the arrival of the service and the customers can lead to queues forming. Such models analyse and improve various aspects of system performance such as average waiting time by considering the inter-arrival distribution of customers (for example by using the Markovian or Erlang distributions), service time distribution and number of channels. Major applications of queueing models in railways calculate scheduled waiting times of trains competing for the available infrastructure and estimate the knock-on delays. This method is widely used in Germany for capacity studies (Wendler, 2008).

As detailed by Yuan et al.(2006), the average scheduled waiting time of trains on open tracks for heterogeneous traffic was studied by the queueing models developed by Schwanhäußer (1974). This application was further developed by Wakob (1985) and Wendler (1999). A queueing model consists of arrival process, service process, service station and waiting area. The arrival process describes the period of time between streams of demands and the service times of a line section is a matrix of minimum headways between trains i and j. Further explanations of using waiting time are given in section 3.1.2.1.

Table 2-2 - Key railway software packages with capacity features

Name	**Producing Company / Country**	**General Features**	**Capacity Related Features**
RailSys	Rmcon /Germany	• Timetable construction and modelling • Running time calculation • Planning of capacities • Infrastructure planning • Scheduling possessions and planning of special traffic • Planning of logistic concepts for large scale projects • Design, investigation and registration of timetables • Validation of nationwide basic interval timetables • Investigation of operational quality, punctuality and guaranteed connections • Completion of disposition strategies in cases of delays and operational disturbances • Cost-benefit analysis • Elaboration of technical documents for transport related tenders	• Planning of capacities (UIC 406 capacity method) • Timetable optimisation by evaluating different timetable alternatives through its Evaluation Manager module

Name	Producing Company / Country	General Features	Capacity Related Features
Open Track	The Institute for Transport Planning and Systems of the ETH Zürich / Switzerland	• Infrastructure planning and comparing different options • Timetable construction, analysis and simulation • Rolling stock analysis and planning • Signalling analysis (including different levels of ERTMS) • Power and energy analysis	• Determining capacity of stations and lines • Analysing the effects of infrastructure or train failures and delays caused
DONS (and its SIMONE module for capacity)	Railned/ Netherland	• Generating cyclic timetables • Routing trains trough railway stations	• trace and quantification of bottlenecks in a network (SIMONE module)
PETER	Delft University/ Netherland	• Calculating timetable performance indicators	• Identifying bottlenecks with tightest schedule

Name	Producing Company / Country	General Features	Capacity Related Features
VIRIATO (and its CAPRES module for capacity)	EPFL + SMA and partner/ Switzerland	• Regular interval timetable planning • Producing netgraph (schematic representation of a railway network and timetable)	• Identifying bottlenecks • Evaluating the remaining capacity of a network • Comparing different timetables according to capacity • Determining additional trains that can be added • Estimation of the effects on capacity caused by modification of infrastructure • Impacts of new lines added to a network
DEMIURGE	SNCF/ France	-	• Evaluating a network's capacity to absorb additional traffic • Locating bottlenecks to assist in making decisions about infrastructure investments • Optimising current and future timetables • Calculating the residual capacity of a timetable

Name	Producing Company / Country	General Features	Capacity Related Features
RAILCAP	Stratec / Belgium	-	• Calculating the capacity used by a scenario • Analysis of bottlenecks • Calculating the operations program's influence on the available capacity
CMS (Capacity management system)	DeltaRail/ UK	• Timetable planning and validation • Conflict detection • Loads modeling • Resource planning • Visual representation of infrastructure	• Choosing the best timetable among different options based on capacity, resources and demand evaluations
RTC (Rail Traffic Controller)	Berkeley Simulation Software/ USA	• Through train dispatch and conflict resolution at the network level • Integrated train performance calculator • Operating plans	• Diagnosing bottlenecks and recommending schedule changes • Assessing the impact of adding new trains to a network • Evaluating various capital improvement scenarios

2.4 Major comprehensive works on railway capacity

Table 2-3 reviews major comprehensive studies that have a holistic approach to railway capacity and are usually the main (inter)national points of reference. The main themes are measuring, analysing and improving capacity utilisation. This study provides a comprehensive railway track capacity manual which covers all aspects of defining, measuring, analysing, improving and controlling.

2.5 Summary and conclusions

Railway capacity is a seemingly easy but rather inaccessible concept. Various definitions exist in the literature and it is concluded by the International Union of Railways (UIC) that: "A unique, true definition of capacity is impossible." and that "Railway infrastructure capacity depends on the way it is utilised". The four main factors affecting capacity utilisation are the number of trains running on the infrastructure in the unit of time, heterogeneity, reliability and the average speed. It is the balance between these factors that determines railway capacity.

Defining railway capacity is a stepping stone towards analysing and improving it. Existing definitions of capacity tend to focus on the number of trains, hence the concept of 'macro capacity utilisation'. It should be noted that the real 'value' generated by using the infrastructure can not be reflected by considering trains as black boxes. A genuine definition for railway capacity should consider its 'micro' aspects such as load factor.

Past approaches to analysing railway capacity are categorised into: analytical methods, parametric models, operations research and simulation. Analytical methods such as the UIC 406 (used in continental Europe) and the CUI (used in Great Britain) give a general overview of how much the infrastructure used by compressing the timetable to the minimum technically possible and generating a capacity utilisation index. Parametric models analyse the capacity utilisation curve by the relationship between the parameters of infrastructure, timetable, operation, etc. Operations research mainly optimises sub-problems of capacity utilisation (timetabling, train routing, etc.). Simulation can be used alongside other approaches or can be used by software packages to estimate delays in a synchronous or asynchronous manner.

A gap is felt in the approaches toward analysing capacity utilisation for tactical and strategic planning. At one end of the spectrum, simulation and operations research tend to focus on meticulous details of operational planning and they are computationally intensive. At the other end, analytical methods can be helpful for tactical or strategic planning but they are overly simplified and can provide very limited insights. Hence, the thesis will try to develop methodologies to analyse capacity utilisation for tactical and strategic planning purposes based on the concept of the 'value' generated by capacity utilisation.

Table 2-3 - Major comprehensive works on railway capacity

Author(s)	Theme	Main Contributions	Type	Volume (pages)	Country of case studies
Kieran (2001)	Pricing railway capacity	• Comprehensive study of track access charges in Europe and North America • Suggesting a track access pricing process for Canada	Research project	38	Canada
Cambridge Systematics (2007)	Improving capacity utilisation	• Identifying level of service for primary corridors in the US railway network • Estimating future capacity improvements needed	Research project	69	United States
Harrod (2007)	Improving capacity utilisation	• A new practical model for master scheduling of a freight railway by considering line capacity constraints, multi commodity flows and network value	PhD thesis	215	United States
Abril et al. (2008)	Improving capacity utilisation	• Survey of capacity analysis methods • Developing a system called MOM that can produce improved timetables for off-line and on-line scenarios, analyse network capacity utilisation and timetable robustness.	Journal Paper	33	Spain
Lai (2008)	Improving capacity utilisation	• Developing a decision support system named RCET that can optimise investing in different capacity expansion schemes	PhD thesis	184	United States

Author(s)	Theme	Main Contributions	Type	Volume (pages)	Country of case studies
Landex (2008)	Measuring and analysing capacity utilisation	• Thorough investigation of the UIC 406 method • Studying trade-offs in the capacity balance	PhD thesis	218	Denmark
Lindfeldt (2010)	Analysing and improving capacity utilisation	• Developing the SAMFOST mathematical model that can calculate crossing time for single tracks based on infrastructure configuration, rolling stock, timetable and delays. It can be used to assess alternative infrastructure improvements and their effects on capacity utilisation . • Developing the TVEM model that can systematically generate and compare different timetable variants for double track lines to evaluate their effects on capacity utilisation	PhD thesis	228	Sweden
Roberts et al.(2010)	Improving capacity utilisation	• Matrix of capacity interdependencies • New model for choosing capacity enhancement measures	Research project	84	United Kingdom
Pudney et al. (2010)	Measuring, analysing and improving capacity utilisation	• Survey of different capacity interrelated indicators, capacity analysis methods and capacity improvement techniques	Research Project	45	Australia
Kontaxi and Ricci (2011)	Measuring and analysing capacity utilisation	• Comprehensive overview of capacity measuring methodologies since 1950s • Developing RailCAT, an integrated online capacity calculating tool	PhD thesis	Underway	Italy

3 Factors Affecting Capacity Utilisation

Railway capacity is a multidisciplinary area. As illustrated in Figure 3-1, various factors affect capacity utilisation from rolling stock to infrastructure, timetable, human factors and even external factors such as weather conditions. Detailed study of these factors is necessary to provide a suitable foundation for defining, analysing and improving capacity utilisation to better manage ‘the capacity challenge’.

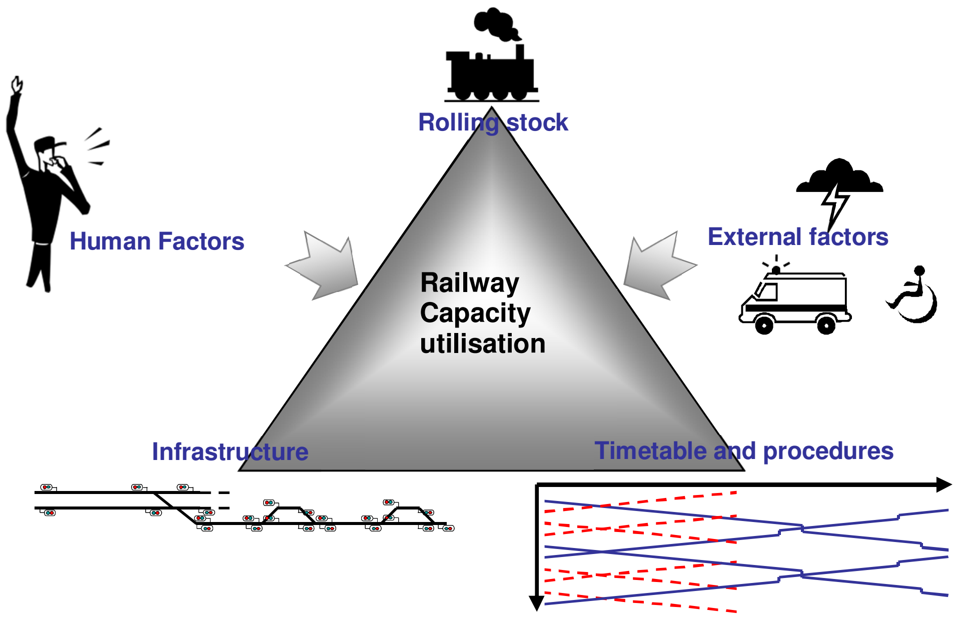


Figure 3-1 - Schematic representation of capacity [utilisation] and what affects it (Landex, 2008)

Some major factors affecting capacity utilisation are summarised and their impact on capacity utilisation is shown by ‘+’ as positive impact or ‘-’ as negative impact in Figure 3-2.The rest of this chapter investigates factors affecting capacity utilisation by broadly categorising them into timetable, signalling, nodal capacity constraints, rolling stock, infrastructure, external factors and governance.

3.1 Timetable

Railway infrastructure is a limited and expensive resource that is allocated to trains; it should be utilised in the best possible way. Wherever there is a limited resource, there is need for scheduling which is defined as the allocation of scarce resources to different tasks. In general, scheduling identifies “Which resources should be allocated to perform each task” and “when should each task be performed?” (Baker and Trietsch, 2009).

The scheduled train path is the product sold by the infrastructure manager to a train operating company and the right to run a train on that path under specified operating conditions. Scheduling is coordinating different train paths ordered by competing train operating companies.

Figure 3-2 - Major factors affecting capacity utilisation (Khadem Sameni et al., 2010)

Train timetables perform the following functions:

- Coordinating trains for optimum and efficient use of the infrastructure
- Ensuring predictability of trains
- Producing timetable data for passengers
- Providing the necessary inputs for train control, rolling stock allocation and crew scheduling

(Pachl, 2008)

The quality of the timetable has a great impact on capacity utilisation. Hansen and Pachl (2008a) identify the characteristics of a high-quality timetable as:

- Representing track and other infrastructure well
- Incorporating signalling constraints
- Considering capacity utilisation
- Estimating train running time precisely
- Applying energy-efficient driving standards
- Analysing stability and robustness
- Using techniques such as analytical or simulation
- Monitoring, analysing and evaluating the timetable regularly (in terms of punctuality and reliability).

There are many complexities involved with train planning and timetabling. Watson (2008) identifies the following:

- Trains having a single degree of freedom for movements on track;
- Trade-off between infrastructure efficiency utilisation against robustness and time taken to produce the solution;
- Congested nature of many rail routes means that it cannot accommodate all business requirements;
- Separation of infrastructure management and train operation; and
- Relatively limited software support available to train planners.

In this regard, extensive research has been carried out on different aspects of train scheduling, rescheduling and the robustness of timetable which is beyond the scope of this thesis. It has been reviewed by Cordeau et al. (1998), Tornquist (2005), Hansen et al. (2008) and Lusby et al. (2009). The characteristics of timetable that affect capacity utilisation are of interest to this thesis and are discussed in the following sections.

3.1.1 Heterogeneity

Trains of different type, speed, characteristics and stopping pattern often share the same infrastructure. If there is a significant difference between the running time and stopping pattern of trains, it is referred to as heterogeneity of traffic in the railway literature as opposed to homogeneous traffic. The capacity of the line is best utilised when all trains run under harmonised schedules without speed differences (Pachl, 2009). Heterogeneous traffic adversely affects capacity utilisation due to irregularities caused in the flow of trains and more complex timetable planning. (Dingler et al., 2009b) investigate the effect of heterogeneity on capacity utilisation. They study delays for different levels of traffic (number of trains) and heterogeneity (different combinations of train types including passenger, intermodal, coal and manifest). The results confirm that delays caused by heterogeneity are disproportionately more than a comparable increase in the amount of homogeneous traffic. For example adding more coal trains generates much more delay when there is a considerable percentage of manifest[1] and intermodal trains. They conclude that the speed ratio of two trains, rather than absolute differences in speed, contributes to heterogeneity effects

Heterogeneity has severe effects on network capacity utilisation: “The greatest constraint on our (Chiltern) line capacity is thus the impact of differential train speeds” (Dare, 2009). Lindfeldt (2009) proposes a simple deterministic method of evaluating capacity utilisation which clearly shows the effect of heterogeneity on long double track lines in Sweden. To decrease the adverse effects of heterogeneity on railway services, Vromans et al. (2004) suggest “homogenisation” through adding an extra running time supplement to long distance trains; decreasing the running time supplement for short distance services; considering running time differences only between two consecutive stations when overtaking is planned and equalising the number of stops per train (if possible and in order to decrease heterogeneity). Some of these measures, such as slowing down long distance trains or adding more stops, may seem against effective capacity utilisation. If heterogeneity of traffic causes several overtakings and considerable delays to other trains, these measures have a positive overall effect. However, they should be adopted only where necessary and with careful studies of the consequences.

A simple measure of heterogeneity can be the speed ratio of the fastest train to the slowest train as suggested by Krueger (1999):

$$Speed\ ratio = \frac{speed\ of\ fastest\ train}{speed\ of\ slowest\ train}$$

However this measure does not consider how many trains deviate from the average speed. For example if at the time period considered there are 9 fast trains and just 1 slow train, it yields the same speed ratio as in the case of 5 fast trains and 5 slow trains. Obviously traffic is more heterogeneous and the standard deviation from the average speed is higher in the latter.

With the same number of heterogeneous trains, the way trains are sequenced on the timetable also affects capacity utilisation. This is part of the study by Abril et al. (2008) and it is concluded that even spacing provides minimum capacity utilisation (i.e less capacity of infrastructure is occupied by the same number of trains). Vromans et al (2004) suggest some heterogeneity measures between two railway nodes. The sum of shortest headway reciprocals (SSHR) is defined as:

$$SSHR = \sum_{i=1}^{n} \frac{1}{h_i^-}$$

[1] A freight train carrying goods not hauled in single commodity trains or intermodal ones. TRAINS: THE MAGAZINE OF RAILROADING. 2012. *Railroading Glossary* [Online]. [Accessed 25/06/2012.

Where h_i^- is the smallest scheduled headway between two consecutive trains (i and i+1) on the track section. As stated by the authors, this measure can reflect distribution of the trains on an hour as well as their heterogeneity. The drawback is that arrival and departure headways are treated the same whereas headways at arrivals are more important than headways at departure (at arrival headways are usually larger and fast trains catch up with slow trains at the end of the blocks). Hence they suggest another measure based on arrival headways as the sum of arrival headway reciprocals (SAHR):

$$SAHR = \sum_{i=1}^{n} \frac{1}{h_i^A}$$

Heterogeneity of traffic is caused by variations in speed and stop patterns as well as variations in headways. Comparing the ratio of headways at departure and arrival can also accommodate both sources of heterogeneity. In this regard, Landex (2008) proposes to use the ratio of the headway at departure station ($h_{t,i}^D$) to the following headway ($h_{t,i+1}^D$) multiplied by the ratio of headways for arrival at stations ($h_{t,i}^A$ and $h_{t,i+1}^A$). To provide a formula independent of the number of trains, the result is divided by the number of headways minus 1 (h_{N-1}).

$$Heterogeneity = 1 - \frac{\sum \left(\min\left(\frac{h_{t,i}^D}{h_{t,i+1}^D}; \frac{h_{t,i+1}^D}{h_{t,i}^D} \right) \times \min\left(\frac{h_{t,i}^A}{h_{t,i+1}^A}; \frac{h_{t,i+1}^A}{h_{t,i}^A} \right) \right)}{h_{N-1}}$$

With this heterogeneity measure that varies from 0 to 1, the interactions between heterogeneity and capacity utilisation can be depicted as in Figure 3-3.

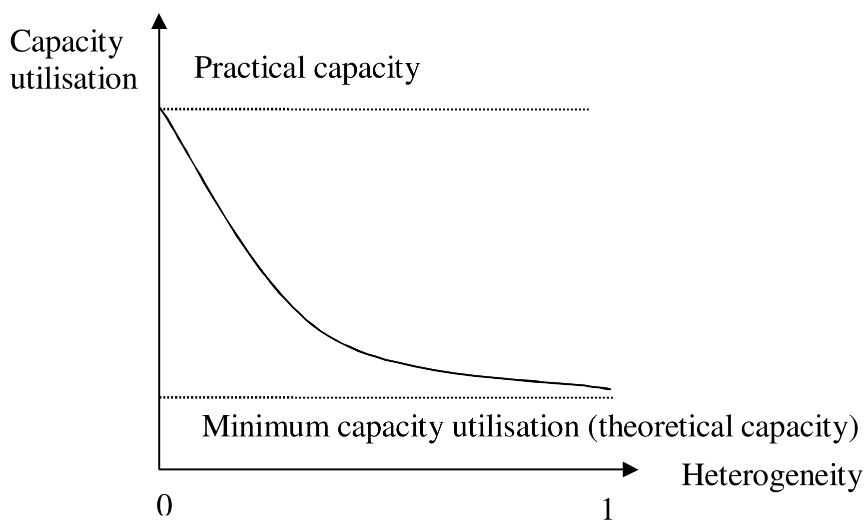


Figure 3-3 - Effect of heterogeneity on capacity utilisation

3.1.2 Stability and reliability

Punctuality is defined as "the percentage of trains which arrive to/depart from/ pass a location with a delay less than a certain time in minutes" (Hansen and Pachl, 2008b). This threshold for considering trains as delayed varies between countries. In most European countries trains that arrive less than five minutes late are not considered delayed. This threshold is four minutes for Switzerland, three minutes for the Netherlands and 10-15 seconds for Japan (Yuan, 2008). Train delays can be initial (primary/original) delays or knock-on (consecutive /secondary) delays which are caused by other trains due to the network/domino effect. Bush (2007) and Daamen et al. (2009) investigate primary and knock-on delays further. The former studies on-time performance parameter ranges and the latter develops a tool for identifying route conflicts in the event of delays and estimation of knock-on delays.

Stability of the timetable is "its ability to compensate for delays and returning to the desired state" (Hansen and Pachl, 2008b). Stability can be regarded as punctuality multiplied by reliability where reliability is the percentage of trains actually operated and punctuality is the percentage of trains operating "on time" (Khadem Sameni et al., 2010a). This is also the essence of the UK's Public Performance Measure (PPM). Stability of timetable is provided by means of recovery time and buffer time. Slack, running time supplement, standard allowance or recovery time is the extra time added to the running time of trains in order to compensate for the delay of a train. Buffer time is the extra time added to the minimum line headway to avoid propagation of small delays.

Delays occur for a multitude of reasons as shown in Figure 3-4 for Great Britain. Therefore, timetables must be robust enough to recover from a certain level of delays to be reliable. As can be seen in these two sample railways, the infrastructure manager is mainly responsible for the delays that have occurred.

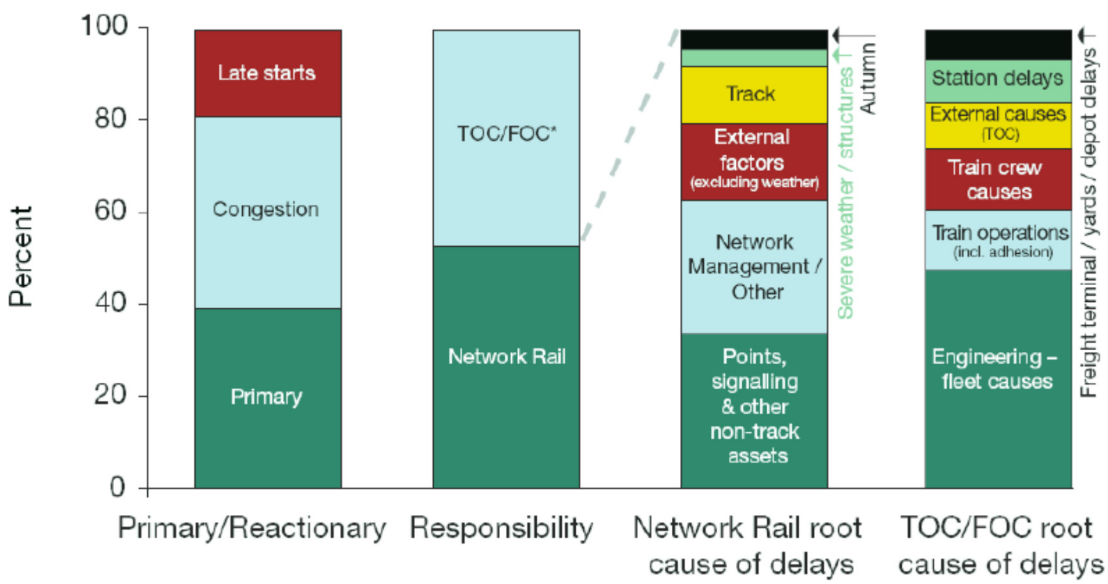


Figure 3-4 - Causes of delay in Great Britain for 2005-2006 (Department for Transport, 2007a)

Figure 3-5 illustrates that the reliability of a timetable is inversely related to capacity utilisation (theoretical capacity is k_{max} and practical capacity is k_f as discussed in section 2.3.1.2). To achieve a very high level of reliability, capacity utilisation must be very low to avoid risk of delays. In this regard, a balance is needed to keep the level of service at a desirable reliability and stability level while utilising the capacity of the infrastructure as far as possible.

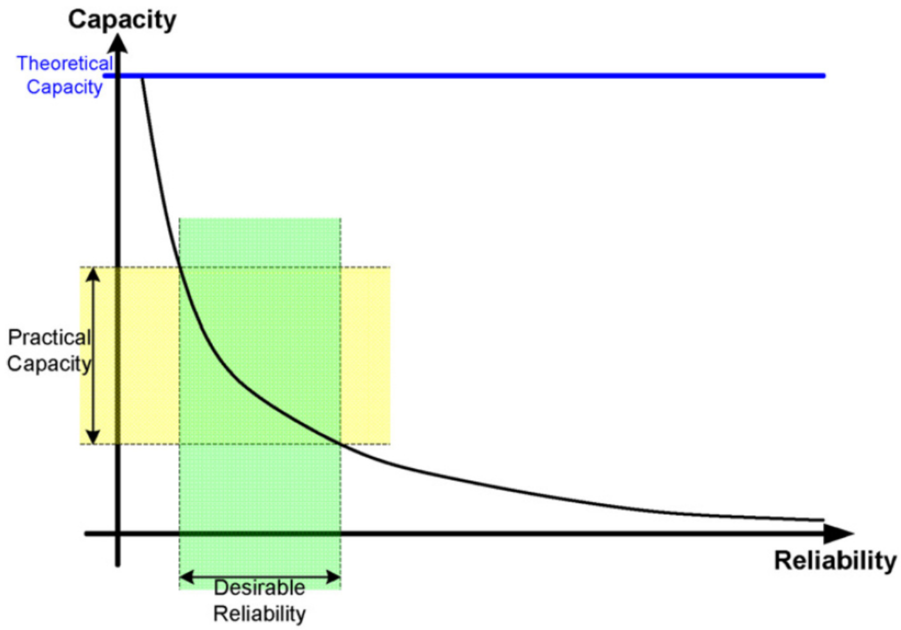


Figure 3-5 – Effect of reliability on capacity utilisation (Abril et al., 2008)

3.1.2.1 Waiting time

Two kinds of waiting time for trains exist: scheduled waiting time and delays in operations (Pachl, 2009). Scheduled waiting time is "an artificial increase in the overall timing of a train which is caused by the resolution of conflicts during the scheduling process" (Hansen and Pachl, 2008b). Total waiting time[1] approaches a vertical tangent which is the theoretical capacity of the line. The maximum number of trains that can be scheduled without buffer time is timetable capacity (Pachl, 2009) (Figure 3-6).

By the waiting time curve, the recommended area of traffic flow can be determined by the following methodology described by Pachl (2009) which is based on the concept of traffic energy. Hertel (1992) first introduced traffic energy by applying the analogy of kinetic energy to railway traffic. It was defined as mass (trains) per unit of length multiplied by the square of average speed or simply the traffic density multiplied by average speed as in the following equation:

[1] Total waiting time= scheduled waiting time + delays

$$E_{traffic\ energy} = \frac{n}{s}.v^2 = \frac{n}{t}.\frac{t}{s}.v^2 = \frac{n}{t}.v$$

$E_{traffic\ energy}$: *traffic energy*

n : *number of trains*

s : *length of the line*

t : *time*

v : *average speed*

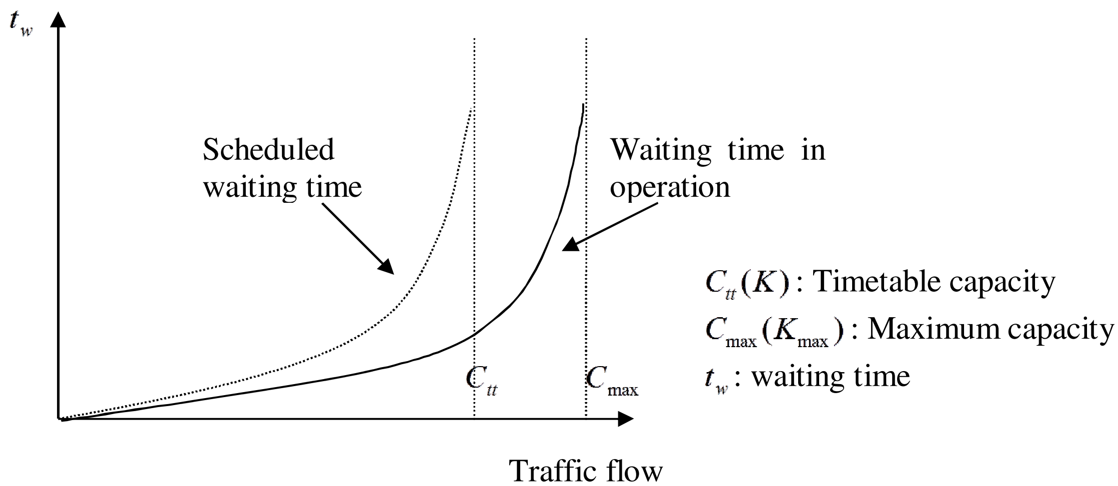


Figure 3-6 - Scheduled waiting time versus timetable capacity (Pachl, 2009)

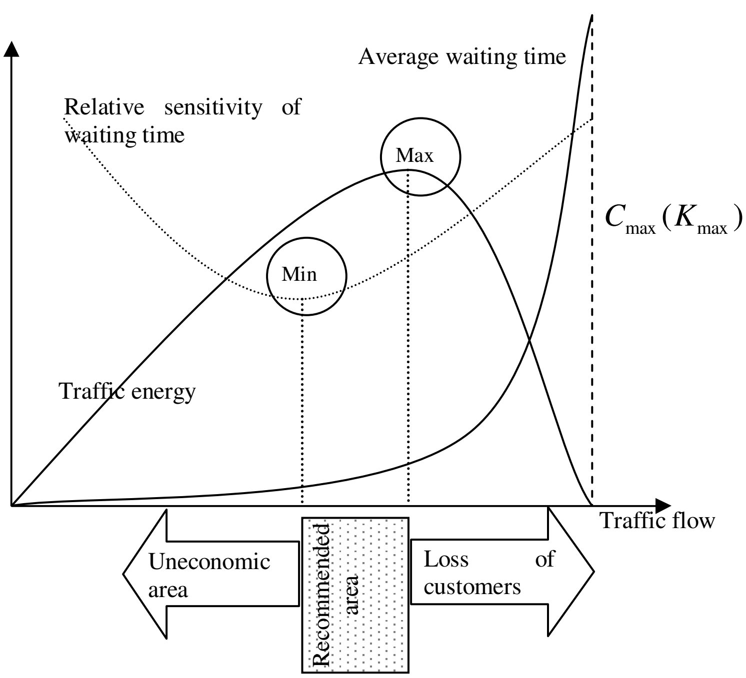


Figure 3-7 - Recommended area of traffic flow (Pachl, 2009)

As illustrated in Figure 3-7, the sensitivity of the waiting time is the differentiation of the waiting time curve. The relative sensitivity of the waiting time is the ratio of the sensitivity divided by absolute waiting time which has a minimum. This is the point up to which, increasing traffic causes reasonable increase in waiting time. Any traffic level less than that point is wasting the capacity. After this point, waiting time considerably increases if extra traffic is added until the theoretical capacity is reached and the exponentially increasing average waiting time approaches the vertical tangent line. There is a point where traffic energy reaches a maximum after which it begins to decline due to the congestion and decreasing average speed. The recommended area for traffic is between the minimum of the relative sensitivity of waiting time and the maximum of traffic energy. According to simulations of various European railways, the minimum of relative sensitivity of the waiting time is reached at about 50% of maximum capacity (60% for more homogenous traffic) and the maximum of traffic energy is from 60% rising to 80% for more homogenous traffic (Pachl, 2009).

3.2 Signalling

Capacity utilisation is dependent upon the number of trains that can safely pass a line. With steel wheels running on steel rails, the coefficient of adhesion in railways is on average eight times less than road transportation which necessitates long braking distances (Pachl, 2008). The heavy mass of trains and their speed result in such high kinetic energy that even a slight accident can have very severe consequences. Railway traffic on the network is regulated by means of signalling which pursues six major goals:

- Controlling trains in a safe manner
- Maintaining safe distances between trains
- Preventing conflicting movements of trains
- Ensuring that points are locked in the correct position
- Enabling running of trains at the required headways
- Enabling operations of trains with minimum disruption

(Bonnett, 2005)

The evolution of signalling technology in railways can be simplified as in . At all stages, rules play an important role in the safe operation of trains as well as railway capacity utilisation. (e.g. How to pass a red or black signal, what to do in case of system failure, etc.)

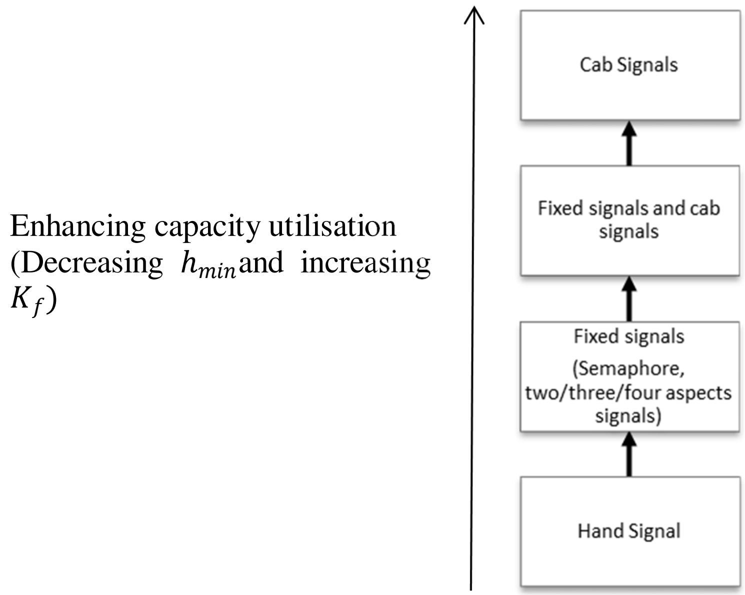


Figure 3-8 Evolution of signalling in railways and impact on railway capacity - Author's own illustration)

In the early days of railway transportation (when speed was very low), policemen were employed by railways to arrange the traffic at stations and level crossings. Many railways adopted fixed signals by the middle of the nineteenth century. Early fixed signals were semaphores which later on were replaced by traffic light signals (Bonnett, 2005). They can be two aspects (green and red), three aspects (green, yellow and red or double green, green and red) or four aspects (green, double yellow, yellow and red) that provide necessary information to the train drivers. In modern signalling, line-side signals are not required as the necessary information is displayed in the driver's cabin. Automatic Train Operation (ATO), Automatic Train Protection (ATP) and Automatic Train Control (ATC) are complementary systems that ensure safer operation of trains and can enhance capacity utilisation.

Technical details about various signalling and control systems are beyond this thesis, but the interactions of signalling with capacity utilisation are of interest. In order to understand how modern signalling affects and can improve capacity utilisation, the concept of the 'blocking time stairway' is critical. Blocking time is defined as "the time interval in which a section of track is exclusively allocated to a train and therefore blocked for other trains". The blocking stairway is "a graph displaying the blocking time of all block sections that a train passes into a time-distance graph" (Pachl, 2009). As shown in Figure 3-9, blocking time can be summarised as:

$$Blocking\ time = $$
$$Physical\ occupation + clearing\ time + approaching\ time + switching\ time + reaction\ time$$

(Wendler, 2007)

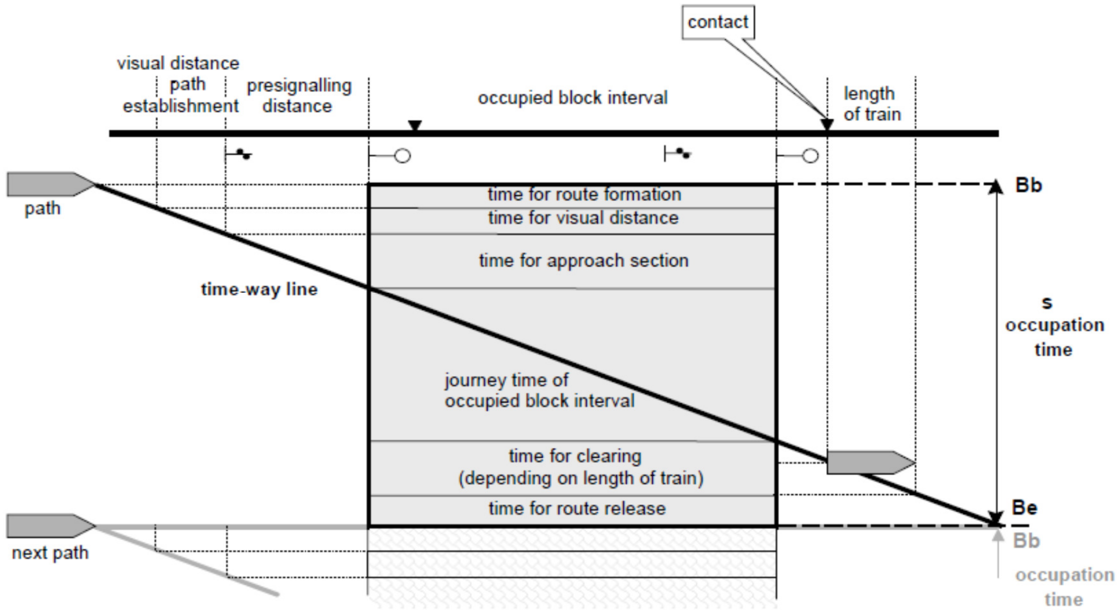


Figure 3-9 - Blocking time and its components (UIC, 2004)

Reducing the safety distance between trains and decreasing headways can improve capacity utilisation to a great extent. The European Cab-based signalling (ERTMS) is a signalling system that can significantly improve capacity utilisation by more efficient train control while increasing safety. ERTMS level one (Figure 3-10) is based on track-train communication where trackside equipment reads the signals and passes the information to the train. The on-board computer controls the speed, authority and limit of movement. At this level, trackside signalling is still in use. At ERTMS level two (Figure 3-11), there is no need for trackside signalling as there is continuous communication between the train and the radio block centre that authorises train movements. Continuous information keeps the driver updated about the traffic and signals ahead. While maintaining a safe braking distance, this enables higher operational speeds and reduced headways thus more efficient capacity utilisation. ERTMS level three (Figure 3-12) is in the conceptual phase and intends to introduce moving blocks which can decrease headways further. (UNIFE, 2009a, UNIFE, 2009b, Climent, 2009, Railway Safety and Standard Board, 2002)

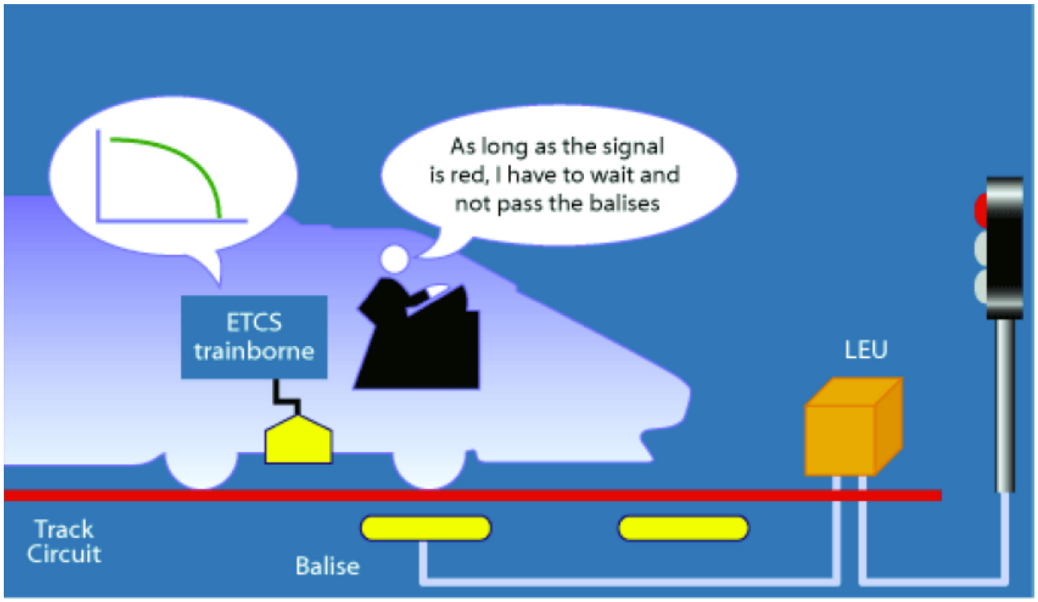


Figure 3-10- ERTMS level 1 (UNIFE, 2009a)

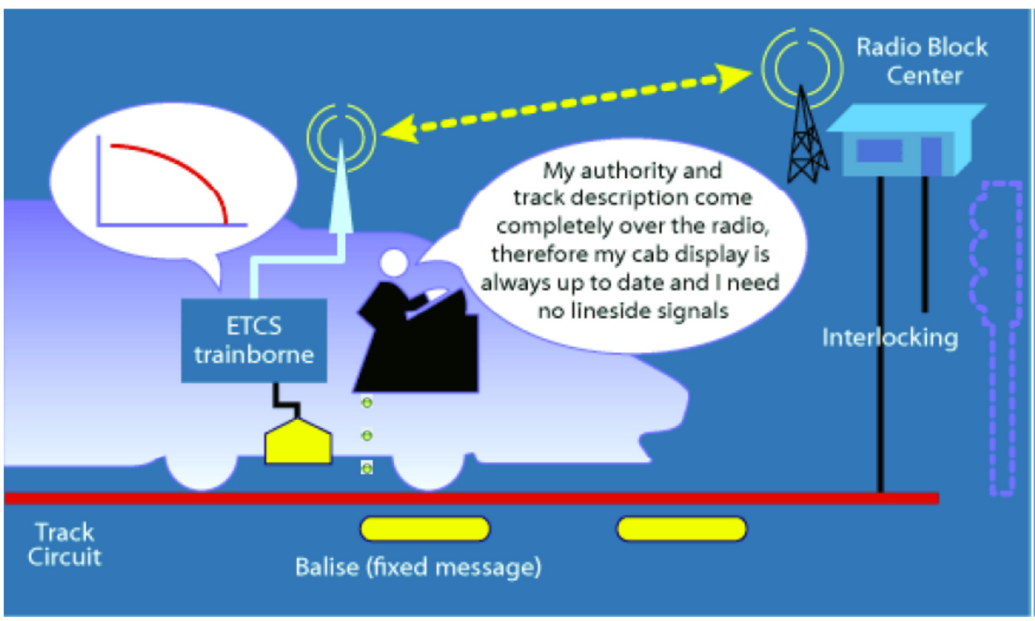


Figure 3-11- ERTMS level 2 (UNIFE, 2009a)

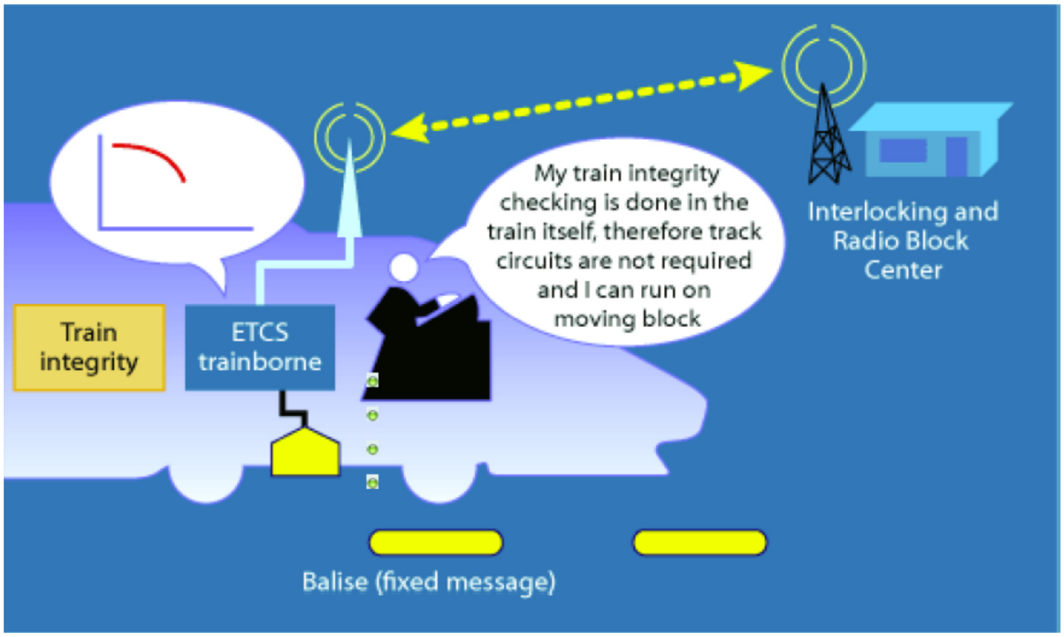


Figure 3-12- ERTMS level 3 (UNIFE, 2009a)

ERTMS level one (Figure 3-13), just like the traditional signalling, is based on 'fixed block' and fixed braking distance. However, the lower deceleration rate in ERTMS level one results in lower capacity utilisation.

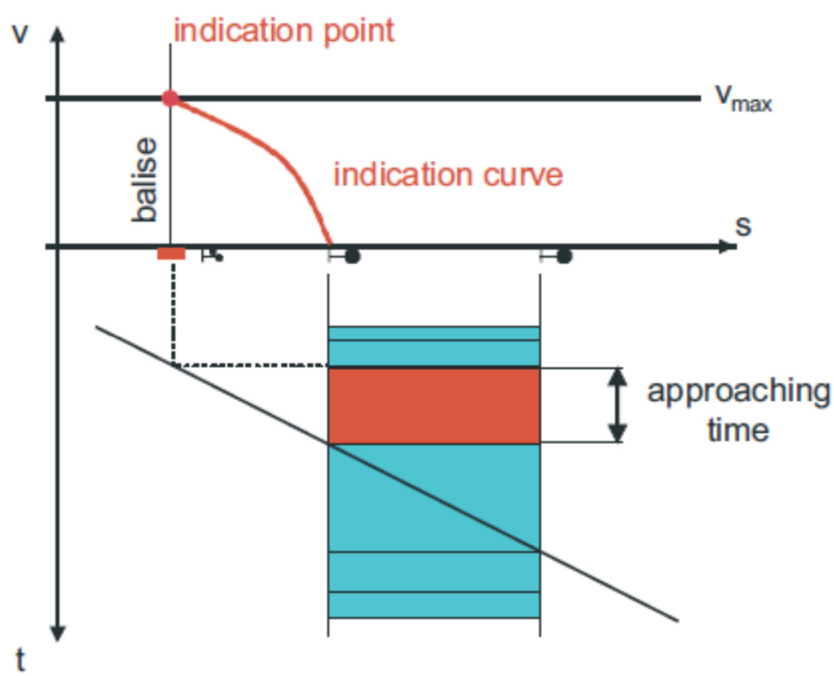


Figure 3-13 - Effects of ERTMS level one on blocking time (UIC, 2008)

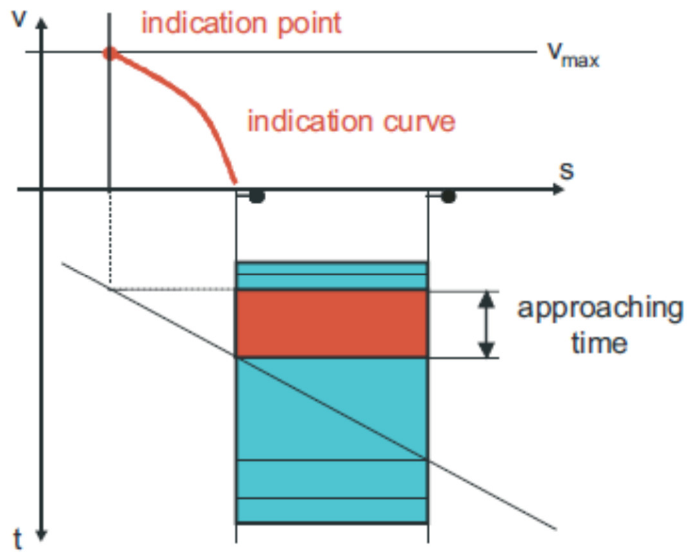
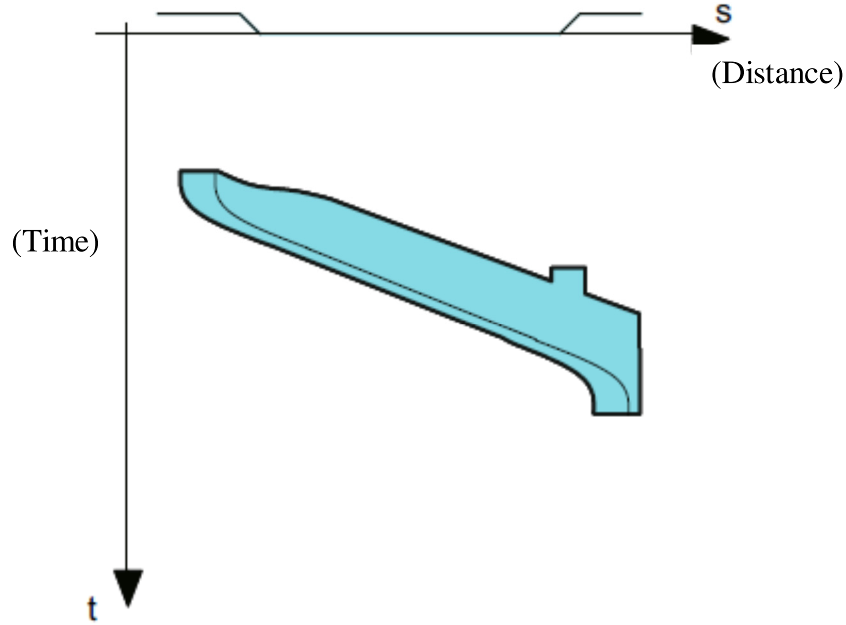


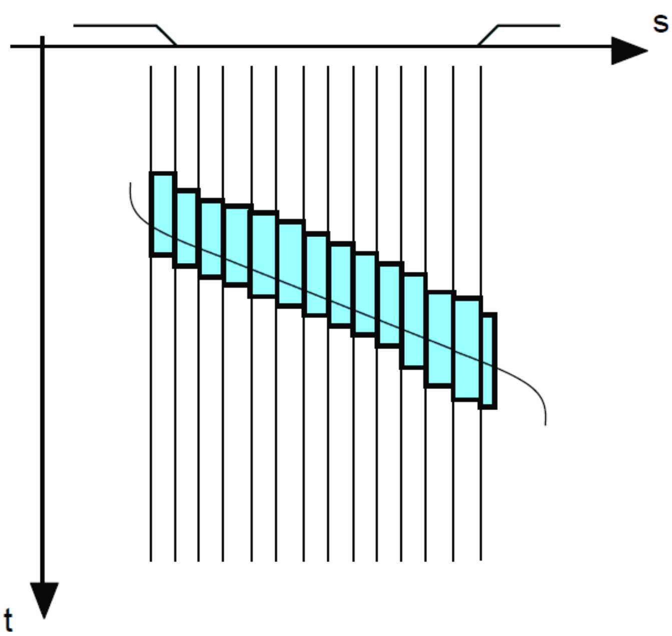
Figure 3-14 - Effects of ERTMS level two on blocking time (UIC, 2008)

ERTMS level two is also based on a 'fixed block system' but it eliminates the time needed for the driver to see the trackside signals as there is cab signalling which is displayed on-board. ERTMS level three also eliminates the time needed for the driver to see the trackside signal because of cab signalling. On the other hand, moving blocks are dynamic and can be much shorter than fixed blocks (Figure 3-15). Higher speed and shorter blocks decrease the journey and clearing time of the occupied block. Therefore more trains can be accommodated on the track which improves capacity utilisation.

In an email on 19th October 2012 Professor Joern Pachl summarised the impact of cab signalling on blocking time. He states that "In cab signalling, the approach time is no longer the running time between distant and main signal but the running time within the dynamic braking distance. Since the braking distance is a function of the square of the speed, the approach time is a function of speed. Running at higher speeds will reduce the running time within the block section but increase the approach time. Depending on the block and train length, at which the blocking has a minimum. By making the block length zero, ERTMS Level 3 eliminates the running time within the fixed block section. The approach time, like in ERTMS level 2, is the running time within the dynamic braking distance."



a) Blocking time band (moving block)



b) Graduated blocking-time band

Figure 3-15 - Effects of ERTMS level three on blocking time (UIC, 2008)

Impacts of different ERTMS levels and variants on capacity utilisation are summarised in Table 3-1. A later study by RWTH Achen University, commissioned by the UIC, quantifies the effects of ERTMS for typical high speed, conventional and regional case studies. As discussed by Wendler (2007), it is concluded that the most effective increase in capacity utilisation with the aid of ERTMS is achieved for high speed lines. ERTMS level two shows a slight increase in capacity compared to using level one. However, using ERTMS level two while optimising block sections can significantly increase capacity utilisation. ERTMS level three has the highest potential for increasing capacity which can be slightly more than ERTMS level two when block sections are optimised simultaneously. Figure 3-16 shows the impact of different levels of ETRMS on capacity utilisation.

Table 3-1 Summary of ERTMS variants and their impacts on capacity utilisation (Railway Safety and Standard Board, 2002)[1]

ERTMS Level	System Variants		
Level 1 - Movement authority from existing line-side signal passed to train via switched balise	**System A** (No infill) - ***significant reduction in capacity*** due to severely limited signal update	**System B** (single infill) - ***smaller reduction in capacity*** compared with System A (further infill points would mitigate capacity reduction but will significantly effect maintenance and costs)	
Level 2 - Movement authority from Radio Block Centre passed to train. Radio network provides continuous infill	**System C** (Lineside signals retained and ERTMS "overlaid") - ***broadly neutral effect on capacity***	**System D** (Lineside signals removed or minimised - cab signalling) - ***offers potential for significant capacity increase***	**System E** (Low density application with much reduced trackside infrastructure, for use on regional lines with low frequency services) - ***capacity benefit is not the main reason for implementation***
Level 3 - Builds on Level 2 - train provides safety critical position data	**NOT MODELLED** - Because there is no direct work on Level 3 in Europe at the present time		

[1] A later study analyses the impacts of ERTMS level three on capacity utilisation. INTERNATIONAL UNION OF RAILWAYS 2009. *Compendium on ERTMS: European Rail Traffic Management System*, Eurail Press.

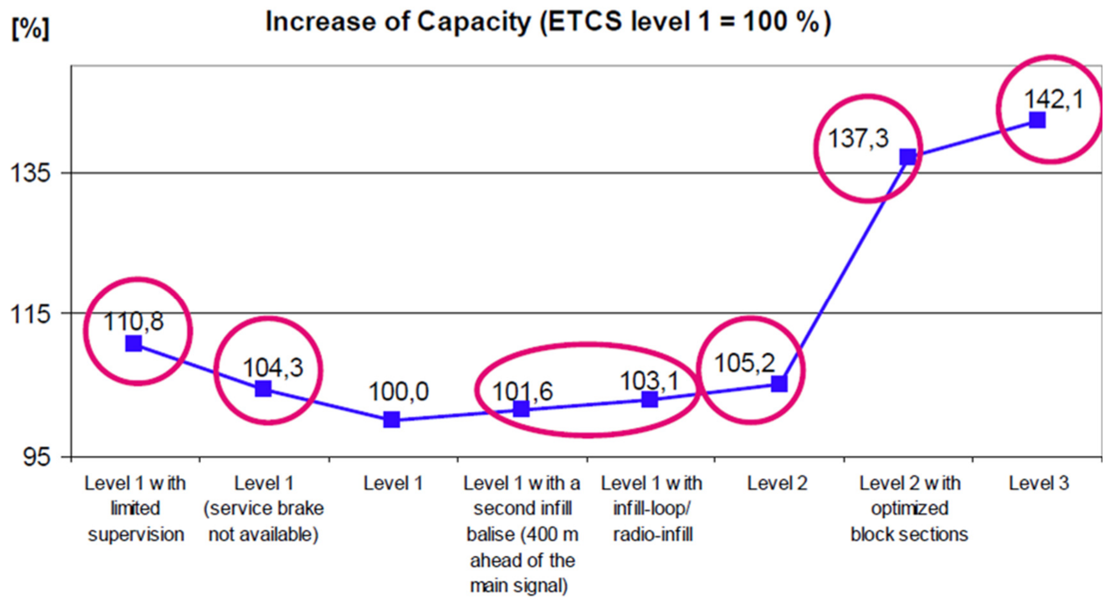


Figure 3-16 Impact of different levels of ERTMS on conventional main line (International Union of Railways, 2009)

3.3 Nodal capacity constraints

Stations and junctions are usually the bottlenecks of railway networks as traffic merges, stops, originates or terminates there whereas in links traffic flows. Passengers board and alight and change trains at stations which needs adequate platform allocation and dwell time. All the trains passing or stopping at nodes need to be properly routed and conflicting movements that limit capacity unitisation should be minimised. Overtaking of trains can occur at stations. Trains may be coupled or de-coupled if necessary. As shown in Figure 3-17 there are various capacity constraints at nodes of the railway network. They have been divided into soft constraints and hard constraints. Soft constraints are those related to operation and hard constraints are infrastructure-related constraints.

Currently there is no holistic measure for systematically assessing these constraints and quantifying capacity utilisation at nodes. However, some parts of these constraints have been addressed in the literature and they are presented in sections 3.1.1 to 3.3.6.

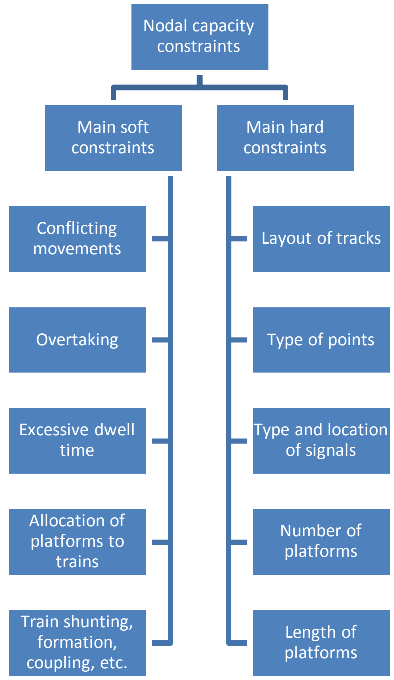


Figure 3-17 - Major nodal capacity constraints

3.3.1 Train routing at stations

There are usually several platforms at big stations and conflicting movements can occur which affect capacity utilisation. In this regard, some studies have focused on routing the trains in complex stations more efficiently.

"Avoiding conflicting movements" is part of Pro Rail's "Triple A" strategy in the Netherlands to improve capacity utilisation[1].
Figure 3-18 and Figure 3-19 show how routing through Utrecht station has been improved which has resulted in fewer dependencies, smoother operations for staff and passengers, better capacity utilisation due to less conflicting movements and punctuality enhancement by 2% (Kraaijeveld, 2009).

[1] The "Travelling Without a Timetable" initiative in the Netherlands is based on triple A or different ("Anders") approaches to "Planning and operations", "Capacity allocation" and "Capacity enhancement". KRAAIJEVELD 2009. Making room on the rails. *Growth and capacity challenge-an international perspective*. London, UK..

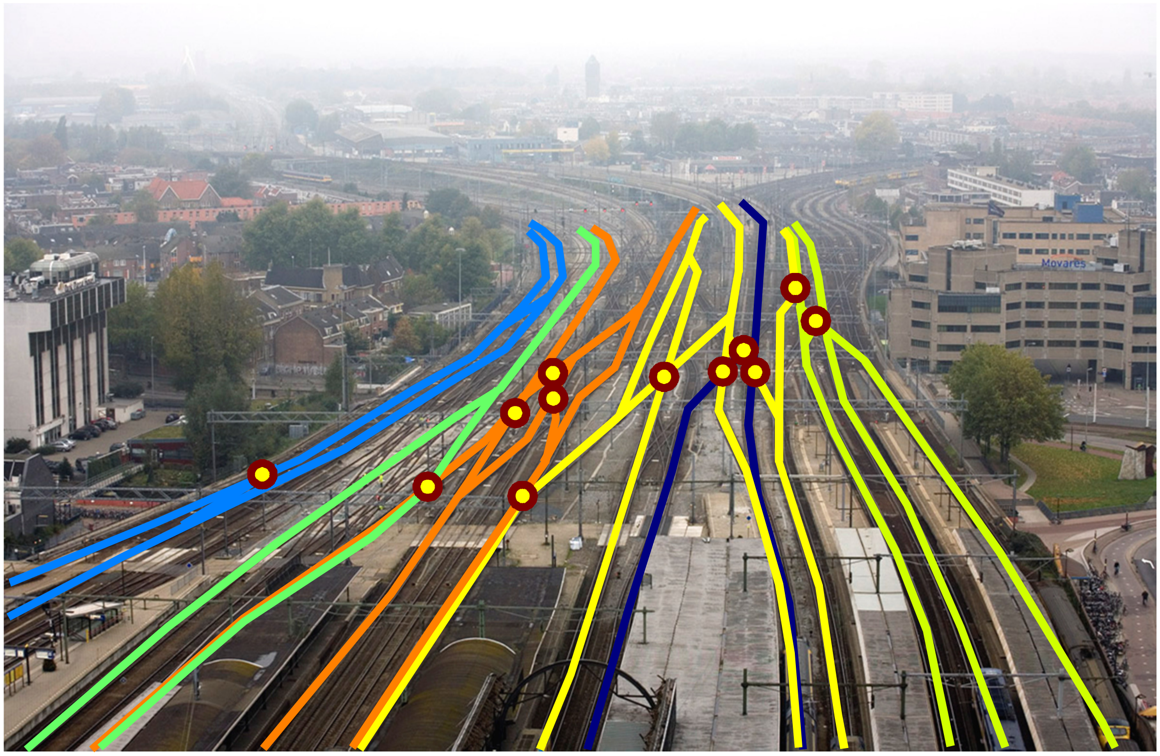


Figure 3-18 - Utrecht station before improvements (Kraaijeveld, 2009)

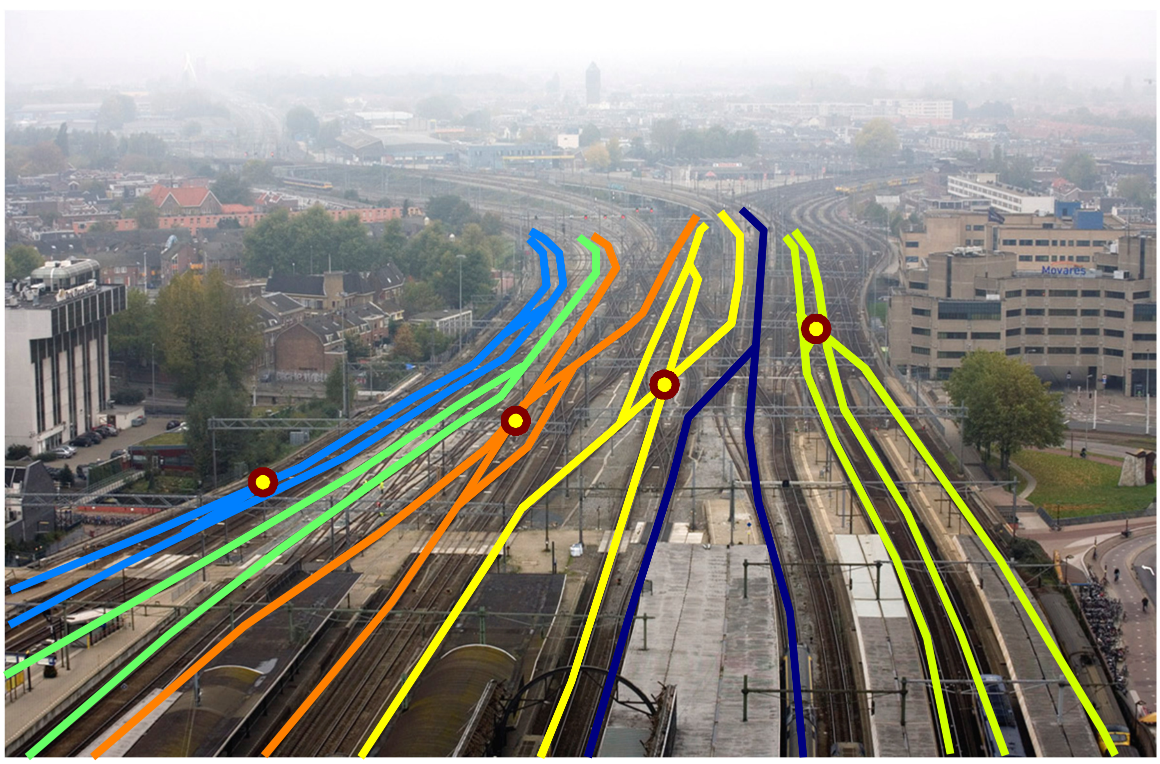


Figure 3-19 - Utrecht station after improvements (Kraaijeveld, 2009)

For strategic planning, the feasibility of allocating platforms to a combination of trains has been studied by Zwaneveld et al. (1996). Kroon et al. (1997) investigate the complexity of this problem which is proved to be NP-complete[1] when trains have more than two options for routing. Follow-up work by Zwaneveld et al. (2001) presents a node packing model to route the trains through stations. The above-mentioned works form the basis of the 'STATIONS' decision support system in the Netherlands that routes the trains through stations based on the available capacity, safety and service requirements.

The manual process of scheduling and routing trains through complex stations is mathematically modelled by Carey and Carville (2003). They introduce binary variables to consider platform feasibility (connection, appropriate length, special needs, etc.) for each train, using platform desirability. Some costs are considered to differentiate between more desirable platforms as well as platform obstruction costs when a platform accommodates more than one train (as one train obstructs the path of the other). If trains cannot be assigned to the most desired arrival and departure times (due to occupation of platforms and conflicts), time adjustment costs are considered. Trains scheduled and routed in chronological order, path/platform conflicts and minimum headways are checked and resolved. Carey and Crawford (2007) extend this work to a network including corridors as well as stations.

3.3.2 Dwell time

Dwell time of trains at stations affects the capacity utilisation of the network. The more stops a train makes at stations, the longer the overall journey takes. Each stop also has an effect on other trains as there must be a safety distance between two consecutive trains travelling in the same direction. Therefore when a train stops at a station, it affects the following train. Moreover, the stop of a train at a station makes the next block after the station idle. Stop of a train is a trade-off between infrastructure capacity utilisation and avoiding lost demand. Unfortunately no research has been found on the value of a stop for a train, to quantitatively determine where it would be worthwhile for a specific train to have a stop by considering overall capacity utilisation as well as the estimated number of passengers boarding and alighting. Existing research on the value of passenger time could help with this aim.

Dwell time at stations depends on several parameters: vehicle type (number, position, width of doors, etc.), infrastructure (platform length and level) and demand (number and distribution of passengers). Buchmüller et al. (2008) study the five sub-processes of dwell time at a station: unblocking the doors, opening the doors, passengers boarding and alighting the train, closing doors and train dispatching. They develop a model to estimate required dwell time at stations in Swiss Federal Railways (SBB).

---

[1] In computational theory, NP-complete problems are a class of problems that can only be solved in polynomial time by non-deterministic methods.

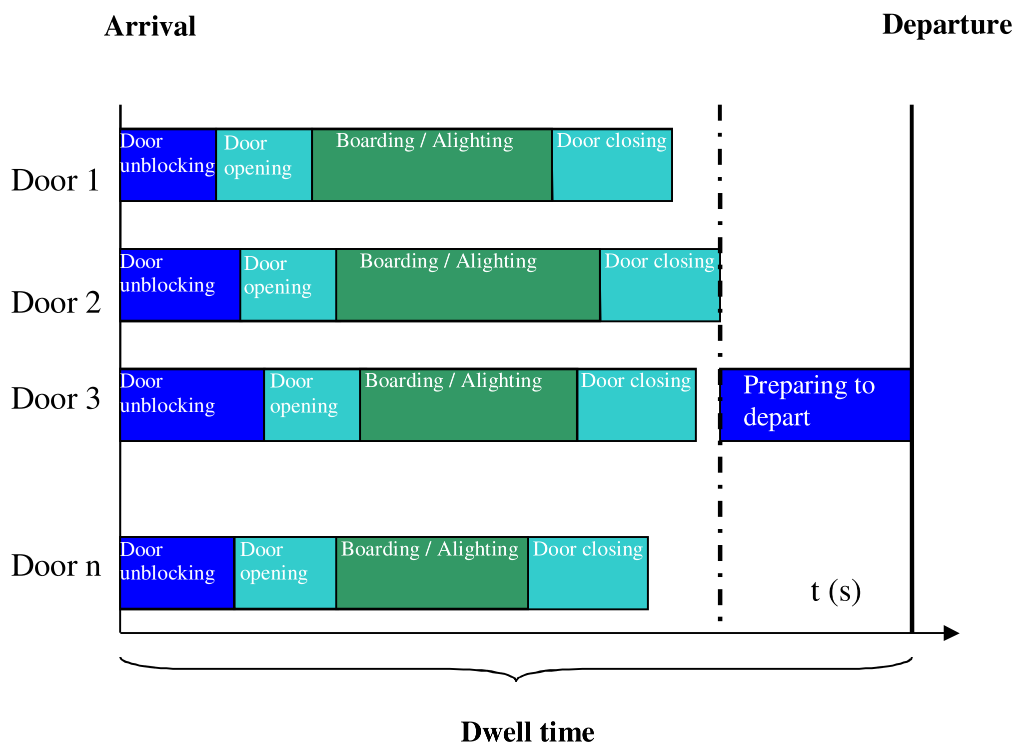


Figure 3-20 - Dwell time at stations (Buchmüller et al., 2008)

Dwell time supplements aim to compensate for excessive dwell time at stations to avoid disruptions to the timetable.

3.3.3 Layout of crossings

At nodes, the layout of crossings affects conflicting movement as well as flexibility of changing tracks. Figure 3-21 shows major crossing layouts.

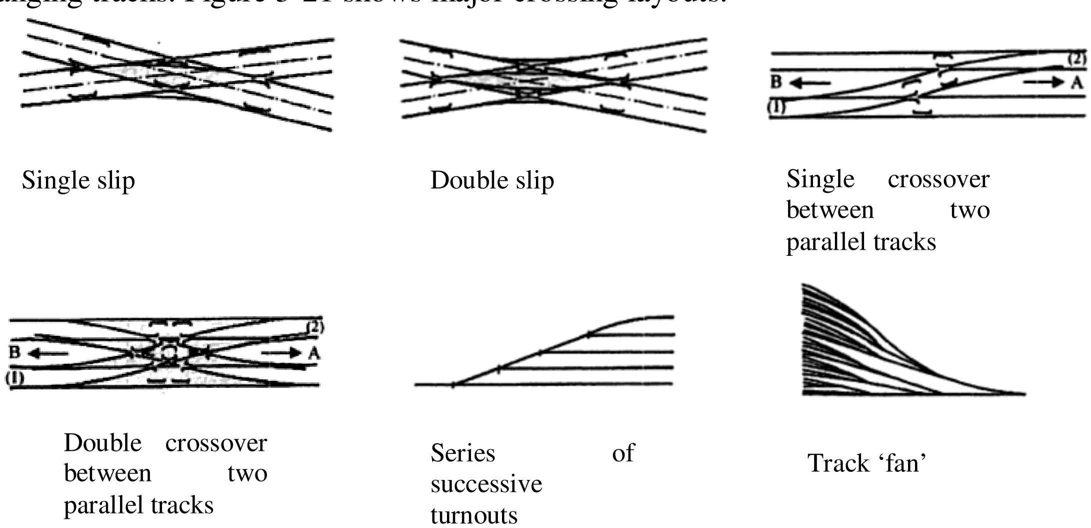


Figure 3-21 - Major crossing layouts (Profillidis, 2006)

When the traffic volume is very high and considerable delay is caused due to conflicting movements at level crossings, it might be worthwhile to build a flyover.

3.3.4 Number of stations

The number of junctions and crossings usually cannot be decreased as they are needed for changing tracks and routes. But it is possible to increase or decrease the number of active stations in a network. As discussed in 3.3.2, the more a train stops at different stations, the longer its journey takes, hence the blocking time of railway infrastructure. Moreover, with the increasing value of time for passengers, slower journeys are not desirable for onboard passengers and might affect their choice of mode and result in reduced system demand. In this regard, existing underutilised railway stations might be closed down and just function as links for the greater good of the overall system. This should be done with careful consideration of not only the freed capacity but also the lost demand. Care should be taken not to reduce accessibility to the railway to the extent that passengers shift to other modes. The same rules apply for adding new stations. In essence, accurate studies should be undertaken to find the optimum balance between attracting new passengers, losing current passengers and the capacity utilisation of the infrastructure (Figure 3-22). Time savings to new passengers (additional revenue) should be greater than time loses to existing passengers (lost revenue).

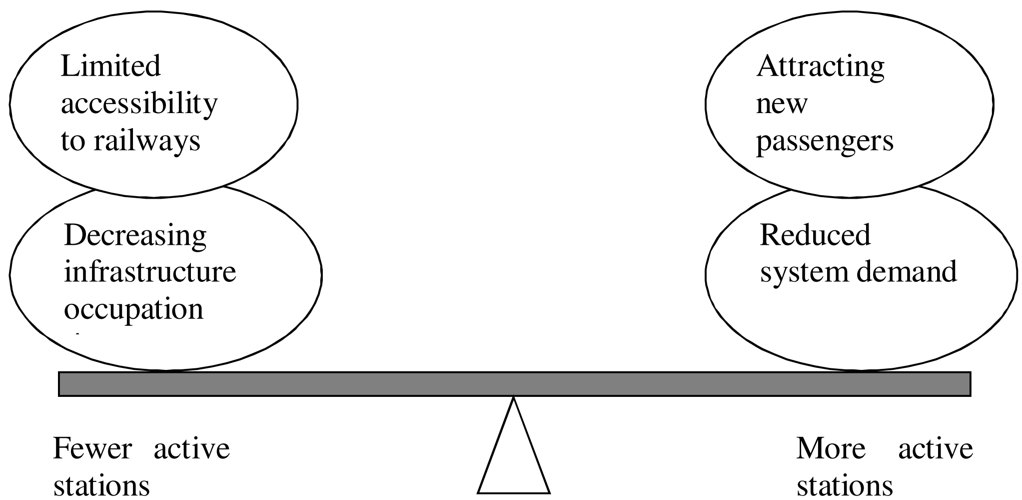


Figure 3-22 - Balance between number of stations and micro/macro capacity utilisation

3.3.5 Number of platforms

As a rule of thumb, the more trains stop at a station, the more platforms are needed to handle traffic as they act like servers to trains. Moreover, there is a ‘minimum platform reoccupation time’ needed between two consecutive trains travelling in the same direction that is mainly determined by type of signalling that is used.

3.3.6 Length of platforms

The length of platforms affects the length of trains that can be accommodated at stations. In some stations, longer trains stop at short platforms. In this case, passengers are asked to move to front coaches to alight the train. Selective door opening might be used as well. Under certain circumstances and according to the operational rules of a railway, two trains might be allowed to share the same platform (to couple, de-couple, etc.). Extending the length of platforms is desirable but costly. It is estimated that one metre extension of a 2.5 metre wide platform would cost between £3,750 (estimated by Office of Rail Regulation) to £5,000 (estimated by Network Rail) (Department for Transport and the Office of Rail Regulation, 2010). These estimations are considerably higher than previous ones estimated by Franklin + Andrews Ltd. (2004).

3.4 Rolling stock

The discrete steps of railway capacity utilisation at macro level are used by trains. Hence, many characteristics of rolling stock affect capacity utilisation which are briefly listed below.

3.4.1 Speed, acceleration, deceleration

The blocking stairway of a train (which is a time-distance graph as discussed in 3.2), is directly proportional to the speed of train which is how much distance can be travelled in the unit of time. The faster the train, the less time it occupies sections of the line and blocks it for other trains, hence increasing opportunities for more efficient capacity utilisation at macro level. Acceleration and deceleration affect the ratio of time it takes the train to change its speed, hence also affecting the blocking stairs, and are especially important for reaching the required speed from standstill and vice versa. Speed, acceleration and deceleration all depend on tractive effort[1] (diesel or electric) as well as the technical specification of the rolling stock.

3.4.2 Door mechanism

For passenger trains, the number of doors, their width and mechanism of unblocking, opening and closing (sliding, bi-parting, etc.) affect the dwell time required for alighting and boarding the trains. For safety reasons, it is important that the doors allow for one-man operation. Potential savings from upgrading doors can be substantial when considering the total number of trains on the network and the total stops at stations.

---

[1] "The effort of a locomotive or a multiple unit which is intended to move the train." HANSEN, I. A. & PACHL, J. (eds.) 2008b. *Railway timetable and traffic,* Hamburg: Eurailpress.

3.4.3 Number of coaches

The number of coaches per train affects the maximum number of passengers that can be carried. For freight trains, the number of wagons affects total tonnage. For locomotive hauled trains this also affects train speed, acceleration and deceleration. In saturated networks it is desirable to have long trains. However, there are limitations for the length of trains. Signalling constraints and the length of platforms limit the allowable length of trains.

3.4.4 Double-deck and single deck

Double deck trains can be considered as two trains merged into one; they nearly double the number of passengers carried per train while blocking the infrastructure just as one train. Double-deck trains need more dwell time at stations but overall they greatly enhance capacity utilisation. However the use of double deck trains is constrained by the loading gauge[1] of the network. In this regard bridges and tunnels are the main determining factors. Double-deck coaches are operational in many counties including European countries (Germany, France, Netherland, etc.) and in North America. However, the loading gauge in Great Britain has ruled out using double deck trains. It should be noted that boarding and alighting for double-deck trains needs longer dwell time.

3.5 Infrastructure

Infrastructure parameters affect the flow of trains on the network. These parameters can limit the operational speed of trains, maximum tonnage, width and height of trains.

3.5.1 Number of tracks

The number of tracks greatly determines the overall line theoretical capacity. On single track lines, trains going in opposite directions conflict with each other which greatly increases the scheduled waiting time. It is estimated that double track line often quadruples the theoretical capacity of the line (Nash, 1982). In the case of multiple-track lines (e.g. 4 tracks), for each direction usually one line is dedicated to slow trains and the other to fast trains. This decreases the effect of heterogeneity on capacity utilisation (section 3.1.1 Heterogeneity).

3.5.2 Line speed

The speed at which trains can travel on tracks is a decisive factor for railway capacity utilisation. The faster the trains travel, the less time the infrastructure is occupied and the

[1] The loading gauge represents the maximum width and height to which a railway vehicle may be built or loaded for ensuring safe operation. MUNDREY, J. S. 2000. *Railway track engineering,* New Delhi, Tata McGraw-Hill Publication.

more passengers and goods can be transported. Trains cannot usually use their full speed due to the constraints imposed by infrastructure or such as line speed. Therefore the operational speed of railway is:

$$\text{Operational speed} = \min \{\text{line speed}, \text{rolling stock speed}\}$$

The constraints affecting line speed are mainly:

- Curvature
- Gradient
- Track and subgrade conditions

Curvature and gradient are as far as possible avoided and kept to minimum in the design of railway lines. However, wherever they exist, they impose considerable speed reduction due to enormous centrifugal force in curves and grade resistance.

Track and subgrade conditions including the age of the infrastructure, how well it is maintained, types of sleepers, subgrade stability, etc. affect line speed. For instance wooden sleepers enforce lower line speeds as compared to concrete sleepers. In Great Britain, where railways originated and thus the oldest infrastructure exists, modern rolling stock (like class 220 Voyager DMUs and class 221 Super Voyager DMUs with the maximum speed of 125 mile per hour or 200 kilometer per hour) can not operate at full speed A major line upgrade as in the West Coast Main Line can significantly increase the line speed, closing this speed gap.

3.5.3 Axle load

Railway lines are designed for bearing a maximum axle load. A line with a high permitted axle load allows the heavier freight trains to pass the line which increases the amount of transported freight. Axle load is interlinked with depth of ballast and subgrade, type of subgrade soil, rail profile, rain fall, sleeper spacing, type of sleeper, etc. (Mundrey, 2000).

3.5.4 Loading gauge

The loading gauge represents the maximum width and height to which a railway vehicle may be built or loaded for ensuring safe operation (Mundrey, 2000). The loading gauge affects the width and height of trains that can use the lines and is mainly determined by bridges and tunnels. This influences both passenger and freight trains. Wider or double-deck trains can carry more passengers. As described in section 3.4.4, double deck trains are not feasible in Great Britain due to the limited loading gauge. The same problem exists for handling new generations of ‘tall containers’ that are 9 foot, 6 inches high (2.9 metre) (Lowe, 2005).

3.5.5 Maintenance and engineering work

In railway transportation, steel wheels running on steel rails cause wear which risks causing derailment. Therefore, much more constant maintenance is needed than for road transportation. In air and marine transportation this need is even less as only the nodes (airports and ports) need maintenance and not the main natural infrastructure (air and water). Railway engineering works interfere with railway operation and affect capacity utilisation. Albrecht et al. (in press) and Macbeth and de Opacua (2010), review various strategies for track maintenance scheduling

3.6 External factors

Capacity utilisation can be affected by external factors, especially weather conditions. Operators can not control weather conditions but they can adopt precautionary and mitigating measures to decrease the negative impacts of severe weather conditions on railway operation and capacity utilisation.

Fallen leaves cover the tracks, are compressed by passing trains and form a teflon-like coating which causes trains to slip and slide and results in delays as well as damage to the tracks and trains. Such a coating can also interfere with the track circuit mechanism and identifying the position of the trains in the signalling system which is why they are nicknamed the “black ice” of the rail industry. They cause problems for many railways especially in Great Britain, the USA, Sweden, Germany and France. Precautionary measures include long-term control of vegetation. Native shrubs are good vegetation and also act as a barrier to trespassers. However, for the safe and reliable operation of railways, trees like sycamore, chestnut, poplar, lime and ash tree should not be close to tracks. Mitigating measures include the use of Multi Purpose Vehicles (MPVs) fitted with laser or sand-based gel, and high-pressure water jets or ‘hot spot’ teams that remove the leaves from the tracks manually. It is estimated that in Great Britain, the annual cost of weather conditions to the railway industry is approximately £50 million including autumn train borne operations, vegetation management, ‘hot spot’ teams and damage to trains and tracks. (Network Rail, 2010a, Network Rail Media Centre, 2006).

Fog, snow and ice reduce the sighting distance of the driver, increase the blocking time and decrease capacity utilisation. (See Figure 3-9). Points may also freeze in cold weather or take longer to operate. Braking distance of trains might also be longer due to the snow on tracks. In addition, ice and snow may insulate the third rail or overhead line causing problems for the traction power of trains. Severe weather conditions can cause major delays and disruptions to train operations. Equipment for tackling these issues include point heaters, a variety of snow clearing machines, anti-icing sprays, miniature snowploughs fitted on the front of trains, Beilhack snow ploughs fitted on individual locomotives and independent drift ploughs. Infrastructure managers and train operations might also agree on an ‘emergency timetable’ in the event of severe disruptions (Network Rail Media Centre, 2006).

Once again the need for a holistic approach to railway capacity should be emphasised; all the factors affecting railway capacity utilisation are important. It is always the 'weakest link of the chain' that determines the overall capacity utilisation. For example the impact of weather can be the main constraint in autumn and winter (Figure 3-23) and should be addressed effectively. Lower Public Performance Measure[1] is an indication of either considerable delay or cancelled services which both adversely affect capacity utilisation.

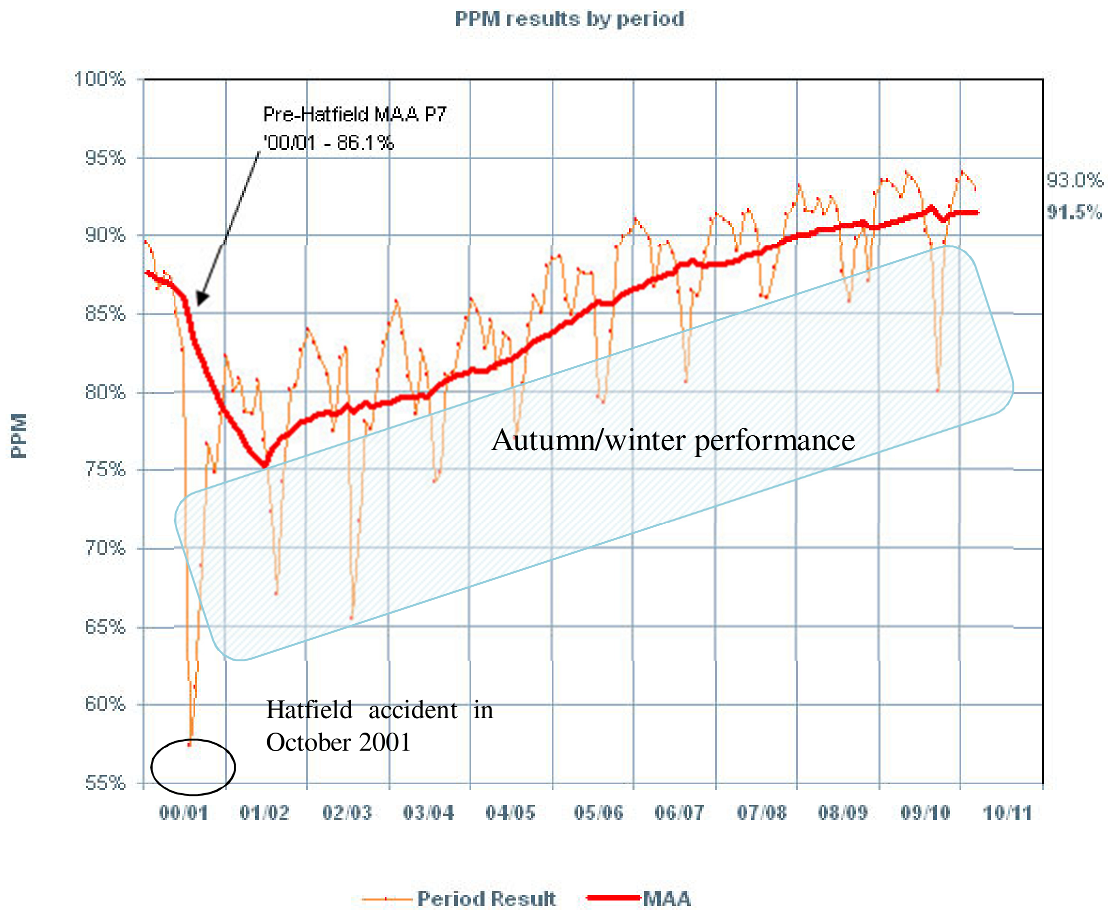


Figure 3-23- Effect of weather on Public Performance Measure and capacity utilisation (Network Rail, 2010d)

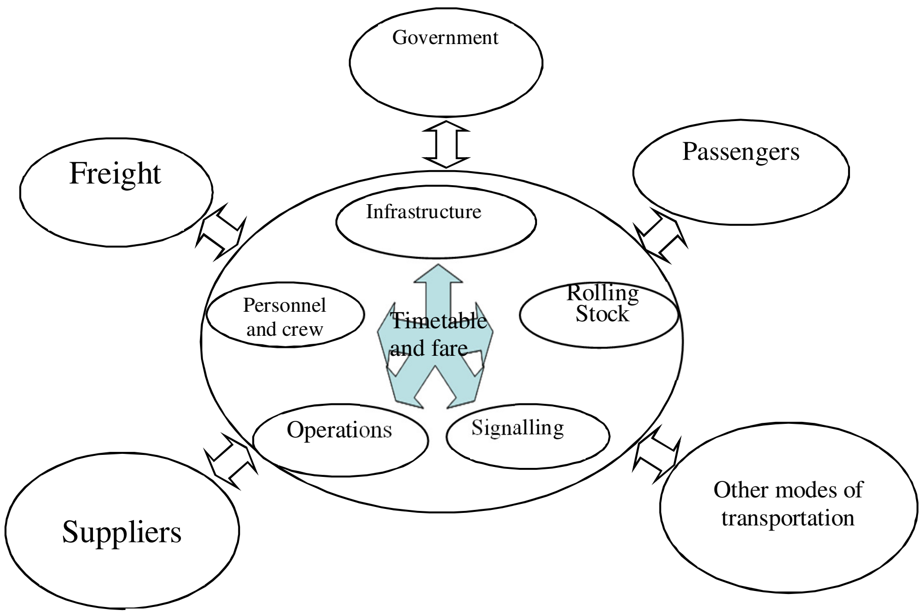
3.7 Governance

Railways are run in different ways in different parts of the world and capacity utilisation naturally varies too. This section studies the structure of railways, access charges and white papers defining policies for capacity utilisation.

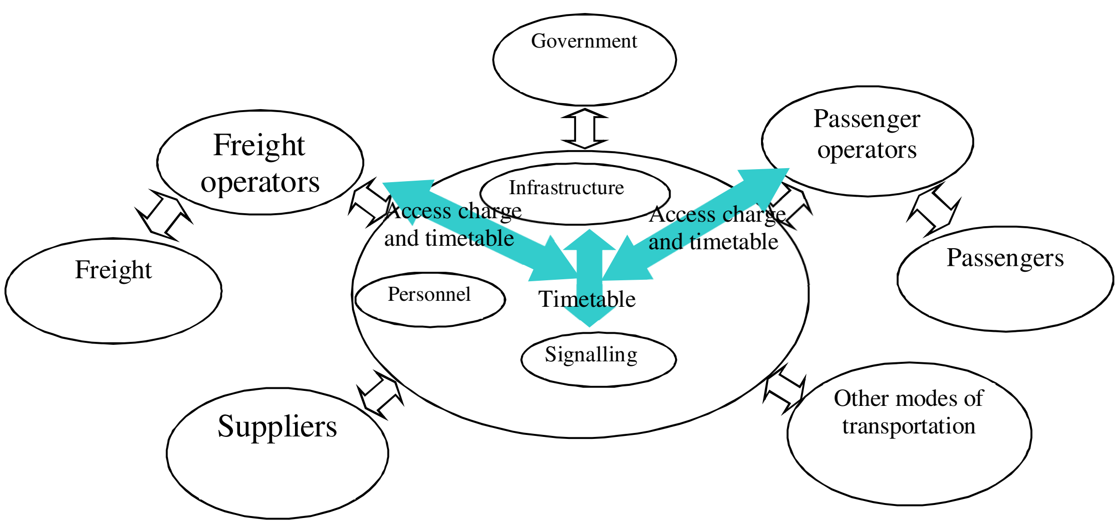
[1] Public Performance Measure (PPM) is the percentage of passenger trains that arrive at their destination on time (not later than 5 minutes for local services and not later than 10 minutes for inter-urban trains). If a train is cancelled or is later than the threshold, it has not met the criteria. The national Public Performance Measure for the year ending 30 April 2011 is 90.8%. NETWORK RAIL. 2011c. *How we measure up* [Online]. London. Available: http://www.networkrail.co.uk/aspx/699.aspx [Accessed 25/07/2011.

3.7.1 Structure

European railways and North American railways are governed in contrasting ways. According to an EU directive (91/440), in European railways the operational and infrastructure sides are separated and the timetable is the financial interface between them (Pachl, 2008). This is usually referred to as “vertically separated” structure. Figure 3-25 illustrates a simplified representation of railway structure before and after privatisation in European railways. In vertically *separated* structure, policy dictates how the government-owned infrastructure is efficiently allocated to private train-operating companies. In North America, railways are vertically *integrated* and lines are usually privately owned by freight operators. Table 3-2 provides an overview of railways in Europe and the USA.



A - Before privatisation (vertically integrated operation)



B - After privatisation (vertically separated operation)

Figure 3-24 - Vertically integrated versus vertically separated railway structure

Table 3-2 - Overview of railways in Europe and the USA (Khadem Sameni et al., 2010)

	Europe	**USA**
Main focus	Passenger	Freight
Timetable	Thorough timetable	Most often no exact timetable
Infrastructure owner	Most often state or state owned infrastructure manager	Mainly privately owned by the operator
Operation and infrastructure	Railway operation is separated from infrastructure management as a requirement of liberalisation stated by the European Union laws. (vertically segmented railways)	Railway operation and infrastructure management are merged together. (vertically merged railways)
Signalling	High technical level – often with ATC/ATP	Often simple signalling
Distance	Short/medium distance	Long distance
Length of trains	Varies	Usually very long
Traction	Electric, some diesel	Diesel

Even within vertically segmented European railways, the market share and degree of competition between operators can affect capacity utilisation. The benchmarking study of four European railways by Civity Management Consultants (2011) suggests that Great Britain has the most competitive market structure and that market shares are distributed among different operators (Figure 3-26) whereas in Sweden and the Netherlands state-owned companies still dominate. However recent studies contend that "Britain's rail infrastructure manager faces an efficiency gap of 40 per cent against European best practice and that train operating costs have also risen substantially, both because of rising factor prices (wages and fuel) and because of deteriorating productivity" (Lovell et al., 2011). Hence "vertical integration" is suggested (Department for Transport and Office of Rail Regulation, 2011).

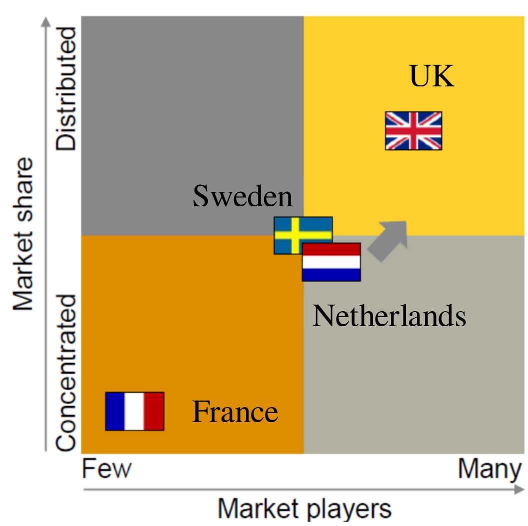


Figure 3-26- Market structure of 4 European railways (Civity Management Consultants, 2011)

3.7.2 Access charge

Access charge and potential franchises economically relate infrastructure owners to freight and passenger operators in a vertically segmented railway. Capacity allocation does not rely on explicit prices or values as it was practised in British Rail before privatisation: Following privatisation a set of decision criteria was formalised as listed by Gibson (2003):

- Sharing capacity in the most efficient and economical manner
- Enabling compliance with the Passenger Service Requirements[1] (first/last trains, frequency, capacity etc.)
- Maintaining and improving reliability
- Carrying out necessary maintenance and renewals
- Maintaining and improving connections
- Avoiding deterioration of service patterns
- Ensuring the pattern of rail services reflects the pattern of demand
- Reserving capacity for short-term bidders
- Enabling operators to utilise their assets efficiently
- Facilitating new commercial opportunities
- Avoiding frequent timetable changes

Charging for scarce capacity can be done through different methods each of which has been studied by different researchers:

- Auctioning and bidding for scarce slots (Nilsson, 2002)
- Calculating the opportunity costs (Johnson and Nash, 2008)
- Including incremental investments needed for increasing the capacity at bottlenecks (Hylen, 1998)

A short review of rail infrastructure charging in different European countries is done by Hylen (1998). He categorises them into:

- Scandinavian approach: practised in Sweden, Finland and Denmark. It is characterised by: low variable charges based on short run marginal cost; Infrastructure charges are estimated by comparisons with other modes of transportation; governments contribute the difference between incomes and infrastructure costs.
- Adjusted average cost: practised with some variations in Germany, France and Austria. Targeted revenue through adjusted variable costs (substantially more than

[1] These are now included in franchise specifications by Department of Transport.

short run marginal costs) is raised depending on the level of government contributions.
- Great Britain approach: Very high fixed costs but variable costs at or below short run marginal costs.

Affuso (2003) studies the mechanism for allocating railway capacity through auctioning train paths and reviews its pros and cons. She states the advantages of auctioning railway capacity as: identifying where improvements in capacity utilisation are needed by receiving signals from the market, revealing the true value of railway capacity by enough bidders and repeated auctioning, providing incentives for the operators to improve efficiency through repeated auctioning and better match of rail supply and market demand. On the other hand, the negative sides of auctioning train paths are expressed as: returns of the auction being only a fraction of the costly investments needed for capacity utilisation and the need for combinatorial systems to enable the complex auctioning process. Public regulators are needed to make the monopolist track operator invest the high scarcity rents of bottlenecks in resolving the bottlenecks. She stresses that harmonising the economic regulation of infrastructure and operation is poorly researched.

Nash et al. (2004) address the track access charges in Great Britain and emphasise the need to include the opportunity cost of scare track capacity. They calculate this as the sum of additional traffic attracted to rail multiplied by the paid price, consumer surplus due to additional quality and external cost savings. They conclude that considering this opportunity cost is a complex issue but provides a good incentive for operators to improve the quality of their services.

Nash and Johnson (2005) describe three issues for railway capacity based on their case study of the East Coast Main Line:

- Physical characteristics of the infrastructure are interlinked with each other; overcoming one constraint may activate another constraint.
- Although passengers may value regular interval timetables more (Wardman et al., 2004), satisfying commercial requirements of passenger services (such as periodic timetables) can prevent the most efficient capacity utilisation.
- Complex relationships between different train services might lead to some paradoxical consequences: increasing one type of services enables an increase in the frequency of another type of train and hence better capacity utilisation.

In continuation of their previous work, Johnson and Nash (2008) emphasise the importance of proper charges for scarce track capacity by considering the opportunity costs. They conclude that operators do not have enough incentives to make efficient use of capacity after they have been awarded the time slot. Similar observations have been stated by Smith (2009) in the South of England.

According to economic theories and as advised by European Commission, access charges to the rail infrastructure should equal short run marginal costs[1] (Nilsson, 2002, Hylen, 1998). Marginal costs of adding one more train can include additional costs to the infrastructure manager, external costs, disruptions and delay costs to other trains and opportunity costs of trains that could have run instead (Hylen, 1998). According to his work, if:

- Access charge < marginal cost: Some operators that offer less valuable and efficient services are allowed to use the track capacity.
- Access charge = marginal cost: optimum access charge [2]
- Access charge > marginal cost: Some operators are priced off the network even though they might be able to pay more than marginal costs.

As discussed for definitions of capacity in section 2.1, railway track capacity is used in discrete steps. For effective capacity utilisation, the capacity must be used effectively at the levels of both quantity and quality.

3.7.3 White papers, policy papers and strategy documents

Major White papers that affect the government's policy on railway capacity utilisation are provided below for the case of Great Britain as an example of a vertically separated railway. It will be followed by some policies from the United States as an example of vertically integrated railway. Capacity challenge in Great Britain is being met nationally but in the US it is met by individual class I railroads.

3.7.3.1 Great Britain (vertically segmented railway)

White papers are published by the Department for Transport, usually before legislation. The Office of Rail Regulation and Network Rail publish policy papers and strategy documents all of which affect capacity utilisation. These documents are briefly reviewed. The overall structure of the passenger rail industry in England and Wales is shown in Figure 3-27. The basic structure was established by the 1992 white paper and the 1993 Railways Act. It was modified by the 2004 white paper and the 2005 Railways Act. All the interactions shown in the structure affect capacity utilisation at macro and micro levels.

1 Marginal cost is the change in total cost when the quantity produced changes by one unit.

2 It should be noted that just charging marginal cost may not raise enough revenue to finance capacity enhancement renewals.

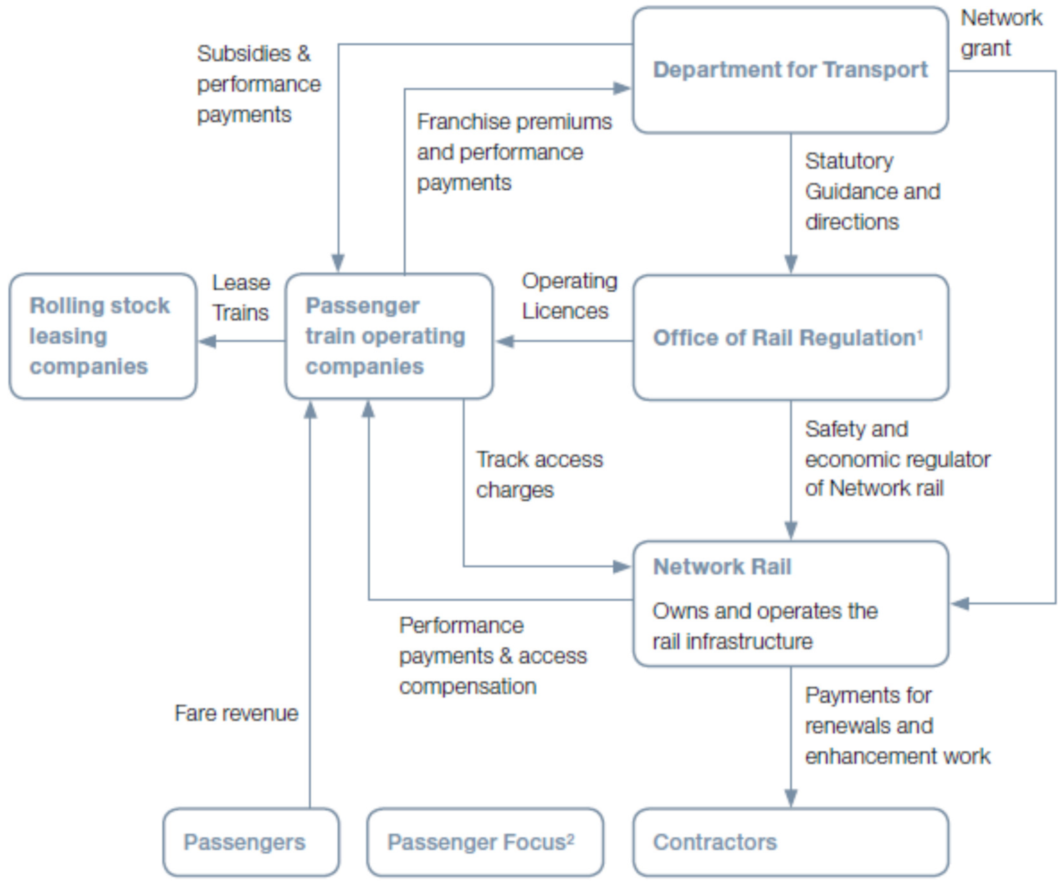


Figure 3-27 - Structure of passenger rail industry in England and Wales (Department for Transport and the Office of Rail Regulation, 2010)

The Department for Transport reviewed the rail industry in 2004. “The Future of Rail White Paper” reached conclusions regarding the structural and organisational changes needed to increase safety, control costs and improve performance. Under its proposals, the Government is responsible for setting the strategy, Network Rail operates the network and is responsible for performance, and the Office of Rail Regulation regulates safety, performance and costs. (Department for Transport, 2004)

The Eddington Transport study is Sir Rod Eddington’s advice to Government published in 2006. It includes volumes on understanding how transport can contribute to economic success, defining the challenge and identifying strategic priorities for the UK transport, meeting the challenge and prioritising the most effective policies and taking action and enabling the system to deliver (Eddington, 2006). However, it was sceptical of “grand projects” such as Maglev.

A strategic modelling tool was presented in “The Network Modelling Framework” in 2007 to support the testing of railway schemes by the Department for Transport (DFT), the Office of Railway Regulation (ORR) and Transport Scotland. It is divided into modules. The demand module investigates how changes in exogenous factors, fares and timetable affect demand. It models crowding and impact of performance on demand as well as providing methods for calculating passenger revenue. The Train Operating Company (TOC) operating cost module estimates change in operating costs by taking into account various physical and financial parameters like staff, rolling stock, Network Rail costs, station costs, etc. The Performance module predicts Average Minutes Lateness (AML) and Public Performance Measure (PPM) based on use of the Capacity Utilisation Index. (Department for Transport, 2007b) A general overview of the Network Modeling Framework is presented in Figure 3-28.

The growth and development of railways in Britain is discussed in the "Delivering a sustainable railway" White Paper by the Department for Transport. It sets long-term strategies in safety and security, reliability, tackling capacity challenge, providing urban, regional and international services, improving environmental performance, reducing costs, etc. 'Capacity challenge' is discussed in a separate chapter covering demand forecasting, load factor and setting out the government's high level output specification for 2013/14. The main approach suggested is lengthening trains and platforms. (Department for Transport, 2007a)

Three key issues are covered by High level Output Statement (HLOS): reliability, safety and capacity. The first HLOS sets the strategic output that Government wants the railways to deliver during control period 4 (April 2009- March 2014). SoFA declares the public funds available. Capacity issues include an increase in the volume of demand to be accommodated and the maximum acceptable level of crowding for planning purposes. (Department for Transport, 2008)

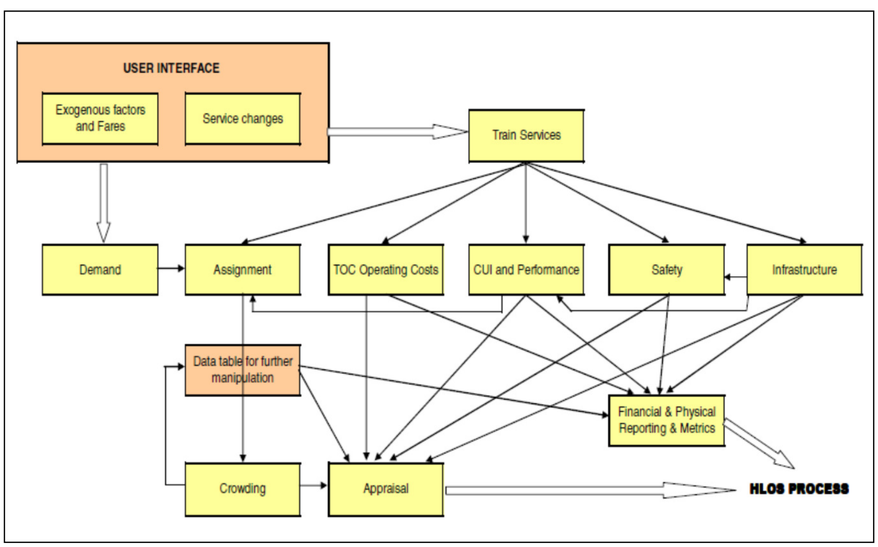


Figure 3-28- High level model structure of Network Modelling Framework (Department for Transport, 2007b)

Network Rail develops "Route Utilisation Strategies" to cover the rail network, in conjunction with rail industry partners and wider stakeholders. Route Utilisation

Strategies (RUSs) try to balance capacity, passenger and freight demand, operational performance and cost, and to address the requirements of funders and stakeholders. Twelve Route Utilisation Strategies have been published so far. (Figure 3-29)

Figure 3-29-UK network as divided by Route Utilisation Strategies (Network Rail, 2009b)

Rail value for money, the groundbreaking analysis undertaken by Sir Roy McNulty (Department for Transport and Office of Rail Regulation, 2011), is based on over 2600 pages of consultancy reports commissioned on 7 different aspects of railways: industry objectives, strategy and outputs; leadership, planning and decision-making; structures, interfaces and incentives; revenue; asset management and supply chain; safety standards and innovations and people.

The main barriers to efficiency and value for money in Great Britain railways are identified. The role of government as the main entity responsible for costs has caused industry not to take responsibility for cutting the costs down. Incentives are either being ineffective or misaligned as the whole-system approach has been neglected. Fragmentation of structure in the railway industry contributes to inefficacies and high costs. Hence, train-operating companies focus on very short-term goals and Network Rail working in a heavily centralised manner. Franchise periods are short and franchises overly-prescriptive. The fare structure is also extremely complex and not based on efficient pricing. (Department for Transport and Office of Rail Regulation, 2011)

Recommendations are proposed in three categories of creating an enabling environment, delivering greater efficiencies and driving implementations. To create an enabling environment it is suggested that government focuses primarily on setting overall goals and objectives for the industry and leading franchising procurement while industry accepts a higher level of responsibility for delivering the goals and cutting costs. Devolved decision-making is recommended by decentralisation and reform of franchises allowing more freedom to train operating companies. Closer alignment of incentives are needed for train operating companies and Network Rail. In terms of regulation, it is suggested that the rail industry is regulated by Office of Rail Regulation while the Department for Transport does a review of fares to better manage demand and capacity utilisation. To deliver greater efficiencies, the report suggests improving asset management, project management and the supply chain. Improving safety culture, establishing a Rail Systems Agency (RSA) to drive innovation, reviewing staffing and human resource practises, better cross-industry information systems and more effective procurement of rolling stock are among other suggestions. It is also stressed that regional railways need to lower their costs as regional franchises constitutes 61% of total net franchise costs. To these recommendations, an independent team for change management is suggested to take the lead for reporting the progress. (Department for Transport and Office of Rail Regulation, 2011)

3.7.3.2 United States (vertically integrated freight railway)

In the United States, railway infrastructure is owned by the freight operators so they have some level of autonomy on how to use railway capacity. There are also some reports and reviews about current and future trends in national capacity utilisation.

The American Public Transportation Association (2006) discusses in the “US rail capacity shortage” study the ‘US rail capacity crunch’ and emphasises the importance of railways for the US in the global economy. It discusses how tight capacity has affected commuter railways and emphasises the importance of rail corridors in urban areas.

“National rail freight infrastructure capacity and investment study” was conducted by Cambridge Systematics (2007) and studies the concept of track capacity utilisation for Class 1 railways in the United States. It analyses current capacity consumption and level of service on the main freight corridors as well as projected capacity utilisation for 2035. Based on these analyses, the investments required in infrastructure expansions and improvements are estimated to be $148 billion (in 2007 dollars). Without improvements, 25 percent of primary corridors would operate at or near theoretical capacity and 30 percent above practical capacity.

“The State of U.S. Railroads: A Review of Capacity and Performance Data” by RAND Corporation (2008) investigates the concerns about the ability of the US railroad system to absorb predicted huge increases in freight demand without degrading the level of service[1]. It studies the concept of capacity in a very broad spectrum and its interactions with industry structure, infrastructure, motive power, operating strategies and crews.

---

[1] Railway freight traffic and revenue in the US has constantly been increasing over the past few years with the exception of year 2009 compared to year 2008 which was due to the economic downturn.

3.8 Summary and conclusions

Railway capacity utilisation is a multidisciplinary area. A chain of various factors stemming from timetable, singling, nodal capacity constraints, rolling stock, infrastructure, external factors and governance affect capacity utilisation. Timetable links rolling stock, infrastructure and signalling together. Much research has been focused on studying and improving different aspects of timetable. Signalling ensures the safe running of trains on the infrastructure. Advanced signalling systems reduce headway and blocking time of infrastructure. Stations and junctions are usually the bottlenecks of the railway network. The routing of trains at stations, dwell time, the layout of crossings, the number of stations and the length and number of platforms all affect capacity utilisation at stations.

Rolling stock characteristics such as speed, acceleration and deceleration affect blocking time of the infrastructure. Improved door mechanism can decrease dwell time at stations and the possibility of running double deck trains increases the number of passengers that can be transported in the unit of time. Infrastructure characteristics such as the number of tracks, line speed, axle load and loading gauge affect train operations. Constant maintenance and engineering works which are needed to keep the steel tracks and substructure in good condition can interfere with train operations. External factors such as falling leaves or extreme weather conditions can interrupt train services and railway authorities usually take precautionary measures to manage their consequences. The way railway is governed including its structure, access charge and white papers affect capacity utilisation at strategic or tactical levels.

Taking into account the wide range and multidisciplinary nature of factors affecting capacity utilisation, it is concluded that efficient capacity management needs a holistic and systems approach. This will be adopted and developed in the following chapters.

4 Holistic approach to railway capacity utilisation

As can be seen from the factors affecting capacity utilisation in chapter 3, railway capacity utilisation is the outcome of complicated interactions of various parameters. Therefore it is crucial to have a holistic approach toward railway capacity, see the big picture and consider railway as a system. This has been neglected in the literature to a great extent. In this chapter railway capacity is studied by adopting a systems approach which results in a new definition for used capacity.

4.1 Introduction to systems thinking

The word system is used in daily conversations for referring to organised wholes in the fields of for example transportation, computer science, medicine, sociology and communication. In scientific terms, one of the early definitions for system was formulated by Fredrich Hegel (1770-1831). He summarises the main characteristics of systems as:

- The whole is greater than the sum of parts
- The whole defines the nature of the parts
- The parts cannot be understood by only studying the whole
- The parts are dynamically interrelated or interdependent

(Skyttner, 2001)

One frequently cited definition of the system is “a set of objects together with relationships between the objects and between their attributes” (Hall and Fagen, 1969). Bertalanffy in his ‘General System Theory’ book defines a system as “a whole that consists of interconnected parts” (Bertalanffy, 1968). The International Council on Systems Engineering (INCOSE) defines system as “an integrated set of elements that accomplish a defined objective” (International Council on Systems Engineering, 2004). According to Blanchard (Blanchard, 1991), a system:

- is contained within some form of hierarchy
- is usually influenced by the performance of the higher-level system and the external factors
- may be broken down into subsystems
- must have a purpose and be able to respond to identified need in a cost-effective manner

A system constitutes a complex combination of resources and different entities in the form of human beings, materials, equipment, facilities, data, money, etc (Skyttner, 2001). The main properties of systems according to General Systems Theory can be summarised as:

- Hierarchy: Systems are complex wholes that are usually nested within each other and a system constitutes smaller subsystems.
- Interrelationships: Interactions between elements and subsystems exist and they affect each other.
- Organisation and regulation: Interrelated objects in the system must be organised in an effective manner in order to achieve the goal.
- Holism and synergy: Due to the organisation in the system, the whole is not just the sum of its parts. Holistic properties exist that are not possible to detect by reductionism.
- Boundaries: Boundaries distinguish systems from their environment.
- Goal seeking: There is a purpose, final state or goal to achieve.
- Inputs and outputs: The system is in interaction with its external environment through inputs and outputs
- Transformation process: The existing processes in the system transform some inputs into outputs.
- Feedback: Information about the output is back as input into the system to change the transformation process if necessary to better achieve the goal of the system.
- Entropy: There is some disorder or randomness in any system. If order is not maintained, the entropy of the system increases.
- Differentiation: In a complex system, each subsystem performs specialised functions.
- Equifinality and multifinality: In an open system[1] the same objective can be achieved from different initial states (equifinality) or different objectives might be attained from the same initial state (multifinality).

(Skyttner, 2001, Litterer, 1969, Bertalanffy, 1968, Kast and Rosenzweig, 1972)

There are two main approaches in system thinking: Hard System Thinking (HST) and Soft System Thinking (SST). Systems engineering and (hard) operations research fall into Hard System Thinking, which is more suitable where there are well defined systems to engineer. Soft System Thinking is more suitable for the fields that involve complex human activities, business, sociology, etc. where there is ‘complexity’, ‘confusion’ and no apparent system; thus the process of inquiry should be systematic (Yan and Yan, Checkland, 1999). For railway transportation, a mixture of Hard and Soft System Thinking is appropriate as it entails distinct engineering subsystems and is also a socio-technical system. Operations research has been long used in railway transportation for optimising and solving different railway problems like train (re)scheduling, train routing, crew scheduling, train formation, etc. However, system engineering has been neglected.

Systems thinking “helps to see patterns in the world and spot the leverage points that, when acted upon, lead to lasting beneficial changes” (Haines, 2000). In order to shift into better patterns and tackle poor results it is necessary to see the relationship between

---

[1] An open system is “a system that is dependent upon environment with which it can exchange matter, energy and information” SKYTTNER, L. 2001. *General systems theory : ideas & applications,* Singapore; River Edge, N.J., World Scientific.

structures and how the system works. In the rapidly changing and complex systems of today, "it is systems thinking that enables us to identify root causes of problems, manage, adapt and discover new opportunities" (Meadows and Wright, 2008). Systems engineering is not just a tool but a paradigm of thinking that provides better grounds for railway practitioners to grasp the goal of the system, underlying processes and complex interactions involved. This is crucial for enhancing railway capacity utilisation. It is important to realise that having a systems approach will not solve the problem itself. However, it "does reframe how we think about what we view as a problem in the first place, and what solutions might look like" (Cabrera et al., 2008).

System engineering can provide robust means for better managing transportation systems, and researchers in recent years have paid attention to it. Larsson et al. (2010) have successfully applied system engineering to the concept of road safety to improve existing approaches. Bojovic (2002) applied general system theory to the problem of railway car fleet size to minimise total costs while satisfying the demand. Wang (2008) has used the theory of system engineering for developing integrated multi-modal transportation.

4.2 Railway as a system

Following this general introduction to systems theory and thinking, we can describe the railway as a system with its own terminology. This helps us in developing a complete definition for used capacity and better managing it. For a typical European railway which is vertically separated and is passenger-focused the concepts of systems theory are discussed in sections 4.2.1 to 4.2.6.

4.2.1 Stakeholder of the railway system

A stakeholder "is anyone or an organisation having a vested interest in a system and its outcomes" (Wasson, 2006). In most European railways where the infrastructure is a public asset, the general public is the stakeholder of the overall railway system but the government represents the interests of the general public. It should be noted that different subsystems of the railway have their own stakeholders. In the case of track capacity, the stakeholder is not the infrastructure authority but the government, which applies vital safety and cost controls when public funds are used. As for passenger services, both infrastructure authority and private operator receive funds from government, the system might reach a state that is not optimum and public funds are not efficiently used. Such a state might be acceptable for the infrastructure authority as well as the private operator but would not be efficient in a larger context. An example of this was described by Smith (2009) for some train services in the south of England:

- Overloading at peak hours (125- 150 percent)
- Overall load factor: 25 percent

- A lot of empty seats being hauled around off-peak[1]
- Hauling empty seats long distances to satisfy short distance demand (eg. South West Trains from Weymouth and Exeter to meet the Woking demand)
- Trains carrying few passengers around the fringes of the country while there is overcrowding in the central parts of the network[2].

He states that "timetabling is the culprit in the empty seats problem". However, by a systems approach it can be discussed that this is not the real cause, but a symptom. The real culprit is an inaccurate approach toward used railway capacity, its definition and its stakeholder. For both the infrastructure authority and the passenger operator such a situation works: the bid has been successful, the infrastructure authority has received access charge while total subsidies, performance payments and fares outweigh the costs of the passenger operator. However, a nearly empty train is not an effective use of railway infrastructure as a public asset. For example instead of a nearly empty passenger train, track capacity could be allocated to a freight train that generates more revenue and also eliminates many trucks from the congested roads. The costs of inefficiencies are ultimately met by the government and the passengers who pay higher fares. This is also due to a macro approach toward railway capacity (e.g. the UIC 406 and CUI approaches) that just measures macro capacity utilisation: whether a train is empty or fully loaded, makes no difference for the capacity utilisation index. The relative value of the train is not reflected in macro approaches. Therefore, not only it is critical to identify the right stakeholder for railway capacity utilisation but also a combination of macro and micro approaches should be used to measure capacity utilisation for the stakeholder. To tackle the above-mentioned issues and increase efficiency, the recent value for money study suggests merging stakeholders by "vertical integration though a concession of infrastructure management and train operations combined" (Department for Transport and Office of Rail Regulation, 2011).

4.2.2 Goal of the railway system

The goal of a system, as summarised by McMullen (1998), is determined by its owners or stakeholders and should be measureable. The goal of the railway system can be defined as providing a sustainable mode of transportation, with an acceptable level of service and safety in a cost-effective manner. Currently there is no single measure to quantify the goal of the railway system. Therefore different measures exist for its sustainability, level of service, safety and cost effectiveness. This is an issue for railway capacity as it encompasses them all. It will be discussed further in section 4.3.1 - Lack of a holistic measure to analyse efficient capacity utilisation.

1 Some of these are positioning movements

2 70% of all rail users either start or finish their journey in London. NETWORK RAIL. 2010c. *General facts* [Online]. London. Available: http://www.networkrail.co.uk/ [Accessed 13/09/2010.

4.2.3 Hierarchy

The railway system is nested in different levels of hierarchy that are nested. At each level there are different regulators and regulations. For example in the case of Great Britain, the national regulators are the Office of Rail Regulation and the Department for Transport. At the international level it is the European Union and the International Union of Railways (UIC). Laws from the higher level dominate lower level ones.

4.2.4 Boundary

Defining the boundary of a system is critical. A system boundary is a physical or conceptual separation of the system from its environment while encompassing all the essential elements and subsystems to address the decision problem (Parnell et al., 2008). The decision problem affects what is outside the boundary and what is inside. As our decision problem is railway capacity challenge, the boundary of the system is the whole railway infrastructure (tracks, stations, junctions, etc.). The railway infrastructure in the UK is maintained by public funds and allocated to private passenger and freight operators. Due to this complex combination, choosing the right boundary for analysis of capacity utilisation is very important.

4.2.5 Inputs, processes and outputs

A system receives inputs, processes them under control mechanisms, and generates outputs. There are various inputs for the capacity problem. Government pays subsidies and performance payments so that passenger operators can run their services on the railway infrastructure (Usually passenger services are not profitable so they need support from government). Through the infrastructure authority, the government also invests to maintain and expand the infrastructure, and manage its utilisation. By means of bids from operators, track capacity is allocated and appropriate access charge is received.

The railway tracks needs to be managed, maintained, built and allocation of time slots for using it to be . Through these processes, the inputs can be turned into outputs. The outputs of the railway system are the different passenger and freight services offered. These outputs are in the form of seats/seat kilometres, passengers/ passenger kilometres, tonnes/ tonnes kilometres, etc.

4.2.6 Control mechanism

The quality of the outputs is monitored through control measures. In Great Britain, the safety and cost controls are carried out by the Office of Rail Regulation (ORR) and the Department for Transport. Performance control is the responsibility of Network Rail in conjunction with the Department for Transport. The major performance control is Public Performance Measure (as defined in section 3.6). Other performance control mechanisms include regulations of crowding, fare regulation and franchise specifications.

Figure 4-1 summarises inputs, processes, outputs and control mechanisms of railway capacity utilisation and the box in the middle represents processes. Financial inputs include subsidies, investments, bids from operators, performance payments and access charges. Main physical inputs for track capacity utilisation include network kilometres, rolling stock fleet size, stations, platforms and goods yards. Regulatory controls cover safety, cost and performance aspects. Freight services and passenger services can be measured by train kilometres, passenger kilometres tonne kilometres, total passengers, total freight tonnes etc.

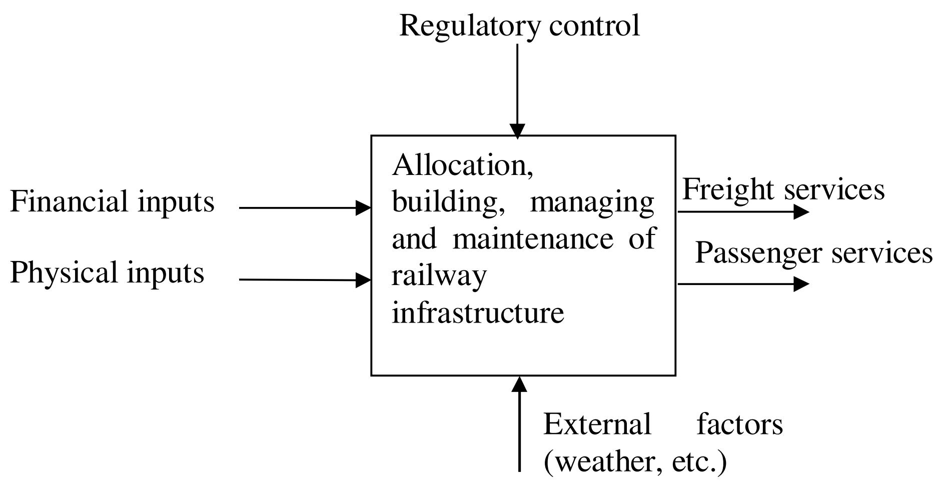


Figure 4-1 - Railway infrastructure as a system

4.3 Need for a systems approach to railway capacity

There are several issues that underline the importance of considering railway as a system and having a systems approach toward railway capacity. They are discussed in the following sections (4.3.1 to 4.3.8).

4.3.1 Lack of a holistic measure to analyse efficient capacity utilisation

There are various metrics that quantify different aspects of capacity utilisation. However, they are "index-numbers" and each of them considers just one aspect of capacity utilisation. As the co-winner of the Economics Nobel prize in 1969 has put it:

"The index-number problem arises whenever we want a quantitative expression for a complex that is made up of individual measurements for which no common physical unit exists. The desire to unite such measurements and the fact that this cannot be done by using physical or technical principles of comparisons only, constitutes the essence of the index-number problem and all the difficulties centre here." (Frisch, 1936)

Figure 4-2 schematically illustrates some different metrics related to capacity utilisation and their non-aggregated nature.

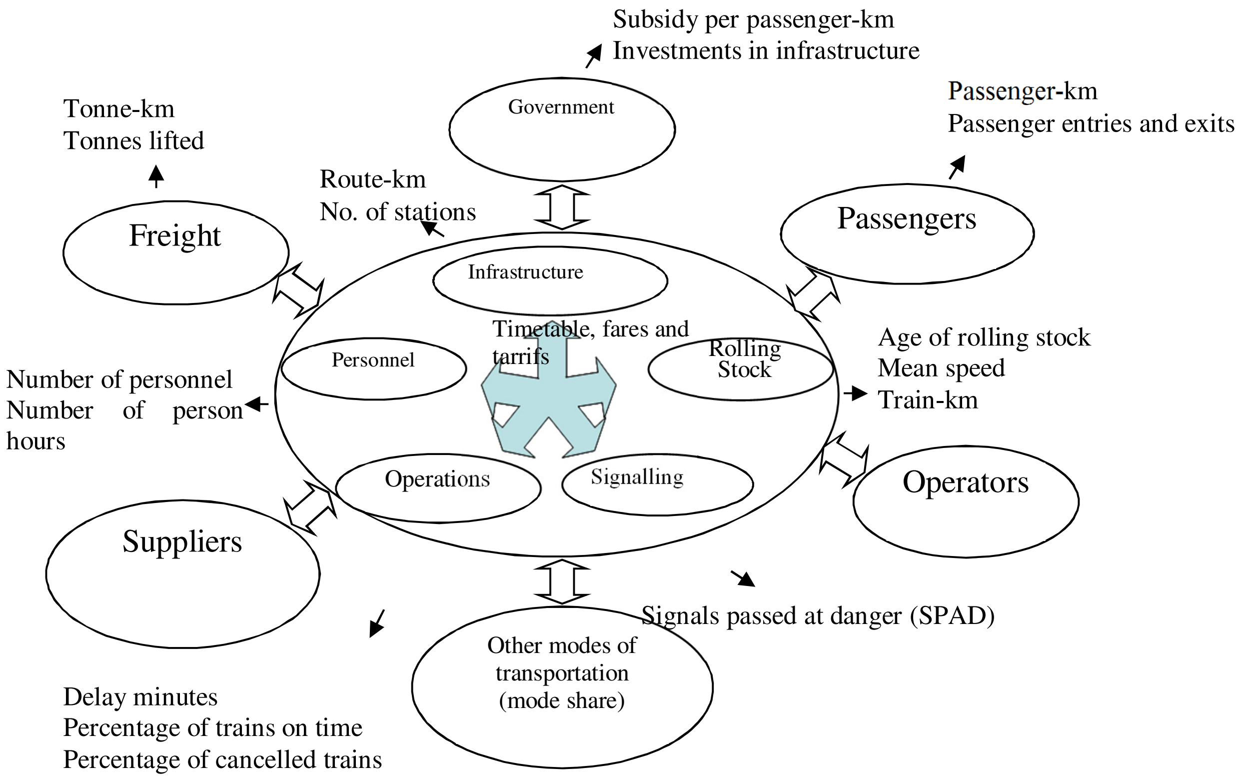


Figure 4-2 - Non-aggregated metrics in railway sub-systems based on (Khadem Sameni et al., 2010a)

Dingler (2010) categorises metrics of capacity into throughput, level of service and asset utilisation. Based on his categories, Table 4-1 summarises the strengths and weaknesses of these metrics.

Table 4-1 - Analysing metrics of capacity utilisation (Khadem Sameni et al., 2011b). Based on Dingler (2010) and (Khadem Sameni et al., 2011a)

Category		Metric	Description	Strengths	Weaknesses
Throughput	Macro	Number of trains, train-km	How many passengers can be transported over a period of time	Easily measurable and understandable	Does not reflect quality of service
	Micro	Number of passengers, passenger-km, seat-km			

Level of service	Average delay, percentage of cancelled or late trains (e.g. Public Performance Measure in Great Britain)	Measures reliability and timeliness	Important for general public	Indirect measure heavily depends on how saturated the network is. Does not take into account scheduled waiting time and timetable supplements which are a waste of time for passengers
Macro Asset utilisation	Capacity Utilisation Index (CUI), Total time utilisation of infrastructure (UIC 406 method), Number of trains per km of infrastructure in a given time period	Estimating how saturated the network is	Important to estimate how efficiently the infrastructure is utilised	A measure of macro capacity utilisation, does not reflect the actual value of trains, load factor and how close the passengers are standing (micro capacity utilisation)
Micro asset utilisation	Train load factor	Estimating how crowded the passenger trains are	Important to estimate how efficiently the rolling stock is utilised and the level of comfort for passengers	A measure of micro capacity utilisation, does not reflect how saturated the network is (macro capacity utilisation)

As Table 4-1 suggests, each of these metrics is suitable to measure one aspect of capacity utilisaiton. Therefore one of the main aims of the present thesis is to develop more holistic measures of analysing capacity utilisation hence addressing research question of "How to measure capacity utilisation from a systems point of view?"

4.3.2 Segmentation after privatisation

In a typical vertically segmented railway, such as the post-privatisation railways in Europe, objectives, interests and concerns are segmented as well. The government, passenger operators, freight operators and infrastructure authority have different responsibilities, objectives and concerns. Figure 4-3 shows a 'rich picture'[1] of railway capacity utilisation based on Figure 4-1. As different players in the capacity utilisation have different goals, inefficiencies in capacity utilisation may occur. This point is one of the main conclusions of the value for money report in the Great Britain Railways: "Fragmentation by which is meant the fact that the structures within an industry which has many players, and the interfaces between those players, have not worked well in terms of securing co-operative effort at operational interfaces or active engagement in cross-industry activities which need to be undertaken for the common good. One of the principal barriers, if not the principal barrier, is the lack of an effective supply chain that starts with the customer (passenger and freight) and taxpayer, and focuses the efforts of all concerned on meeting these needs in a cost-effective manner." (Department for Transport and Office of Rail Regulation, 2011)

---

[1] Rich picture is a system tool introduced by Peter Checkland, developer of Soft Systems Methodology, to graphically represent a complicated situation, relationships and concerns CHECKLAND, P. 1999. *Soft Systems Methodology: A 30-year Retrospective,* Chichester, Wiley.

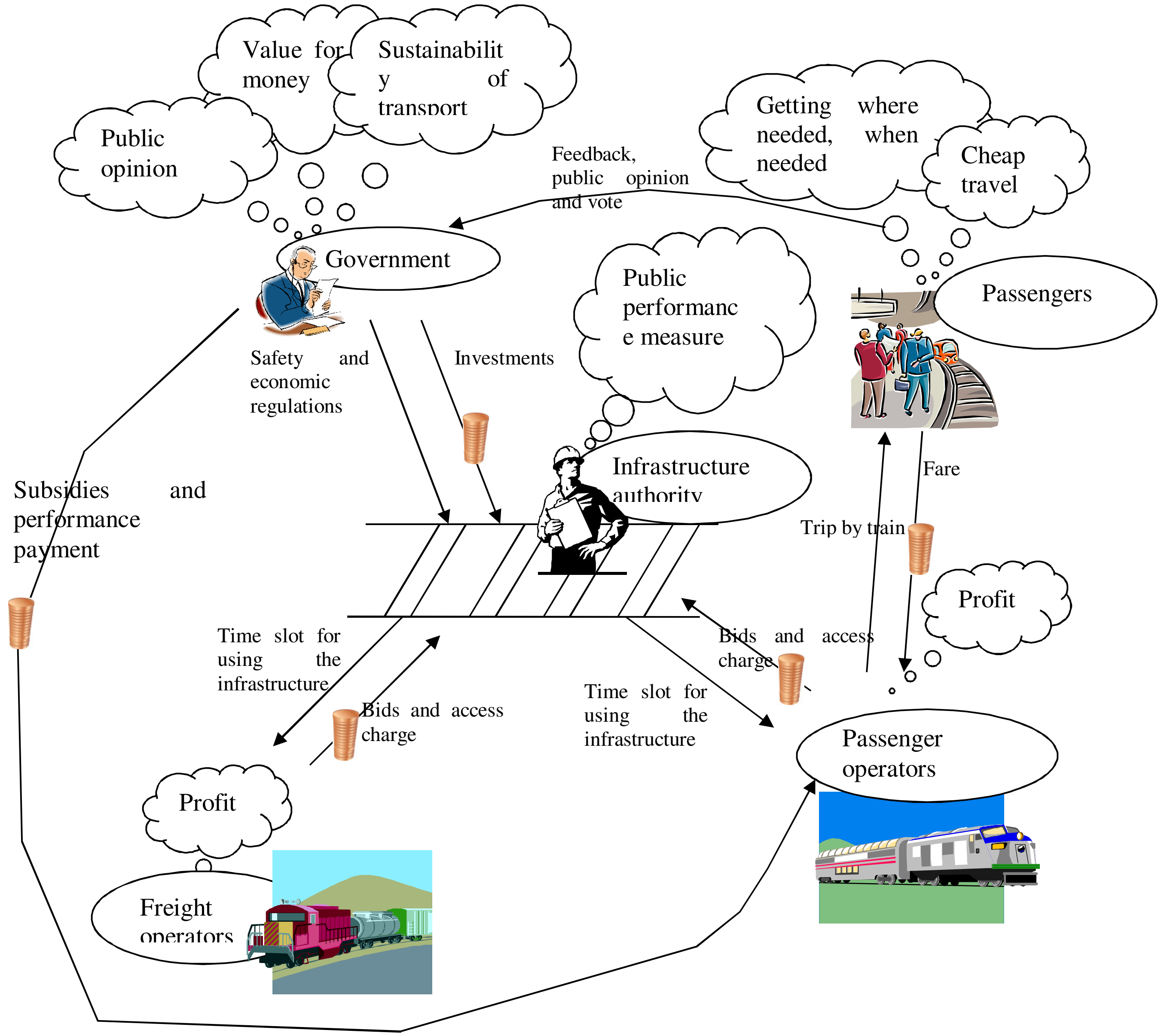


Figure 4-3 - Rich picture of railway capacity utilisation

4.3.3 Reducing complexity

Different systems have different levels of complexity. The complexity of the system is mainly dependent on relationships rather than on its constituent parts (Manson, 2001). These relationships can be between the parts as well as with the environment. The main characteristics of complex systems are:

- Large number of elements
- Many interactions between elements
- Attributes of elements are not predetermined
- Interactions between elements are loosely organised
- They are probabilistic in their behaviour
- The system evolves/deteriorates over time
- Subsystems are purposeful and generate their own goals
- The system is subject to behavioural influences
- The system is largely open to the environment

(Flood and Jackson, 1991, Skyttner, 2001)

No matter how complex the system is, in order to manage and optimise it efficiently, it is essential not to be lost in its complexity. Complexity in the system should be tamed by simplicity in thinking (Haines, 2000). Simplicity in thinking can happen "at the near side of complexity" by lack of knowledge and ignoring important details, or it can happen at "the far side of complexity" by supreme knowledge over the system, going to a higher level and efficiently simplifying it (Haines, 2000). This concept is depicted in Figure 4-4. These are closely related to different levels of public transportation planning as defined by Van de Velde (1999) and discussed in section 1.5.

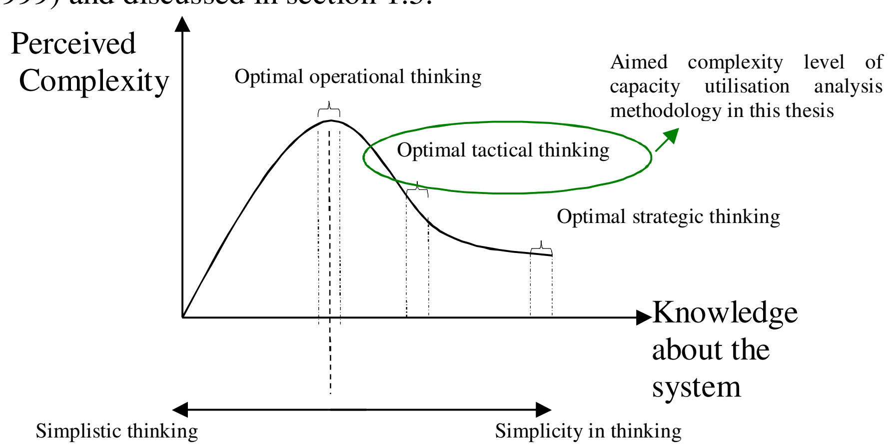


Figure 4-4 - Perceived complexity versus knowledge about the system. Based on (Haines, 2000)

The railway literature is replete with complex tools for operational planning of railways. Simplifying the complexity of railway transportation would greatly enhance the chances of more efficient decision-making which is needed for tactical planning and is the aim of this thesis. The purpose of the railway system is to carry passengers and freight reliably, safely, and efficiently and an appropriate tool for tactical planning of railway capacity should be able to accommodate these concerns.

Railways manifest all the characteristics of complex systems listed in 4.3.3 above. The railway system is composed of subsystems - infrastructure, rolling stock, operations, signalling and personnel - each with its own goals and objectives. For major European railways, the quantity of rolling stock, personnel, infrastructure, passengers and freight is immense. Inside the system, there are intricate interrelationships between operation, infrastructure, planning, signalling, rolling stock and personnel which adds to the complexity. There are complex interactions between passengers and freight demand, other modes of transportation, suppliers, governmental policy and railway transportation supply. There is a factor of probability involved for delays, reliability and stability of services. The railway system can deteriorate or evolve over time based on how well it is maintained and managed. The system is subject to the behaviour of passengers and even the strikes of personnel. The railway system is open to its environment: leaves and snow on the rails and other external factors affect railway services.

Boulding (1956) proposes a hierarchy for different levels of complexity. The first level of complexity starts from a static structure or frameworks and relationships and gradually moves towards higher levels of complexity. The second to the ninth level of complexity are: clock works (level 2), cybernetics (level 3), open systems (level 4), genetic-societal (level 5), animal (level 6), human (level 7), social organization (level 8) and transcendental (level 9). The highest level of transcendental complexity is when the structure or its relationships are unknown. Using this concept, Figure 4-5 presents the complexity of the railway capacity. The identified level of complexity shows how predictable the behaviour mechanism of that entity is.

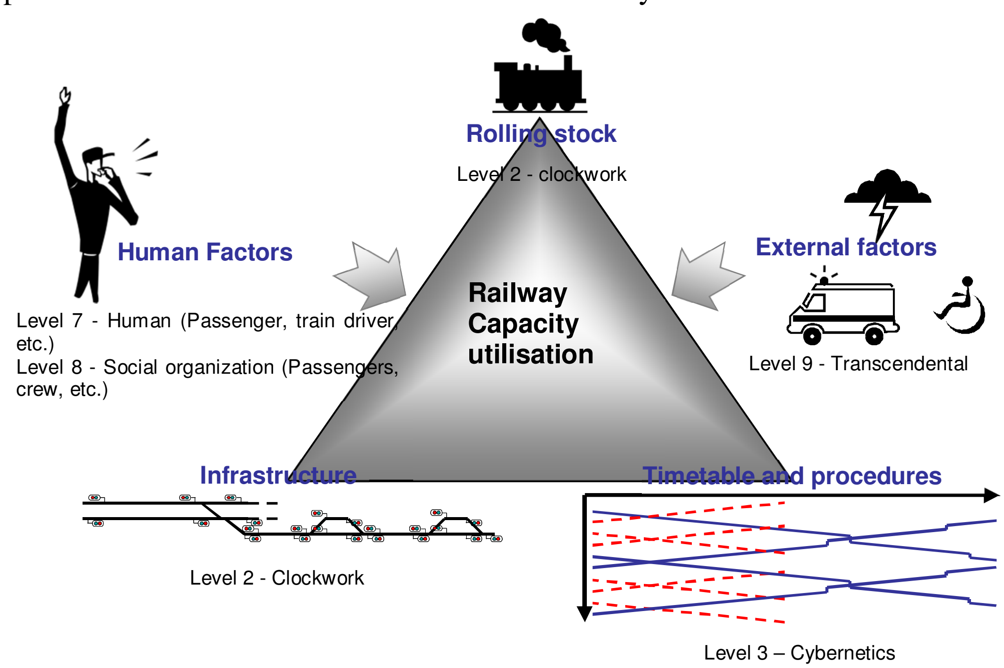


Figure 4-5 – Identifying behaviour complexity of railway subsystems according to Boulding levels of complexity - Based on Landex (2008)

For managing a complex system, the golden rule is that 'control complexity should be equal to the design complexity' otherwise the system or its outputs becomes uncontrollable (Casti, 1986). Existing mechanisms for measuring and managing railway capacity utilisation such as the UIC 406 method (UIC, 2004) and the Capacity Utilisation Index (CUI) (Gibson et al., 2002), while being important initiatives to start work in this field, may be too simple to efficiently control railway capacity utilisation; the control mechanism is simpler than the design complexity of railways. These methods of analysing capacity utilisation (level 2, clockwork) are static. However in railway capacity utilisation everything does not always go like clockwork. Hence these methods of capacity utilisation analysis cannot encompass the stochastic factors that exist in a higher level of complexity and therefore ignore the profound effects of the probability and costs of delay, and the reliability and stability of services on capacity utilisation.

This thesis aims to develop a suitable methodology for analysing capacity utilisation at the tactical level as indicated that encompasses various factors overarching capacity utilisation as indicated in Figure 4-5.

4.3.4 Fragmentation of knowledge in railways

It is not just the way railways are run after privatisation that causes segmentation as described in section 4.3.2. There is an absence of specialised railway engineers who are academically trained to grasp the multidisciplinary nature of the railway system[1]. Instead, civil engineers, electrical engineers, mechanical engineers and economists run different subsystems of railway transportation from their own perspective and field of expertise without adequate knowledge of the other subsystems. They specialise in just one of the subsystems - infrastructure, rolling stock, signalling, operations or regulations – without understanding that these all interact closely for railway capacity utilisation.

A good example of an effective approach to a multidisciplinary field of knowledge is in medicine. The human body is a system where each organ must work in coordination with the others. Doctors are trained firstly in overall general knowledge of the whole system (the human body) before they specialise in one of the subsystems (the organs). The lack of a multidisciplinary approach to the railway education and research causes problems in a topic like railway capacity utilisation. Like medicine, a good general knowledge of various subsystems and a holistic approach is needed for railway capacity utilisation.

The relationship and organisation in the system forms its identity. Railway professionals have limited common language, understanding and interaction: civil engineers don't know about signalling, mechanical engineers cannot figure out the infrastructure concerns of a civil engineer, etc. This is intensified by the way many railways are run which separates different aspects of the railway, mainly infrastructure from operation.

[1] For more information on railway education please refer to MARINOV, M., PACHL, J., LAUTALA, P., MACÁRIO, R., REIS, V. & EDWARDS, J. R. 2011. Policy-Oriented Measures for Tuning and Intensifying Rail Higher Education on both Sides of the Atlantic *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

Meanwhile, they ignore the close link between different aspects of capacity utilisation for example signalling and infrastructure. Infrastructure requirements determine where to locate the signals and the possibility of allowing trains to enter crossing stations from both sides at the same time. This heavily affects capacity utilisation of the infrastructure. Such close interactions between different subsystems of the railway determine capacity utilisation to a great extent.

General systems theory provides good grounds for "the unification of science" (Kast and Rosenzweig, 1972). This common language and holistic approach is very much needed in complex multidisciplinary area of railway capacity utilisation as shown in Figure 4-6 Adopting a systems engineering approach would lead to a multidisciplinary analysis and improvement of railway capacity utilisation.

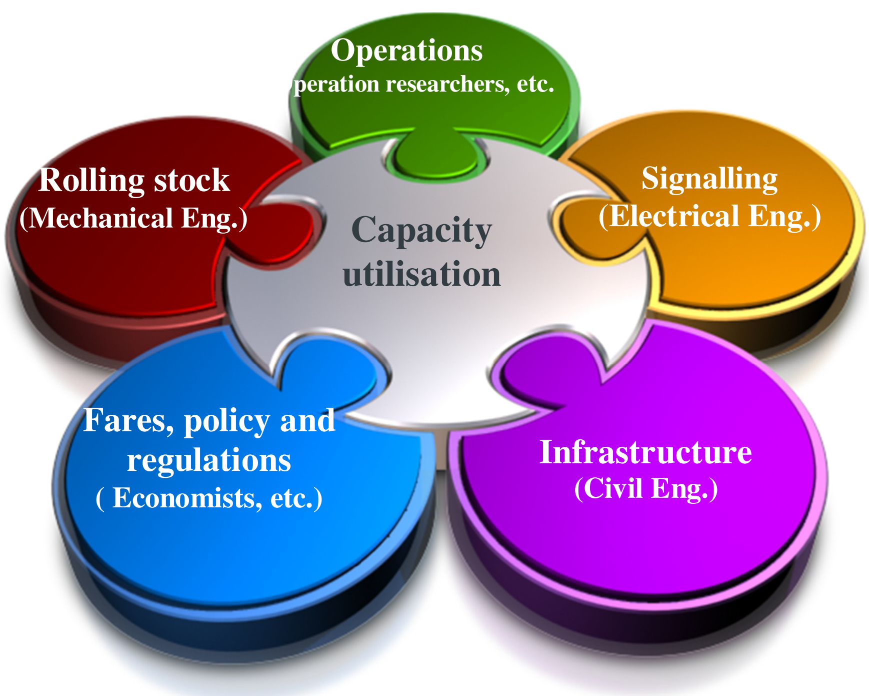


Figure 4-6 - Railway capacity utilisation and fragmentation of knowledge

4.3.5 Weakest link of the chain

To improve any system, three fundamental questions should be constantly asked and answered:

- What to change?
- What to change to?
- How to cause the change?

(McMullen, 1998)

In the railway context these questions should refer to safety and level of service. In the improvement process, finding what to change in the system is the first and foremost step. If 'what to change' is not identified rigorously, much time, effort, and investment may be wasted with no improvement in the system performance at all. Just as the strength of a chain is governed by its weakest link, the overall output of the system is limited by a constraint which is 'anything that limits a system's higher performance relative to its purpose' (Scheinkopf, 1999). The fact is often ignored in the railway industry due to the reductionism approach (breaking down the system into isolated parts) and fragmentation of knowledge. To improve capacity utilisation, the weakest link of the chain should be identified by considering the subsystems holistically. Introducing faster rolling stock with enhanced braking and acceleration will not improve capacity utilisation much if the trains cannot go at their maximum speed on old rails; the constraint is the infrastructure. If rolling stock and infrastructure are in good condition but signalling is the weakest link and the constraint, adding more rolling stock won't help. A system engineering approach would help to identify the constraints of the railway systems more efficiently. This would ensure applying the improvement process as:

- Identifying the system's constraint(s)
- Exploiting the system's constraints(s)
- Subordinating other subsystems
- Elevating the system's constraint(s)
- Returning to step one and avoiding inertia

(Scheinkopf, 1999)

For instance for the case of railway capacity utilisation, the bottleneck of a route (e.g. the most crowded station) should be identified and improved. It is needed that other subsystems interacting with that bottleneck (e.g. the line sections to and from the station) are arranged accordingly. After the constraint is elevated, the analysis should be done again to find the next constraint that is now the weakest link of the chain.

4.3.6 Need for creative and innovative problem solving

Two main tools that have been used in railway research for decades are 'operations research (OR)' and 'simulation'. These powerful tools that have been tremendously successful for improving numerous systems in diverse disciplines have some flaws as well. Although they can relax and modify constraints to some extent, their main aim is finding the best configuration of a system with existing constraints. This inherent characteristic limits their potential to optimising a system rather than improving it. In the long term, the system becomes saturated and improvements would be minimal. Sometimes simple and creative ideas can exploit and change a constraint and improve the system's efficiency a great deal; operations research and simulation cannot accommodate such an approach as they cannot change underlying constraints. However, when new horizons are explored by creative problem-solving, operations research and simulation can help the system find the optimum arrangement. This concept is schematically shown in Figure 4-7 . The radius of circles represents the system constraints within which OR and simulation optimise the system. An innovative solution may increase a system's limitation or decrease it. Innovative problem-solving changes the system's constraints and generates a modified set of constraints within which OR and simulation can optimise the system again.

For example, as previously explained (section 3.3) , nodal capacity constraints greatly affect capacity utilisation. Improving train routing through stations can improve capacity utilisation. A normal OR and simulation approach toward train routing in a busy station takes the existing constraints (e.g. number of platforms, arrival and departure times of trains, dwell times, etc.), models them and optimises routing of trains through stations. An innovative approach tries to change the limiting constraints. For example, an innovative approach to train routing might be allocating one platform to fast trains and another to slower trains. This method of problem-solving changes the system constraints, taps human creativity and provides a robust mean of improvement when used along with OR and simulation.

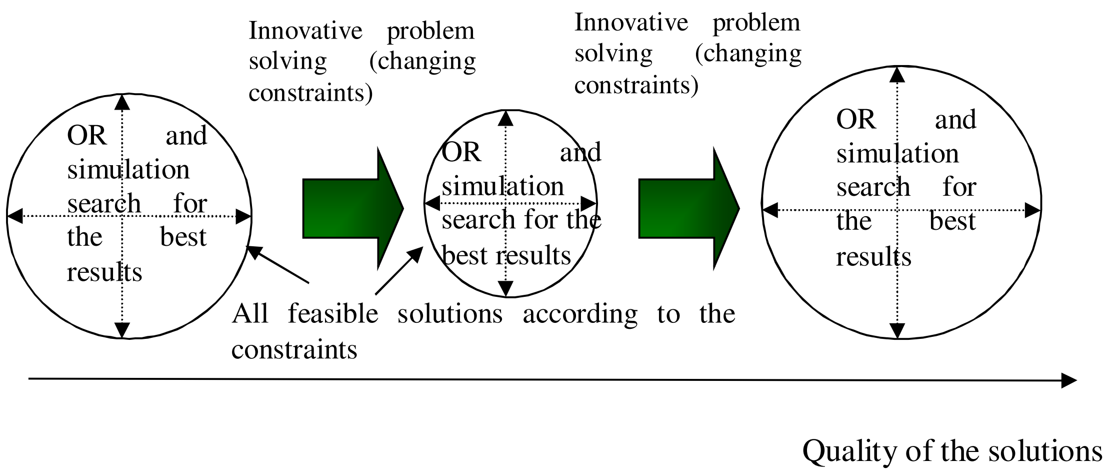


Figure 4-7 - Need for innovative problem solving along with OR and simulation

4.3.7 Holistic decision-making for utilising railway capacity

The railway infrastructure is usually shared by trains with a range of priorities and values. Passenger trains have a variety of stop patterns, destinations, load factor and priorities. Different freight trains carry commodities of different values and time sensitivity. Railways in Europe are mainly passenger-focused whereas in North America freight trains are dominant: in the year 2009, the total tonne-km of railway freight in the United States (2,468,738 million) exceeded the total tonne-km of railway freight in the whole of Europe[1] (2,454,437 million) (UIC, 2011c). The case is very different for the passenger

[1] This includes Turkey and the Russian Federation.

section where the total number of passenger-km in Europe was 624,249 million compared to 9,518 million passenger-km in the United States (UIC, 2011c). Choosing the optimum mix of trains that can maximise the value generated by using the railway infrastructure is a critical issue as different passenger and freight trains generate value in distinctive ways. Time-based capacity utilisation analysis methods can provide limited clues for effective utilisation of capacity as they consider trains as black boxes and ignore the socio-economic value of a train. In mixed traffic, they are biased towards (fast) passenger trains whereas in some cases a freight train can use the capacity of infrastructure more effectively than a passenger train with a very low load factor. More rigorous methodologies for analysing and allocating the scarce and invaluable resource of railway infrastructure are needed due to concerns for CO_2 emissions, global warming, energy crisis and campaigns for sustainable environments.

4.3.8 Tough economic situation: efficiency or deficiency

The existing economic crisis around the world has brought budget cuts to public sectors including the railway infrastructure, which has limited many planned investments to increase theoretical railway capacity. Therefore railways must use the current practical capacity available in the most efficient way possible. Investments must target 'the weakest links of the chain' to bring about best results. As a recent Network Rail report (Network Rail, 2010e) puts it: "With constrained public finances, the taxpayer needs a system that gets the best value for any public money that is spent and one that helps to deliver the highest possible levels of economic return from transport investments". In this regard the report suggests using "real economic returns per pound of net cost to the tax payer" instead of the traditional welfare maximising approach and also prioritising investments, increasing efficiency at lower cost and delivering more for less. None of these are possible unless a holistic systems approach is adopted toward railway capacity utilisation.

4.4 Implication of system laws for railway capacity utilisation

According to General Systems Theory and as previously mentioned (section 4.1), although different systems have different natures, goals and functions, some common themes can be identified among them. These characteristics can be described in the form of laws, theorems and hypotheses in various systems. Some of these laws have been adopted from other disciplines like physics to explain some fundamental system concepts. Table 4-2 summarises some of these laws gathered from different sources by Skyttner (2001) and we investigate them in the context of railway capacity utilisation.

Table 4-2 - System laws for railway capacity utilisation (First two columns are extracted from (Skyttner, 2001)

Laws of System Theory	**Description**	**Implication in railway capacity** utilisation
The second law of thermodynamics	In any closed system, the amount of order can never increase, only decrease over time.	Infrastructure and other subsystems must be maintained in good order. Delays and disturbances violate the order in the railway system and must be controlled.
The law of requisite variety	Control mechanism should be equal to the design complexity	Current methods of railway capacity utilisation analysis are simpler than the complex subsystems that affect capacity utilisation. Methods of analysing capacity utilisation that can accommodate the complexities and probabilities involved must be developed.
System holism principle	The whole is greater than the sum of parts	The railway transportation system has holistic properties over and above those of its subsystems: infrastructure, rolling stock, personnel, etc. For instance, the higher-order network (domino) effect means that total delays are usually more than the sum of primary delays as trains share the infrastructure and if blocking time takes longer than planned, this affects other trains that were not delayed in the first place.
Darkness principle	No system can be known completely	In the railway context, as external factors with highest level of complexity (level 9: transcendental) exist and affect the system, delays etc cannot be known or predicted completely. There is always a factor of probability involved.
Eighty-twenty principle	In any large, complex system, eighty percent of the output is produced by only twenty percent of the system	Nodal bottlenecks of railways (junctions and stations) which are small fraction of the railway network, determine the majority of capacity utilisation. A fraction of stations and junctions determine the capacity constraints hence overall capacity utilisation.

Redundancy of resources principle	Maintaining stability under conditions of disturbance requires redundancy of critical resources	Slack, running time supplement, standard allowance or recovery time is the extra time added to the running time of trains to compensate for the delay of a train. Buffer time is extra time added to the minimum line headway to avoid propagation of small delays. Choosing the right balance between the quality of service and consumption of resources is critical.
Redundancy of potential command principle	Need for different channels of information and feedback to maintain a complex system	Railway control centres take command in the event of disturbances, restore order, monitor the situation, and are free to take action when necessary.
Relaxation time principle	To maintain the system's stability, the relaxation time of the system must be shorter than the mean time between disturbances.	Buffer time, running time supplement, etc. provide enough relaxation time to keep the timetable stable.
Homeostasis principle	System must be maintained to survive with internal and external changes	The railway industry has an historic infrastructure which needs constant maintenance to prevent deterioration over time.
Steady-state principles	All subsystems must be in a state of equilibrium for the system to be in equilibrium and vice versa.	For an efficient and stable use of railway capacity, all the subsystems like rolling stock, infrastructure, signalling, personnel and timetable should be stable.
Viability principle	Autonomy of subsystems and integration with the whole system must be in balance.	Infrastructure owner, operators, etc. have autonomy while proper interrelationships should exist so that the system works effectively and capacity is best utilised
First cybernetic control principle	Implicit control is continuous and behavioural characteristics compared to a standard.	Railways use quantitative standards to measure reliability, stability, capacity utilisation, etc. (eg. Public Performance Measure is used in the UK)
Second cybernetic control principle	Communication is vital for implicit control.	In the event of disturbance, communication is crucial for railway control centres to reschedule the trains efficiently and restore order to the system.

4.5 Choosing a holistic methodology for improving railway capacity utilisation

The previous sections discussed the vital importance of adopting a holistic and systems approach for efficient capacity utilisation. In this section we will choose a holistic methodology that can accommodate the above-mentioned issues and different aspects of efficient capacity utilisation.

4.5.1 Quality improvement methods

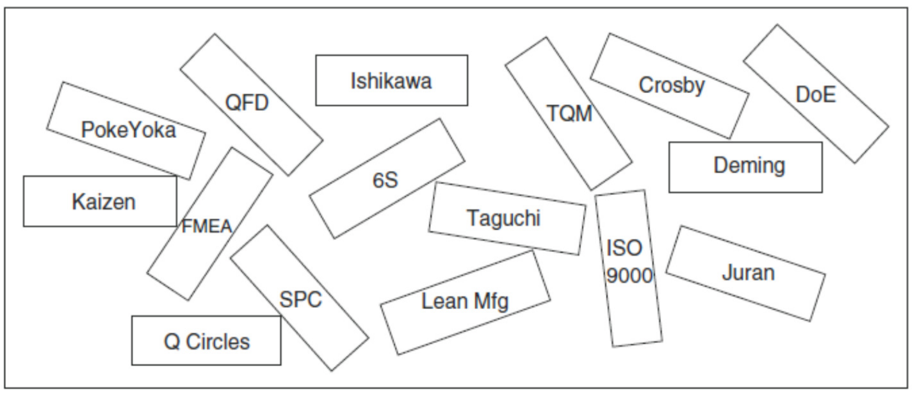
Increasing the quality of products and services, producing the best value for money, increasing efficiency and decreasing costs have always been the primary concerns of industries. Various quality improvement methods have been developed to facilitate reaching these goals which are briefly reviewed in the following paragraphs.

Deming, Juran, Crosby, Taguchi and Ishikawa are some of the best known quality thinkers who laid the foundation of various methods for quality improvements. Deming introduced statistical process control and the concept of quality to Japan and developed a theory for management to improve quality and productivity. Juran's philosophy was based on management commitment to quality improvement. He developed a Quality Trilogy of quality planning, control and improvement. Crosby emphasized on zero defects and developed four absolutes of quality management. Taguchi focused on the importance of reducing variation in quality and introduced a loss function by combining cost, target and variation. Ishikwawa's major contribution was developing cause and effect diagram. (Besterfield et al., 2011)

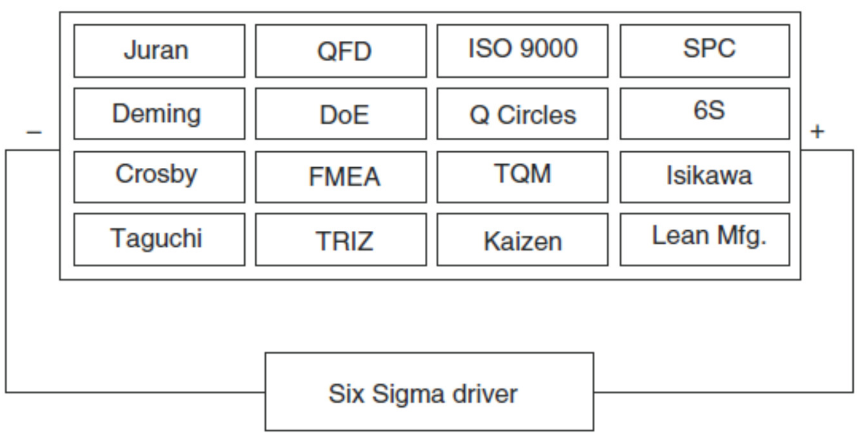
6 S is a method for organizing the work environment which is originated from Japan and is based on sort, stabilize, shine, standardize, sustain and safety (Basu, 2004). Design of experiments (DoE) "is a series of steps which must follow a certain sequence for the experiment to yield an improved understanding of product" (Ross, 1996). FMEA "is a systematic method of identifying and preventing product and process problems before they occur" (MacDermott et al., 2000). ISO 9000 "is a series of quality assurance standards that were created by the International Organization for Standardization (ISO), based in Geneva, Switzerland" (Johnson, 2000). Kaizen "promotes continuous improvement for eliminating waste in machinery, labour or production methods"(Swamidass, 2000). Lean manufacturing "is the production of goods using less of everything compared to mass production: less waste, a less human effort, less manufacturing space, less investment in tools and less engineering time to develop a new product" (Wang, 2010).

PokeYoka is a tool for error prevention and detecting them before they become defects (Seddon, 2005). Quality (Q) circles are "small groups of workers engaged in a continuing cooperative study process to uncover and to solve work-related problems"(Crocker et al., 1984). QFD stands for quality function deployment. Its two main objectives are converting users' needs to quality characteristics at design stage and deploy them to production activities (ReVelle et al., 1998). SPC stands for "statistical process control" and aims for controlling production (Swamidass, 2000). Total Quality Management is a holistic approach by management to improve quality at all levels of the organization (Besterfield et al., 2011). TRIZ is the Theory of Inventive Problem Solving based on 40 general solutions to overcome conflicts (Rantanen and Domb, 2007).

However, as Figure 4-8 suggests, all the above-mentioned methods had a fragmented nature up until the introduction of 'Six Sigma' which organised some of the existing methods towards a common goal (Truscott, 2003). Six Sigma has a more holistic and systems approach compared to other quality improvement methods illustrated in Figure 4-8, therefore we choose it as our underlying methodology.



a) Improvement tools before the introduction of Six Sigma



b) Improvement tools after the introduction of Six Sigma

Figure 4-8 - Six Sigma as an orientating improvement mechanism (Truscott, 2003)

4.5.2 Introduction to Six Sigma

Six Sigma is a process improvement framework which was launched in the Motorola Company in 1987 and helped this company to make huge cost savings while improving quality (Larson, 2003). It has been widely used in various manufacturing and service industries since then. Quality improvements in the service industry are more complicated than the manufacturing companies as services are intangible, perishable, often heterogeneous and are usually simultaneously produced and consumed (Sasser et al., 1978). Although Six Sigma was initially developed in the manufacturing sector, it has been widely used in the service sector as well (Antony, 2006).

The main aim of Six Sigma is to reduce faulty products and services hence increasing their value, reliability and efficiency while decreasing costs. The sigma (σ), standard deviation in statistics, indicates the level of variability. The Six Sigma level of performance is highly stable and expected to meet the required (consumers') expectations, as shown in Figure 4-9 : the quality of service or product very rarely falls outside the acceptable levels.

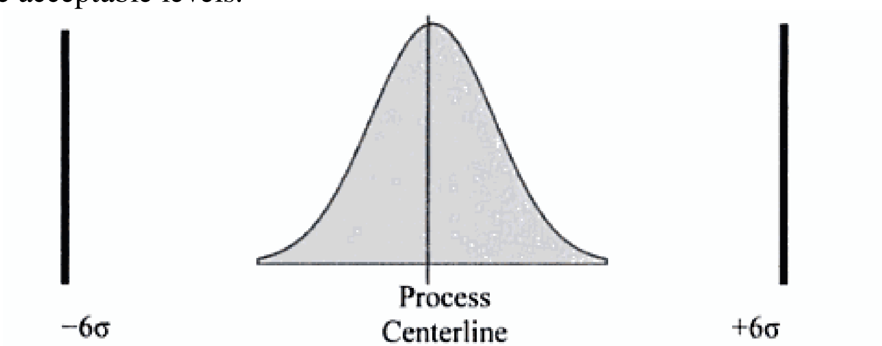


Figure 4-9 - Six Sigma level of performance (Keller, 2011)

The level of service or production in Six Sigma is measured by defects per million opportunities (DPMO) as Table 4-3 suggests. Defect is defined as "any part of a product or service that does not need meet customer specifications or requirements or causes customer dissatisfaction or does not fulfil the functional or physical requirements." (Charantimath, 2011)

Table 4-3 - Sigma levels and defects per million opportunities (George, 2003)

Sigma level	Defects per million opportunities	Yield
6	3.4	99.9997%
5	233	99.977%
4	6210	99.379%
3	66807	93.32%
2	308537	69.2%
1	690000	31%

The Sigma Level of performance is usually calculated by the following formula:

$$DPMO = \frac{Number\ of\ defects\ observed\ in\ the\ sample}{Number\ of\ units\ in\ the\ sample \times opportunities\ per\ unit} \times 1000000$$ (Keller, 2011)

Some measures in the rail industry have tried to address and recognize the importance of reliability. For example, the Golden Spanner annual award in the UK, organized by the Modern Railways (2011) recognizes best practices in rolling stock reliability. It uses mileage between any five minute delays related to rolling stock as its criteria.

Six sigma uses a series of techniques to improve the quality of services and reduce defects by using the cycle of defining, measuring, analysing, improving and controlling (DMAIC).

4.5.3 Sigma level of railway operations

We estimate the Sigma Level of performance in the Great Britain railways by calculating DPMO. The major defect for train services that does not meet 'customer requirements' is delay. The percentage of trains that arrive late is the primary index of the level of service in railways. In Great Britain, it is calculated by public performance measure (percentage of passenger trains that arrive at their destination on time which is not later than 5 minutes for local services and not later than 10 minutes for inter-urban trains). The national Public Performance Measure for the year ending 30 April 2011 is 90.8% (Network Rail, 2011c) . There is one opportunity per train for defects (i.e. being late or not) therefore the DPMO would be:

$$DPMO = \frac{9.2}{100 \times 1} \times 1000000 = 92000$$

By using one of the online sigma-DPMO calculators, the Sigma level of performance is estimated to be 2.83 σ (WCM, 2011). We compare it with the Sigma levels of some other industries as in Figure 4-10.

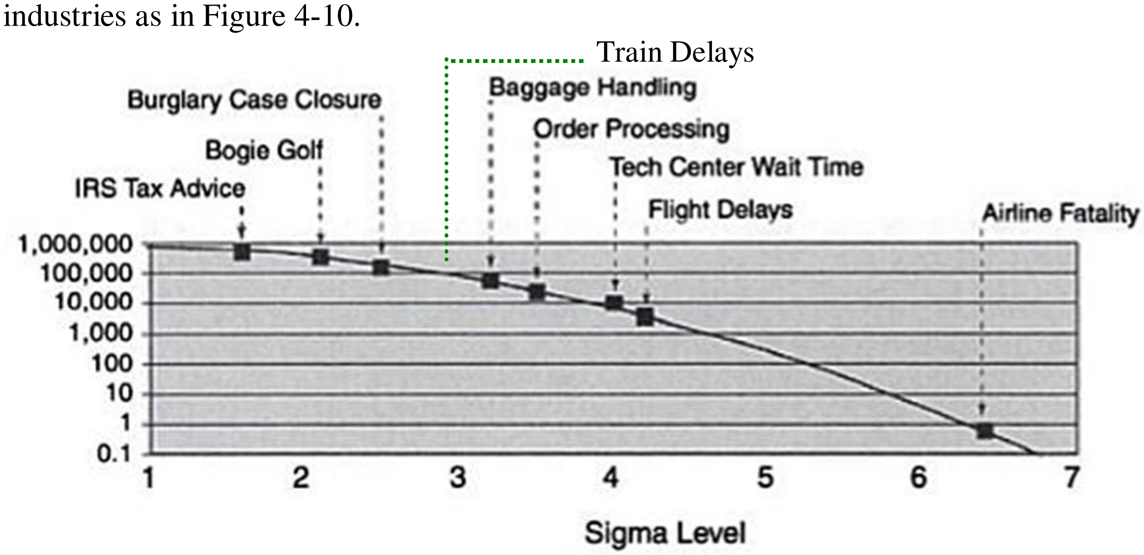


Figure 4-10- Adding Train delays to Sigma levels and DPMO estimations for some industries (Keller, 2001)

Moving from left to right, the quality of services improves. The companies that have quality improvement programs usually operate at 3σ to 4σ. Companies that operate at 2 σ to 3σ "cannot be profitable for very long, so, not surprisingly, only monopolies, government agencies or others with captive customers operate at these levels" (Keller, 2011). Although the Sigma level of performance for train delays is close to 3σ, this concept explains the enormous costs of railways, need for improving efficiency and the value for money as suggested by Department for Transport and Office of Rail Regulation (2011). At this service level, rail is mainly used by the so-called 'captive' passengers or cargo that cannot afford or switch to door-to-door car transportation or use airlines.

4.5.4 Adopting Sigma level for railway capacity utilisation

To improve railway capacity utilisation, we use the underlying concepts and tools of Six Sigma and adopt its DMAIC cycle for the concept of capacity utilisation in the next chapters:

- Defining railway capacity and goals of capacity utilisation
- Measuring capacity utilisation metrics
- Analysing capacity allocation and utilisation
- Improving capacity utilisation
- Controlling

While studying all the above-mentioned aspects of capacity utilisation, the major emphasis of the rest of the thesis will be on developing methods for the analysis stage as it is a key step affecting railway planning decisions. As discussed in 4.3.1, there is a lack of holistic methods to analyse capacity utilisation.

4.6 Summary and conclusions

Railway capacity utilisation is a multidisciplinary area. Hence, it needs a holistic, systems engineering approach. Considering the railway as a system, inputs, outputs, control mechanism, hierarchy and stakeholders are subsequently identified. Adopting a system approach toward capacity utilisation is needed as currently there is no holistic metric for capacity utilisation analysis and each of the existing metrics consider only one aspect of capacity utilisation. A systems approach would also help to overcome the inefficiencies caused by segmentation in the structure of railways as well as the fragmentation of railway engineering knowledge between several disciplines. A systems engineering approach would make it possible to find 'the weakest link of the chain' in capacity utilisation, use innovative improvements for capacity utilisation, increase efficiency and make holistic tactical decisions for track capacity allocation. General system laws were explained in the context of railway capacity utilisation to provide a system thinking foundation.

To address and enhance different aspects of railway capacity utilisation, the Six Sigma methodology is chosen for its holistic and systems approach. Its DMAIC cycle will be adopted to answer the research questions for defining, measuring, analysing and improving capacity utilisation at the tactical level. The main emphasis will be on the analysis stage to develop methods that are holistic, multidisciplinary, avoid 'index numbers', help to find the 'weakest link of the chain' and have the appropriate level of simplicity/complexity for the tactical planning.

5 Defining, Measuring and Analysing Railway Capacity Utilisation for the Passenger Sector

Based on the concept of DMAIC cycle in Six Sigma, in this chapter a comprehensive methodology is developed for defining, measuring and analysing capacity utilisation in the passenger railway operation sector.

5.1 Defining Railway Capacity Utilisation

The heart of capacity utilisation is using the infrastructure efficiently and avoiding its waste. This concept is very close to 'lean thinking' which aims for 'elimination of waste in all forms' (Moore, 2007). To define railway capacity utilisation in a holistic manner, we move toward 'lean capacity utilisation' and will try to measure it accordingly. In order to increase efficiency in railway capacity utilisation, we adopt a 'Six Sigma' approach which is a widely used process management methodology for increasing efficiency and quality while decreasing costs.

5.1.1 Introduction to lean thinking

Although lean thinking has its roots in manufacturing industry, it has proved to be very successful in a wide range of industries both public and private. The paragraphs below provide a brief summary of major lean thinking concepts as reviewed by Womack et al. (2007). Lean manufacturing or simply lean is a practice that originated from the Toyota car manufacturing company in the early 1990s. It aims to improve manufacturing and service processes by 'preserving value with less work'. It defines what is valuable from a customer's point of view and eliminates non-value generating activities. In order to preserve the quality of product and service with less work, every sort of wasting activity must be reduced as far as possible. Muda is a Japanese word that in lean production terminology that means "waste or any activity for which the customer is not willing to pay".

As summarised bt Womack et al. (2007), original seven sources of muda or waste according to Taiichi Ohno are:

- Transportation (moving products that are not actually required to perform the processing)
- Inventory (stacks of work in process and finished product)
- Motion (more movements of people or equipment than are required to perform the processing)
- Waiting (waiting for the next production step)
- Over-Processing (the product with extra steps)
- Over-production (of products that are not needed)
- Defects (in the products)

Womack et al. (2007) suggest another source of muda which is producing goods or services that do not meet customer demand or specifications.

Muda or waste can be avoided by using five principles of Lean:

- Specifying value from the ultimate customer's point of view (not engineers', etc.)
- Identifying value stream
- Flow (making a proper flow of value-generating steps)
- Pulling (letting the customer pull the product)
- Pursuing perfection to reduce costs and time, and to improve the quality

5.1.2 Discrete nature of railway capacity utilisation

Passengers and freight cannot use the railway infrastructure directly; they are packed into trains. Railway capacity is used in discrete steps (as opposed to road capacity that can be continuously used until it is saturated at a standstill level). These discrete steps can be taken in various ways and different combinations of train types, speed and levels of service to generate added value. Value is an expression of "the relationship between function and resources where function is measured by the performance requirements of the customer (such as quality of service) and resources are measured in materials, labour, price, time, etc. required to accomplish that function" (SAVE, 2007).

The railway network can be analysed at three levels - macro, meso and micro - as described in detail by Erol et al. (2008) and Gille et al. (2008) which are schematically shown in It is also important to consider different levels of capacity utilisation. Hereby we define two categories:

- Macro capacity utilisation : Quantity of discrete steps to use railway capacity (e.g. the number of trains and train paths)
- Micro capacity utilisation: Quality of discrete steps to use railway capacity (e.g. Load factor that determines how efficiently the allocated train paths are used)

To efficiently utilise the railway capacity, both aspects should be considered. (Khadem Sameni et al., 2011b)

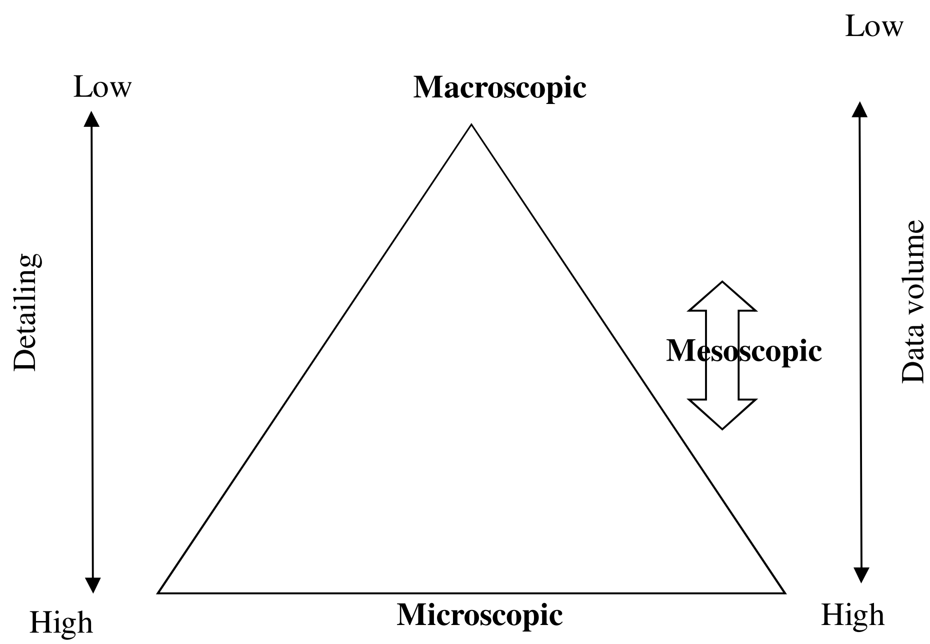


Figure 5-1- Infrastructure Models (Radtke, 2008)

5.1.3 Defining lean railway capacity utilisation

The lean thinking concept has not been studied for railway capacity utilisation. The closest applications are using it for improving a port's performance by Marlow and Paixão Casaca (2003), Loyd et al. (2009) and Cetin and Cerit (2010) as well as applying it to railway classification yards by Dirnberger and Barkan (2007). With our systems approach, we move toward defining and measuring lean railway capacity utilisation to eliminate waste of capacity as far as possible.

We define lean railway capacity utilisation as: "The ability of the infrastructure to generate added value by enabling passengers to reach their destination as planned". To define value, we refer to the concept of transportation itself as presented early in this thesis in section 1.1. Therefore, the more passengers that can be transported further toward their destinations in the unit of time, the more added value is generated by utilising the capacity of the infrastructure. The term 'as planned' emphasises the quality aspects of capacity utilisation such as avoiding delays and ensuring safety. By this definition, whatever does not generate added value, i.e whatever hinders, disturbs or negatively affects this process, is a waste of practical capacity or 'muda' in the lean terminology. This way of defining railway capacity utilisation encompasses both macro and micro capacity utilisation. For example if an empty train moves in the system, as no added value is generated, it is a waste of capacity. Therefore lean capacity utilisation is a function of the number of passengers transported and the distance travelled and is inversely related to the time as the equation below and Figure 5-2 summarise:

$$\textit{Lean capacity utilisation} = f(\textit{macro capacity utilisation} \times \textit{micro capacity utilisation})$$

$$\textit{Lean capacity utilisation} = f(n \times d \times \tan(\alpha)) = f(n.S)$$

d: distance
n: number of passengers

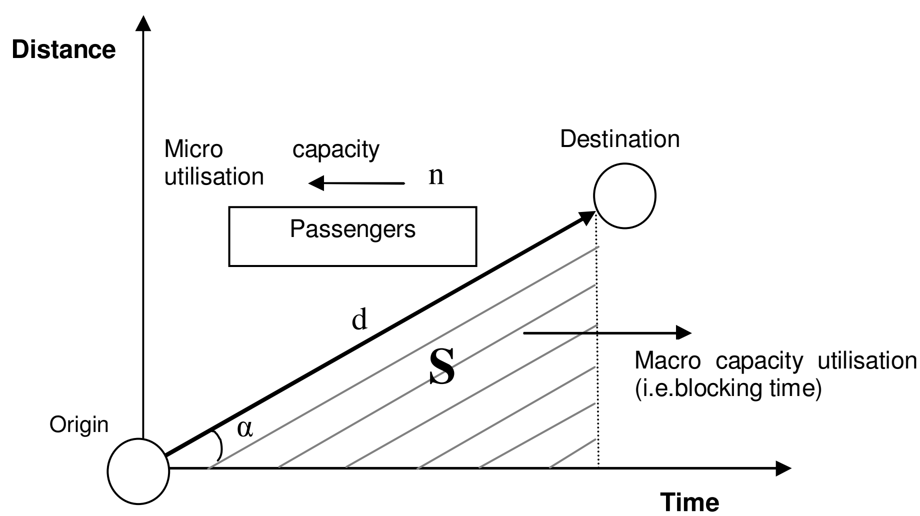


Figure 5-2 - Lean capacity utilisation (Khadem Sameni et al., 2011b)

The area marked as S in Figure 5-2 resembles blocking stairs and how macro capacity utilisation is calculated by the UIC 406 and CUI methods. This definition of capacity utilisation has also some similarities with the concept of 'traffic energy' which Hertel (1992) defines as traffic flow (number of trains per unit of time) multiplied by average speed (Pachl, 2009). This definition of capacity utilisation suggests adding an element of micro capacity utilisation (like load factor) to the above mentioned approaches. Defining lean railway capacity utilisation also paves the way toward new approaches for analysing capacity utilisation by assessing the value that is being generated or wasted.

5.1.4 Sources of practical capacity waste

Following the definition of the lean capacity utilisation, now we can identify some major sources of practical capacity waste or 'muda'. It may occur by means of any underlying factor that was described in chapter two. Some examples are described below:

- Smaller α resulting in macro capacity under-utilisation
    - Scheduled waiting time (although necessary for a feasible timetable)
    - Buffer time (although necessary for quality of service)
    - Delays
    - Dwell time at stations (which is a trade-off between reduced access/egress time for users of the stop and increased in-vehicle time for others on the train)[1]
    - Speed restrictions
    - Inefficient signalling systems
    - Conflicting train routes (in junctions, stations, etc.)
- Smaller d resulting in macro capacity under-utilisation
    - Short run, local services
- Smaller n, resulting in micro capacity under-utilisation
    - Allocating capacity to a service generating less value where it can be allocated to a more valuable service (e.g. low load factor regional services as compared to intercity trains)

In the rest of the thesis we will try to develop methodologies for measuring and analysing lean capacity utilisation.

5.2 Measuring and Analysing Capacity Utilisation by Data Envelopment Analysis

The measuring phase (M) of the DMAIC cycle in Six Sigma "gathers data to establish the current state" and the analyse phase (A) "interprets the data to establish cause-and-effect relationships" (George, 2002). In this thesis, these two phases are combined to develop methods of measuring and analysing capacity utilisation. In this chapter two novel

[1] Optimal stopping patterns have been analysed using OR techniques like the work by VUCHIC, V. R. & NEWELL, G. F. 1968. Rapid transit interstation spacings for minimum travel time. *Transportation Science,* 2**,** 303-339.

methods for capacity utilisation analysis at stations and for passenger train operators are presented based on the data envelopment analysis (DEA).

5.2.1 Introduction to DEA

Data envelopment analysis is a widely-used method of evaluating performance and a breakthrough in analysing relative efficiency. Its building blocks were laid by Farrell (1957) as previous attempts "failed to combine any satisfactory measure of efficiency". DEA is a powerful non-parametric tool that spans the disciplines of management science, operational research, economics and mathematics (Zerafat Angiz et al., 2010). It is especially helpful for evaluating performance where there are complex (or unknown) relations between multiple inputs and multiple outputs.

Efficiency is commonly assessed by the ratio of generated outputs to inputs (Cooper et al., 2006). If it is considered in the wider context of value for money, it can be part of the chain of 'economy', 'efficiency' and 'effectiveness' or "three E's" as described by Booz & Company (2011):

- Economy: how cheaply inputs are provided
- Efficiency: how much output is produced by using inputs
- Effectiveness: the extent of delivering desired outcomes by the cost of producing outputs

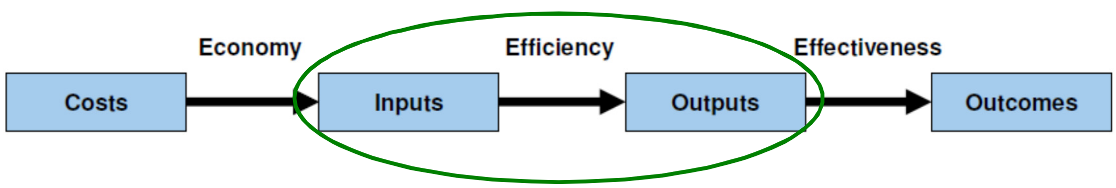


Figure 5-3 - Value for money represented by 3 E's (Booz & Company 2011)

DEA aims to "provide a satisfactory measure of efficiency that takes into account of all inputs yet avoiding index number problems" (Farrell, 1957). As the co-winner of the Economics Nobel prize in 1969 has put it:

"The index-number problem arises whenever we want a quantitative expression for a complex that is made up of individual measurements for which no common physical unit exists. The desire to unite such measurements and the fact that this cannot be done by using physical or technical principles of comparisons only, constitutes the essence of the index-number problem and all the difficulties centre here." (Frisch, 1936)

Farrel's (1957) work was developed further by Charnes et al. (1978a) and Banker et al. (1984). Currently, data envelopment analysis can analyse the relative efficiencies of different units with the same types of inputs and outputs such as different branches of banks, schools, hospitals, etc. (Thanassoulis, 2001). Therefore it has been widely used for many different entities in many different contexts (Cooper et al., 2006). The relative efficiency is identified by analysing a weighted sum of outputs to a weighted sum of inputs (Zerafat Angiz et al., 2010).

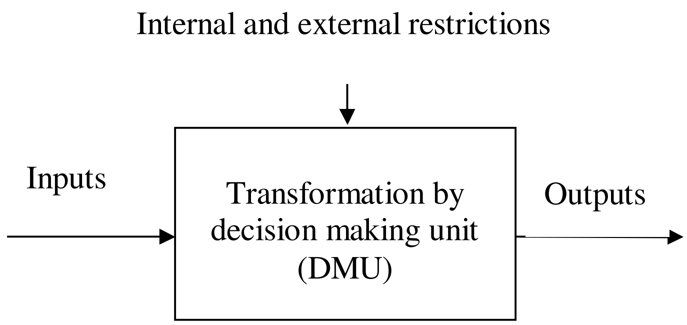


Figure 5-4 - Transforming inputs to outputs by a DMU (Thanassoulis, 2001)

As Figure 5-4 suggests, decision-making problems for an economic agent have three basic features: inputs which are variables chosen by the agent; restriction to choose the set of feasible values and functions that assign values to the outputs generated (Ray, 2004). By taking into account these inputs and outputs, data envelopment analysis can be used for: identifying the most productive and efficient units, the scope for efficient use of inputs or increasing outputs, the marginal rate of substitution between different inputs and productivity change over time (Thanassoulis, 2001).

5.2.2 DEA models

The DEA model maximises the efficiency of each decision making unit (DMU) by maximising the ratio of weighted outputs to weighted inputs subject to satisfying the condition that the weights are positive and that for every DMU, the efficiency score is less than or equal to unity. Considering n DMUs (stations), m inputs and s outputs, x_{ij} as the input i for DMU j, y_{rj} as the output r for DMU j, u and v as the weights for outputs and inputs and ε as non-Archimedean infinitesimal, the formulation as suggested by Charnes et al.(1978b) would be:

$$\max\ h_o = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \qquad j = 1,\ldots,n$$

$$u_r, v_i \geq \varepsilon \qquad r = 1,\ldots,s \qquad i = 1,\ldots,m$$

h_o = efficiency of the unit under assessment

u_r = weight given to output r
y_{ro} = amount of output r for unit under assessment
v_i = weight given to input i
x_{io} = amount of input i for unit under assessment
g_o = efficiency of the unit under assessment
ω_i = weight given to input i in the linear model
μ_r = weight given to output r in the linear model

The above model is a fractional programming and the linear version of the above formulation is:

$$\min g_o = \sum_{i=1}^{m} \omega_i x_{io}$$

$$\sum_{i=1}^{m} \omega_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0$$

$$\sum_{r=1}^{s} \mu_r y_{ro} = 1$$

$$\mu_r, \omega_i \geq 0$$

This is the variable return to scale (VRS) model. The $\sum_{r=1}^{s} \mu_r y_{ro} = 1$ constraint allows for the convex combination and eliminating it results in the constant return to scale model (CRS) (Cook and Zhu, 2008) . DEA models can have two general orientations: input oriented or output oriented. The input oriented model tries to minimise inputs while at least the given level of outputs are produced while the output oriented model tries to maximise outputs while no more than observed level of inputs are used (Cooper et al., 2006).

5.2.3 Application of DEA in railways

There have been two quite isolated trends in railway transportation analysis and planning:

Engineers have been concerned with 'operational efficiency' through different methods of improving capacity utilisation, mainly operations research, simulation, parametric and analytic methods (as presented and discussed in section 2.3 - Methods of estimating railway capacity utilisation). Economists have been concerned with 'cost efficiency' and productivity by partial/total productivity measures, data envelopment analysis and stochastic frontier analysis (as reviewed in Table 5-1).

Table 5-1- Efficiency and productivity studies in railway (Merkert et al., 2010)

Study	Method	Sample	Inputs	Outputs
(Nash and Preston, 1994)	Partial productivity measure (PPM)	14 European railways 1970-1990	Staff/train-km; market share; receipts/total cost	
(Nash and Shires, 2000)	Partial productivity measure (PPM)	11 European railways 1989-1994	Train-km/track-km; train-km/staff; market share; traffic units/train-km; operating cost/ train-km; receipts/traffic units; revenue/costs	
(Oum and Yu, 1994)	Data envelopment analysis (DEA)	19 railways in Europe and Japan	Staff; energy consumption; rolling stock	Passenger-km; freight-tonne-km
(Gathon and Pestieau, 1995)	Stochastic frontier analysis (SFA)	19 European railways 1986-1988	Engines and railcars; staff, length of not electrified/electrified lines	Sum of passenger-km and freight-tonne-km
(Coelli and Perelman, 1999, Coelli and Perelman, 2000)	Data envelopment analysis (DEA) and corrected ordinary least squares (COLS)	17 European railways 1988-1993	Staff; rolling stock; track length	Passenger-km; freight-tonne-km
(Cantos and Maudos, 2001)	stochastic frontier analysis (SFA)	16 European railways 1970-1990	Operating cost; labor cost, energy, material/external	Passenger-km; freight-tonne-km
(Cantos et al., 2002)	Data envelopment analysis (DEA)	17 European railways 1970-1995	Operating cost; track-km	Passenger-km; freight-tonne-km
(Loizides and Tsionas, 2004)	Total factor productivity (TFP)	10 European railways 1969-1993	Staff; capital cost (interest and depreciation); energy cost	Sum of passenger-km and freight-tonne-km weighted with revenue share

Study	Method	Sample	Inputs	Outputs
(Rivera-Trujillo, 2004)	Partial productivity measure (PPM)	14 railways in Europe and 5 American railways 1977- 1999	(Passenger-km + Freight-tonne-km)/ operating staff; traffic units/operating staff (1980-1999)	
(Rivera-Trujillo, 2004)	stochastic frontier analysis (SFA)/ Total factor productivity (TFP)	14 railways in Europe and five American railways 1977- 1999	Staff; rolling stock (four categories)	Passenger-km; freight-tonne-km
(Hatano, 2005)	Partial productivity measure (PPM)	15 railways worldwide	(Passenger-km + freight-tonne-km)/total route Length	
(Cowie, 2005)	stochastic frontier analysis (SFA)	British TOCs 1996-2000	Staff; rolling stock; track length	Train-km
(Growitsch and Wetzel, 2009)	Data envelopment analysis (DEA)	54 railways in 27 countries 2000-2004	Staff; rolling stock; track-km; operating expenditure	Train-km; passenger-km; freight-tonne-km
(Driessen et al., 2006)	Data envelopment analysis (DEA)	14 European railways 1990-2001	Staff; track length; rolling stock	Passenger-km; freight-tonne-km
(Smith and Wheat, 2007)	stochastic frontier analysis (SFA)	26 British TOCs 1996-2006	Staff and rolling stock and other op. cost; wage prices, rolling stock characteristics; policy variables	Train-km/route-km, route-km, vehicle-km/train-km
(Wetzel, 2008)	stochastic frontier analysis (SFA)	31 European railways 1994-2005	Staff; rolling stock; network length	Passenger-km; freight-tonne-km

Study	Method	Sample	Inputs	Outputs
(Cantos et al., 2010)	Data envelopment analysis (DEA)	16 European rail systems 1985-2004	Staff; rolling stock (Passenger vs. freight); network length	Passenger-km; freight-tonne-km
(Merkert et al., 2010)	Data envelopment analysis (DEA)	43 Swedish, German and British train operating firms	Material (Annual amount spent on operation including depreciation and rolling stock lease costs but excluding all staff costs); total staff	Train-km
			Material; managerial and administrative staff; the remaining production staff	Train-km; passenger-km
			Material; managerial and administrative staff; the remaining production staff	Train-km; Tonne-km

Although ‘cost efficiency’ and ‘operational efficiency’ affect each other, they have different concerns. For example, ‘cost efficiency’ revolves round the inputs and outputs that have monetary value like transaction costs, operating costs and income[1] whereas in ‘operational efficiency’, quality and quantity of services are the major goal. These two aspects are closely interrelated for railway capacity utilisation. The powerful tool of data envelopment analysis and its underlying concepts has not been used for assessing capacity utilisation in railways. The main aim of this chapter is to establish a bridge between engineering and economic approaches by using data envelopment analysis to analyse the relative operational efficiency of railway stations and passenger operators in utilising railway capacity.

5.2.4 DEA versus Other Approaches to Railway Capacity Utilisation Analysis

The different approaches to railway capacity utilisation analysis each have their strengths and weaknesses. Table 5-2 compares major aspects of capacity utilisation analysis by simulation, operations research, parametric models, analytical methods and DEA.

DEA does not need in-depth knowledge of the different disciplines that affect railway capacity utilisation. The inputs and outputs alone related to each discipline are sufficient for analysing the relative efficiency of DMUs in using railway capacity. For example for rolling stock, the realm of mechanical engineers, important but simple and easily understandable inputs like the age of rolling stock and the number of trains that affect capacity utilisation can be used. From economic disciplines, revenue, costs and profits can be chosen. From the discipline of civil engineering, the number or length of platforms, etc. can be used. In essence, all the concerns of the engineers, economists, and operation researchers can be accommodated in one single analysis. The real beauty and advantage of data envelopment analysis over other methods of capacity utilisation analysis is that it does not need to know the relationship between these inputs and outputs (e.g. what is the relationship between the number of available platforms and profit or the age of rolling stock and costs, etc.). There is also no need to have a common unit of measurements between variables.

Data envelopment analysis provides an “objective basis for evaluating the performance” and “the outcome at the highest level of efficiency proves an absolute standard for management” (Ray, 2004). By identifying non-efficient units, the weakest link of the chain can be identified on objective grounds. It compares and ranks the relative efficiency of different decision-making units that transform inputs to outputs. Thus the weakest link of the chain (e.g.the least efficient station/train operator, etc.) can be identified and optimum values for their inputs and outputs are determined. Based on the results of analysis, benchmarking techniques from the most efficient units can be used to improve the less efficient ones. Moreover, even positive or negative changes in efficiency can be monitored over years by comparing the results of data envelopment analysis for data sets of different years or control periods. In this way, the performance of individual units (including the most and least efficient ones) and the impact of capacity utilisation improvement measures can be tracked.

[1] Prices are not usually used in DEA models so cost efficiency is indirectly analysed.

The main limitations of DEA are that it is an extreme point method, measurement errors can affect the results and it can only measure relative (and not absolute) efficiency (Cooper et al., 2006). Taking all the above mentioned points into account, DEA is suggested by the present thesis as a meso-tool for assessing the relative efficiency of units that utilise capacity. In the following sections, two novel methods are suggested to use DEA for analysing the efficiency of capacity utilisation by train operators and at train stations.

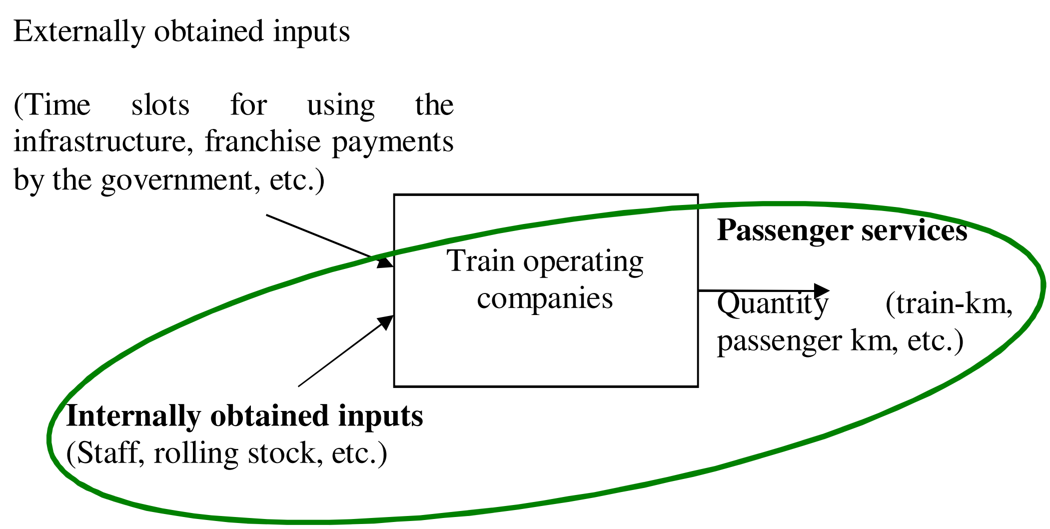
Table 5-2 -Comparing four approaches to railway capacity utilisation analysis

	Operations research and simulation	**Parametric models**	**Analytical methods (UIC 406 and CUI)**	**DEA**
Aim	Optimising sub-problems of capacity utilisation (timetabling, train routing, etc.)	Analysing capacity utilisation curve by the relationship between parameters of infrastructure, timetable, operation, etc.	Giving a general overview of how much the infrastructure is not idle by compressing the timetable to minimum technically possible and generating a capacity utilisation index	Comparing the relative efficiency of different units(e.g. stations and train operators) in capacity utilisation and finding 'frontier of efficiency'
Objective	Usually single objective (mainly minimising delays). Multi objective functions are possible but make solving the model much more complicated	Usually single objective (The curve of train delays v.s. capacity utilisation)	Single objective	Multiple objectives
Characteristics	Operations research: Non-parametric; deterministic or non-deterministic Simulation: parametric or non-parametric; non-deterministic	Parametric and deterministic	Non-parametric and Deterministic	Non-parametric and deterministic
Number of studies done	Very high	Limited	Moderate	Moderate (but mainly used for analysing cost efficiencies of different railways not operational efficiency as needed for capacity utilisation)
Solution time	Time-consuming due to computational complexity	Fast	Fast due to static and deterministic nature	Fast due to non-parametric and linear programming formulation

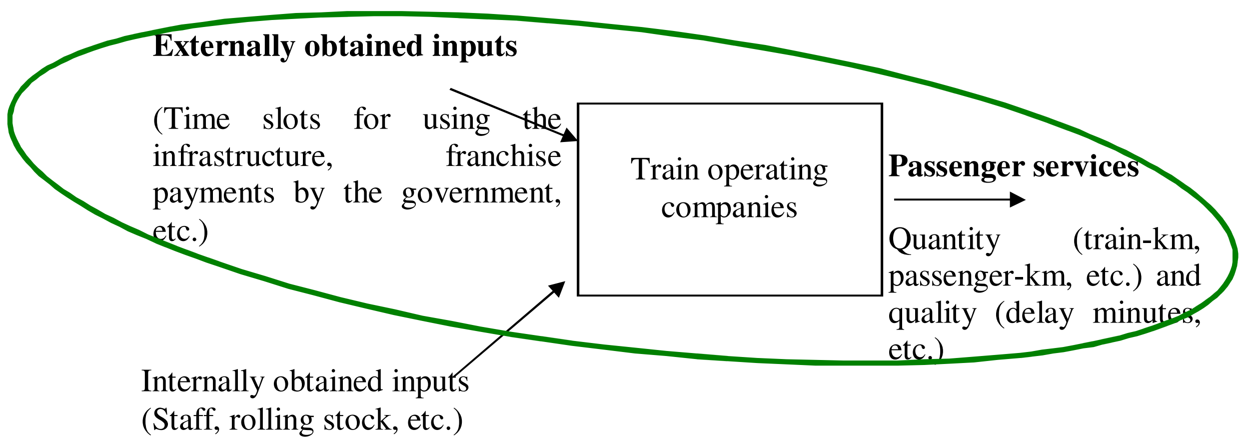
	Operations research and simulation	Parametric models	Analytical methods (UIC 406 and CUI)	DEA
Relation between inputs and outputs	Have a meaningful and known relationship	Have a meaningful and known relationship	Have a meaningful and known relationship	No need to know how inputs and outputs relate to each other
Depth and breadth of details	Many variables within the same discipline	Some parameters of infrastructure, timetable and operation	Limited to blocking time stairs of trains as input and capacity utilisation index as output	No limitation – can handle multidisciplinary inputs and outputs
Geographical Scope of case studies	Usually small parts of the network or as far as computational capabilities allow	Stretch of a line to the whole network	Stretch of a line to the whole network	Stretch of a line to the whole network
Examples in the literature	Sections 2.3.3 and 2.3.4	Section 2.3.2	Section 2.3.1	Table 5-1

5.3 Measuring and Analysing Capacity Utilisation by Passenger Operators

The efficiency studies in railways have never been focused on analysing the efficiency in utilising allocated capacity of the infrastructure to produce reliable and valuable services. In a broader sense, as summarised in Table 5-1, the focus of existing research has been on 'internally obtained inputs' such as staff and rolling stock rather than 'externally obtained inputs' such as track capacity and franchise. The track capacity is limited so it is essential to analyse how well this resource is used when allocated. Moreover, in the outputs, quality of service (e.g. delay minutes) has never been considered, and provides a worthwhile addition to the approach adopted in the current research. Figure 5-5 and Figure 5-6 compare the approach adopted in the current study with the past approaches in the literature.



a) Past approaches to efficiency (with the train-operating company as the stakeholder)



b) The approach of the current research to efficiency (with the government as the stakeholder)

Figure 5-5 - Comparing the current research approach to efficiency analysis with the past approaches

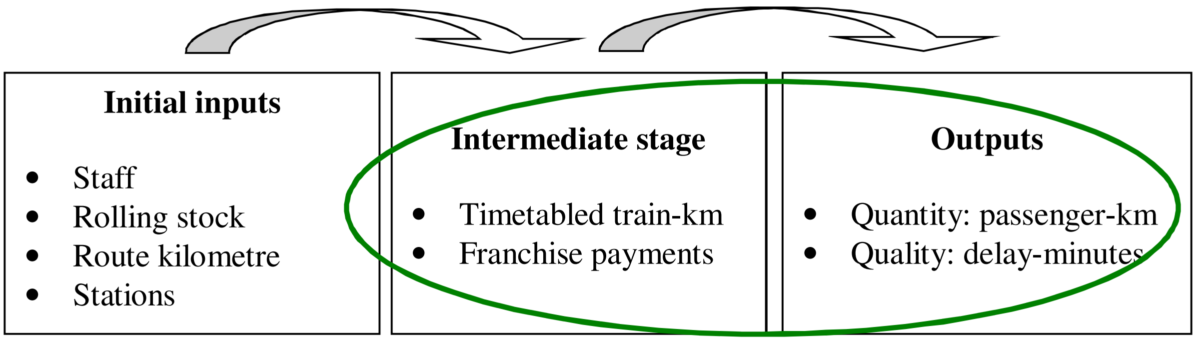


Figure 5-6 - Transformation of inputs into outputs by train - operating companies and the adopted approach of the current research

After privatisation of railways in Europe and in a typical vertically separated railway, objectives, interests and concerns are segmented as well. The government, passenger operators, freight operators and infrastructure authority have different responsibilities, objectives and concerns. As a consequence of this segmentation, analysing efficiency is dependent upon who is chosen as the stakeholder. In this research we consider the government as the stakeholder who is responsible for the socio-economic welfare of society. The efficiency of railway passenger operators will be judged by the extent to which they use public resources as inputs to generate valuable passenger services for society. Both quantity and quality of services will be considered to assess the value of provided services.

5.3.1 Efficiency in Great Britain's Railway Network and Value for Capacity

It is forecast that passenger demand for rail will double and freight demand will increase by 140% over the next 30 years (Network Rail, 2011a). Quality of services has considerably improved too. The Public Performance Measure (PPM) is the index that is usually used to reflect the quality of service which is "the percentage of passenger trains that arrive at their destination on time (not later than 5 minutes for local services and not later than 10 minutes for inter-urban trains). If a train is cancelled or is later than the threshold, it has not met the criteria." The Public Performance Measure for the year ending 8 January 2011 is 90.8% as compared to 78% of 10 years ago (Network Rail, 2011b). However, these achievements have incurred extensive costs. A recent study by Lovell et al. (2011) contends that "Britain's rail infrastructure manager faces an efficiency gap of 40 per cent against European best practice and that train operating costs have also risen substantially, both because of rising factor prices (wages and fuel) and because of deteriorating productivity".

Figure 5-7 shows a breakdown of costs in Great Britain's railway network and Figure 5-8 shows the actual financial flows.

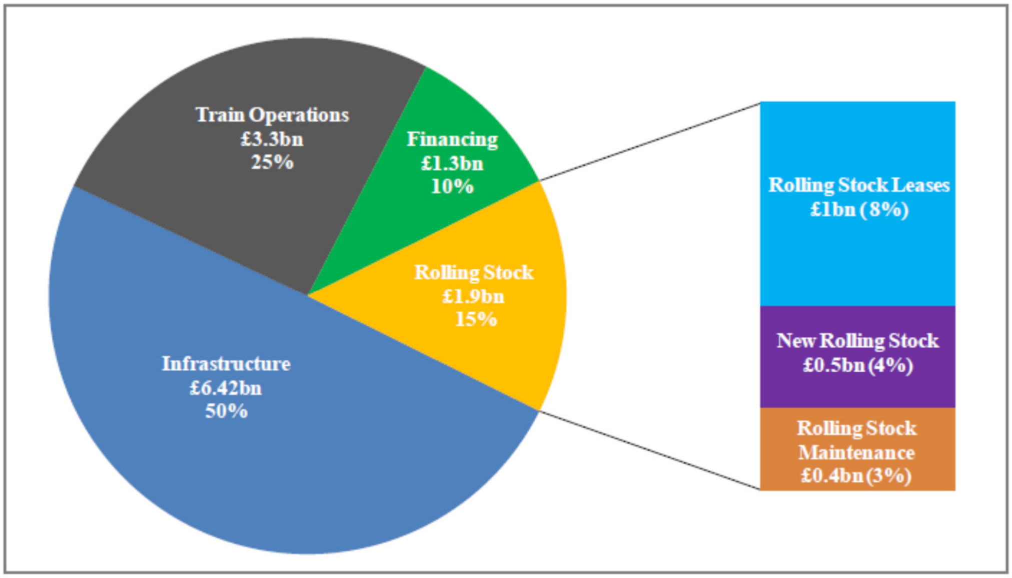
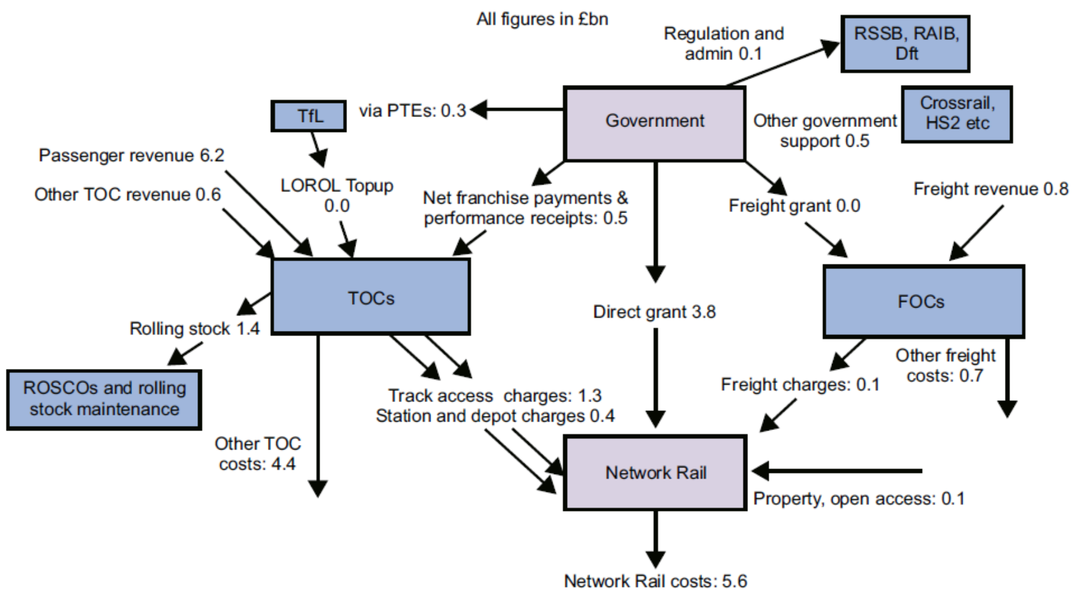


Figure 5-7 -Total GB rail cost breakdown 2009/10 (2009/10 prices) (Atkins, 2011, Arup, 2011)



TOC: Train Operating Company
FOC: Freight Operating Company
TFL: Transport for London
ROSCOs: Rolling stock companies
LOROL: London Overground

PTE: Passenger Transport Executive
RSSB:Railway Safety and Standards Board
HS2: High-speed Two
RAIB: Rail Accident Investigation Branch

Figure 5-8 - Financial flows in GB rail 2009/10 (£ billion) (Department for Transport and Office of Rail Regulation, 2011)

As depicted in Figure 5-7, the infrastructure accounts for the major proportion of costs of the railway industry in Great Britain, making allocated track capacity an expensive resource. Improving the efficiency of utilising the infrastructure along with efficient train operations can be robust means of decreasing costs to achieve the targeted annual cost saving of up to £1 billion as set by Department for Transport and Office of Rail Regulation (2011).

The benchmarking study of four European railways by Civity Management Consultants (2011) suggests that Great Britain has the most competitive market structure and that market shares are distributed among different operators (Figure 5-9) whereas in Sweden and the Netherlands state-owned companies still dominate. However with the above-mentioned massive costs, analysing passenger train operators' efficiency is highly important.

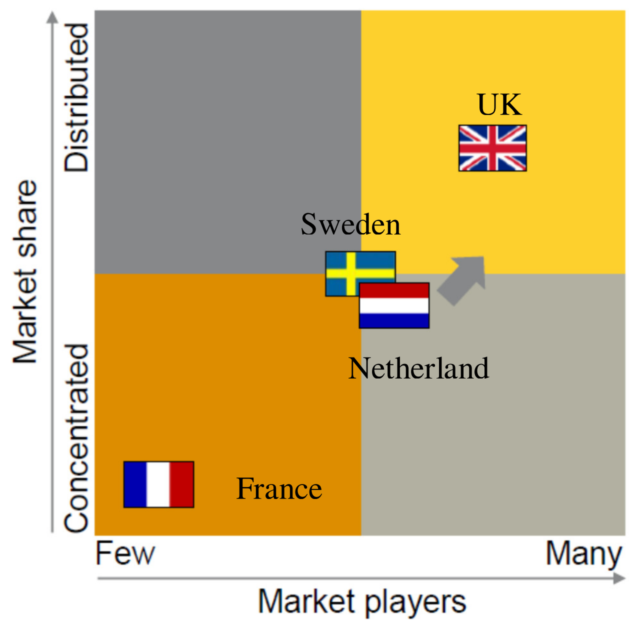


Figure 5-9 - Market structure of 4 European railways (Civity Management Consultants, 2011)

5.3.2 Intrinsic characteristics of the model

In the following sections the intrinsic characteristics of the DEA model to analyse capacity utilisation for passenger train operating companies in Great Britain are summarised.

5.3.2.1 Stakeholder

Government as the regulator of socio-economic welfare of the country provides "net franchise payments" to the train operating companies to run passenger train services. These payments in financial year 2009-2010 were 500 million pounds (Figure 5-8). Therefore the key stakeholder of the model to analyse efficiency of passenger train operating companies in Great Britain is the government. It should be noted that this model is novel and different from the existing DEA models in the literature as reviewed in Table 5-1. As illustrated in Figure 5-5, these models focus on the internally obtained inputs and their stakeholder is the train operating companies. The model proposed in this thesis considers a bigger picture where the government is the key stakeholder. The model can analyse the relative performance of train operating companies for the government.

5.3.2.2 Controllable inputs

For the key stakeholder of the model, the government, there are two main categories of controllable inputs to be given to train operating companies. One is the net franchise payments to enable them to run passenger train services (as these services are not profitable like railway freight transportation). The other main category of inputs is the allocated amount of timeslots to use the state owned railway tracks (Table 3-1).

5.3.2.3 Output priorities

The output priorities for government to assess the performance of train operating companies are the quantity and quality of services they provide. Quantity of services is twofold: number of passengers and length of haul. Public Performance Measure (percentage of passenger trains that arrive at their destination on time) is usually the main quality indicator of the services. Amount of delays is another side of the punctuality coin.

5.3.3 Choosing DEA Inputs: Externally Obtained Resources

In data envelopment analysis, "Inputs are defined as resources utilised by the DMUs or conditions affecting the performance of DMUs (Ramanathan, 2003). Timetabled train-km is the best proxy variable to reflect infrastructure utilisation by a train operator: the more trains it runs on the infrastructure and the longer they run, the more it uses this valuable resource hence the more inputs and chances to generate valuable outcomes there are. It is worth emphasising that the choice of inputs and outputs depends on the process being analysed. Therefore, as analysing the efficiency of capacity utilisation is the main object of our study, unlike the studies mentioned in Table 5-1, train-km is chosen as an input for capacity utilisation analysis. The efficiency of the operators is analysed in terms of transforming this allocated track capacity into passenger services. Few previous studies of efficiency in railways have used 'route-km' as their input for DEA models (as seen in Table 5-1). Route-km is not an exact input to reflect capacity utilisation which is the main goal of this study. It depends how many trains run on this routes. If no train runs on the infrastructure, capacity utilisation is zero according to the UIC 406 capacity leaflet developed by the International Union of Railways (UIC, 2004). The higher the number of trains that run on the infrastructure in the time unit, the higher is the capacity utilisation index.

Franchise payments by government are an external input that can be used for analysing operators' efficiency in capacity utilisation and converting them into valuable train services. It is a public resource which must be used as efficiently as possible.

5.3.4 Choosing DEA Outputs: Public Value of The Services Provided

Outputs are the benefits generated as a result of the operations of the DMUs (Ramanathan, 2003). The value generated by a passenger train operator can be measured in different ways. The first option that comes to mind is to consider the revenue that is generated through ticket sales. However, this cannot be a good index for 'operational efficiency'. Trains running with low load factor but high fares might be 'economically efficient' but they are not 'operationally efficient'.

Passengers transported (the number of passenger journeys) is not by itself informative: one passenger might use the train for a very short distance; one might take the train for a very long distance. Therefore the best measure for estimating the value generated by a train operator through using the infrastructure is passenger-km. Passenger-km is also a very good measure of the environmental effects as CO2 emissions saved by choosing railway as the mode of transportation is likely to be proportional to passenger-km (along with other factors such as train loadings, mode switching, traction energy source, etc.).

Considering timetabled train-km as input and passenger-km as one of the outputs also covers aspects of both 'macro' and 'micro capacity utilisation' as well as 'lean capacity utilisation' as suggested by Khadem Sameni et al. (2011b).

There is a trade-off between railway capacity utilisation and quality of service: higher capacity utilisation increases the risk of primary and secondary delays. Therefore it should be considered for providing a proper capacity utilisation analysis. There is a wide range of data available on the quality of service for each train operator company:

- The number of complaints received per 100,000 passenger-journeys
- National passenger survey results (a detailed survey on quality of services on board and at stations carried out twice per year by Passenger Focus);
- Public performance measure
- Delay minutes

The number of complaints is not a good indicator for quality of service to be included in the DEA model. Complaints can be subjective and mostly originate from train performance. As indicated by the Office of Rail Regulation (2010a), in the financial year 2009-10, 36% of the total complaints were about train service performance, 21% about fares, retailing and refunds and 12% about quality on the train. Therefore a train performance indicator is a better estimate of the quality of service provided by the operator. The quality of services on board and at stations matters but the first priority of passengers is getting to their destinations on time. The Public Performance Measure is a relative index which is why delay-minutes was chosen to indicate the quality of service which is important both for passengers and the network owner. This is also in line with the work of Tongzon (2001) which used delay time (the difference between total berth time plus time waiting to berth and the time between the start and finish of ship working) for analysing the maritime industry through DEA. All the data used in the case study for train operators are extracted from National Rail Trend Year Book 2009-2010 (Office of Rail Regulation, 2010a). Data on train delay minutes for different operators is not included in this comprehensive document but can be found in the 'Annual Return' report published by Network Rail (2010b).

It should be noted that train delay is not a positive outcome. In DEA terminology negative outcomes are called 'undesirable effects' and cannot be used directly in the model as outputs. Methods to handle them have been surveyed by Seiford and Zhu (2002). The most popular methods are: transferring undesirable effects to the input side (as DEA tries to minimise use of inputs) or using the inverse of 'undesirable effects' as outputs (as DEA tries to maximise outputs). Figure 5-10 shows a schematic representation of inputs and outputs for analysing operators' efficiency.

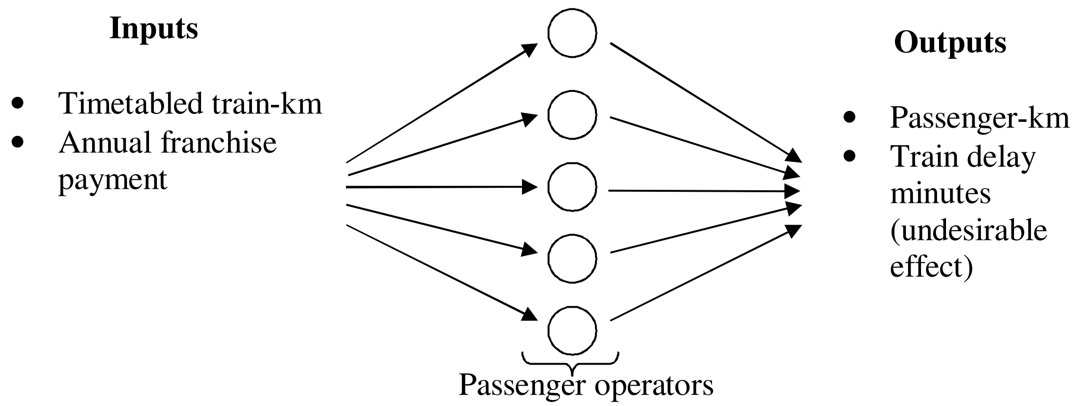


Figure 5-10 - Inputs and outputs for the analysing operators' efficiency in capacity utilisation

Table 5-3 - Descriptive statistics for the operators' case study

	Mean	**SD**	**Min**	**Max**
Delay minutes 09-10 (thousands)	406.26	185.35	69.98	770.55
Passenger kilometres (millions) 09-10	3088.16	1520.71	945.2	5280.9
Timetabled train kilometres (millions) 09-10	29.98	10.17	9.6	44.9
Franchise payment (million pounds) 09-10	201.03	106.64	0[1]	407.3

5.3.5 Analysis of the Results

As the main aim of this case study is to increase the efficiency of railways by cutting costs, the input-oriented DEA model was chosen. The models for constant return and variable to scale (CRS and VRS) were solved using PIM DEA-V3.0 software (Emrouznejad and Thanassoulis, 2011). DEA efficiency scores are presented in Table 5-4.

[1] The Department for Transport was in receipt of franchise payments from few operators resulting in negative data for them. Franchise payments were shifted up to eliminate negative data as Variable Return to Scale (VRS) models are invariant to such transformations COOK, W. & ZHU, J. 2008. *Data envelopment analysis: Modeling operational processes and measuring productivity*, Create Space.

Table 5-4 - Efficiency scores of train-operating companies in the year 2009

Name of the operator	**Geographic area of operation (Network Rail, 2011e)**	**VRS model**	
		Efficiency score	**Rank**
Arriva Trains Wales		0.50	11
Chiltern Railways		1.00	1
Cross Country		0.48	14
East Coast		1.00	1
East Midlands Trains		0.59	8
First Capital Connect		0.71	6
First Great Western		0.97	5
First Scot Rail		0.35	15
London Midland		0.49	12
National Express East Anglia		0.54	9

Name of the operator	Geographic area of operation (Network Rail, 2011e)	VRS model	
		Efficiency score	**Rank**
Northern		0.34	16
South Eastern		0.53	10
Southern		0.49	13
South West Trains		1.00	1
Trans Pennine Express		0.64	7
Virgin Trains		1.00	1

Train-operating companies with the highest average train utilisation (Figure 5-12) tend to get higher efficiency scores. For example East Coast and Virgin Trains which carry the highest number of passengers per train have also received the highest efficiency scores by the DEA model. However when delay-minutes and franchise payments are considered, the ranking is not exactly same as a train-operating company might have not performed well enough to provide punctual services or be cost efficient. For instance First Great Western has the third rank according to the average train utilisation but ranks fifth when the quality of service provided and franchise payments received are considered by the DEA model. To gain 100% relative efficiency, target values as suggested by DEA are shown in Table 5-5. They are calculated by the PIM DEA-V3.0 software (Emrouznejad and Thanassoulis, 2011) based on the distance of Production Possibility Set (PPS) from the efficient frontier as illustrated in Figure 5-11. The efficient decision making units, make the efficiency frontier and provide benchmarks for other units.

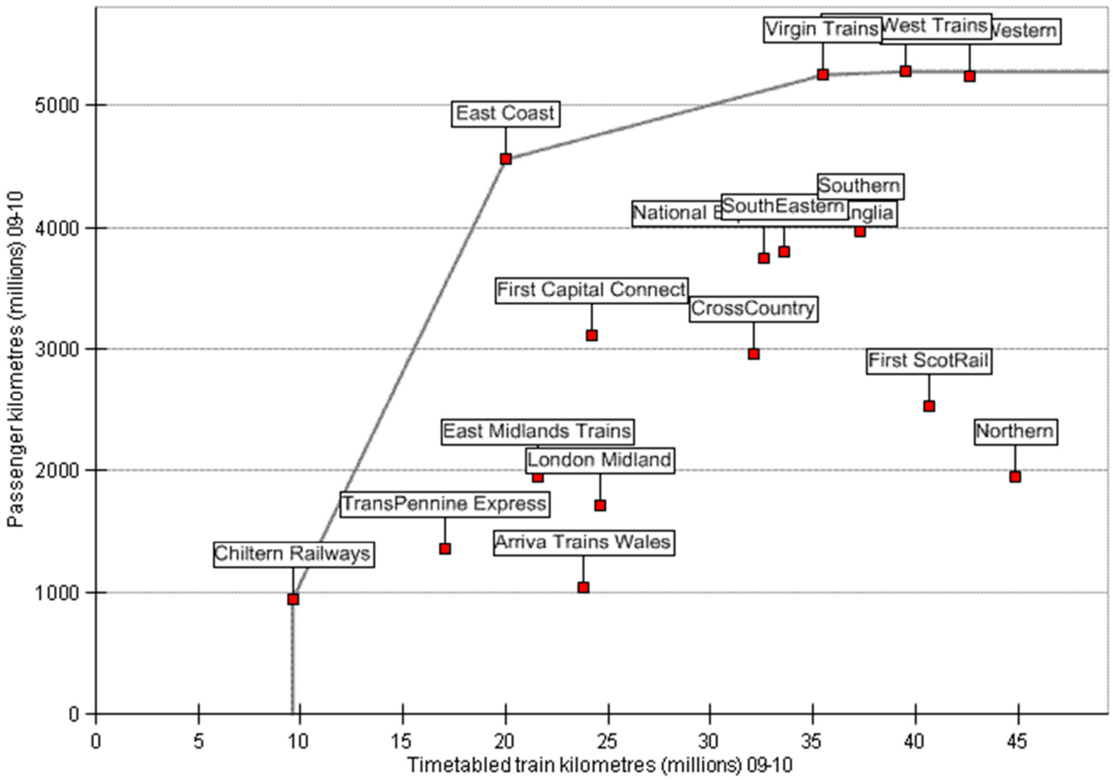


Figure 5-11 Efficiency frontier and production possibility set for passenger operators

Less efficient operators use more track capacity (reflected by timetabled train-km) than necessary to generate passenger-km or are less efficient in producing punctual services or receive more franchise payments than necessary. Reducing non-efficient timetabled train-kilometres (that are not transformed into passenger-km efficiently) would increase train load factor and efficiency of capacity utilisation. Introducing a cap on subsidy would give more incentive to train-operating companies to increase their operational efficiency.

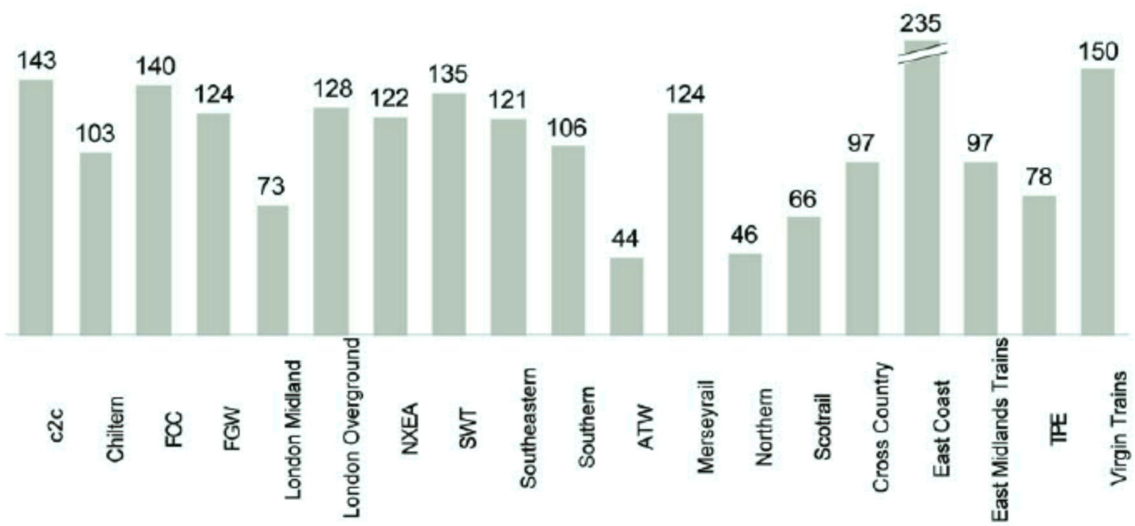


Figure 5-12 - Average passenger per timetabled train (Civity Management Consultants, 2011)

Table 5-5 can provide railway practitioners with insights about how train operators can improve their operational efficiency. The operators that are not operationally efficient, as indicated by Smith (2009), might :

- Have low overall load factor
- Haul a lot of empty seats off-peak
- Haul empty seats long distances to satisfy short distance demand

Two other possible causes might be operating short trains and serving less dense population areas.

Table 5-5- Target values as suggested by DEA

Name	**Delay minutes 2009-10 Gain(%)**	**Passenger kilometres (millions) 09-10 Gain(%)**	**Timetabled train kilometres (millions) 09-10 Gain(%)**	**Subsidy Gain(%)**
Arriva Trains Wales	-61.38	71.86	-49.61	-49.61
Chiltern Railways	0	0	0	0
CrossCountry	-69.99	0	-52	-71.51
East Coast	0	0	0	0
East Midlands Trains	-59.45	4.99	-40.95	-40.95
First Capital Connect	-40.92	15.6	-28.79	-28.79
First Great Western	-13.87	0	-17.63	-2.98
First ScotRail	-78.52	0	-65.2	-73
London Midland	-75.95	7.33	-50.58	-50.58
National Express East Anglia	-72.77	0	-45.8	-49.48

Name	Delay minutes 2009-10 Gain(%)	Passenger kilometres (millions) 09-10 Gain(%)	Timetabled train kilometres (millions) 09-10 Gain(%)	Subsidy Gain(%)
Northern	-83.23	49.54	-66.03	-66.03
SouthEastern	-70.68	0	-46.93	-81.47
Southern	-72.65	0	-50.94	-83.76
South West Trains	0	0	0	0
TransPennine Express	-63.58	3.12	-35.92	-35.92
Virgin Trains	-7.33	0	0	0

Some of the ways train operators can increase their efficiency are:

- Reducing the frequency of their trains to increase their load factor (passenger per train). For example the East Coast operator with the highest relative efficiency has the highest ratio for passenger journeys per trains planned (413.0) and the highest ratio for passenger-km per timetabled train-km (228.6). These ratios for Arriva Trains Wales and Northern were respectively (82.9, 43.5) and (99.9, 43.4).
- Using marketing techniques to attract more passengers to their current services and increase load factor.
- Increasing the reliability of their train services to reduce delays.

The results can also provide helpful insights for railway authorities to plan better for infrastructure and franchise payments. For instance, the results of the model indicate a very low level of operational efficiency for the Cross Country services and the need for drastic cuts in franchise payments and allocated timetable kilometres. A closer look at the geographical area of operation for CrossCountry trains shows an overlap with four other train-operating companies which are operationally very efficient (Table 5-6). This suggests that CrossCountry is not an operationally efficient route and the track capacity and franchise payment for its services should be divided between the other four train operating companies to run the necessary services. This could be a great step toward increasing the efficiency of British railways as targeted by the Department for Transport and the Office of Rail Regulation in the value for money study (2011).

Table 5-6 - Overlap of CrossCountry services with 4 operationally efficient train-operating companies

Name of the operator	Cross Country	East Coast	Virgin Trains	Southwest Trains	First Great Western
Geographical area of operation					
Efficiency score	0.48	1	1	1	0.97

Identifying and reducing inefficient timetabled passenger train kilometres frees up track capacity that can be allocated to freight trains. This generates more revenue which to offset the huge costs of the network and subsequently to invest in improving it.

5.3.6 Tobit Regression

Tobit regression is usually used in the second stage of DEA to assess the relationship between exogenous factors and DEA efficiency scores (Hoff, 2007). Tobit regression is helpful for predicting censored data (when the values are clustered around a threshold) and truncated data (when data is censored below or above some threshold) (Walker and Maddan, 2009). It is named after Tobin (1958) who first applied this model and called it “the model of limited dependent variables” as the dependent variable of his regression model could not be negative (Amemiya, 1985). Efficiency scores range between zero and one and also some efficiency scores are clustered around 1 that is why Tobit regression should be used.

The Tobit model is a linear regression censored below zero with additive error that is normally distributed:

$$y^* = X^{'}\beta + \varepsilon$$
$$\varepsilon \sim N[0, \sigma^2]$$

$$y = \begin{cases} y^* & if\ y^* > 0 \\ 0 & if\ y^* \leq 0 \end{cases}$$

(Cameron and Trivedi, 2005)

In the second stage of the model, a Tobit regression is used to analyze the underlying factors affecting operators’ efficiency. Correlations between efficiency scores and average age of rolling stock, public performance measure, route kilometres operated, passenger satisfaction rates in annual surveys and the number of complaints received are of interest. Tobit regression was done for the VRS model by SPSS V.19, by adding R and Python plug-ins and the ‘SPSSINC_TOBIT_REGR’ application [45]. The results for the Gaussian (normal) assumption are presented in Table 5-7 .

Table 5-7 shows that the efficiency score is positively correlated with serving London (P value < 0. 003). Offering regular services to London was chosen as the criteria hence Scot Rail, that offers a sleeper service to London, received zero for this variable. The efficiency scores are negatively correlated with the average length of journeys for regional services (P value < 0.011). Services that their average length of journeys were less than 40 miles according to the National Rail Trends (Office of Rail Regulation, 2010a) were considered to be regional. The average age of rolling stock and the number of staff employed were found to be insignificant factors.

Table 5-7 Tobit regression results for the Tobit regression

	Coefficient	Std. Error	z Value	Sig.
(Intercept)	.664	.159	4.181	.000
Serving London	.332	.110	3.007	.003
Regional (Average length of journeys less than 40 miles)	-.294	.115	-2.558	.011
Average age of rolling stock	-.003	.009	-.314	.754
Number of employees 09-10	.000	.000	.076	.940

5.3.7 Systems engineering and real world implications

For preserving the quality of outputs in the system a control mechanism is needed which was discussed in 4.2.6. In the tough economic situation, efficiency of using public resources should be controlled more than ever before (section 4.3.8),. A holistic measure (section 4.3.7) is needed to analyse the performance of the train operating companies whereas currently Public Performance Measure is used as the main index for performance analysis. Segmentation after privatisation (4.3.2) has made it more complicated to analyse the performance of all train operating companies in one go. It should be emphasised that UK has the highest number of market players and the most distributed market share for them (Figure 5-9). In such a fragmented system, it is important to find the weakest link of the chain (section 4.3.5) which is in this case train operating company. DEA makes it possible to consider multidisciplinary inputs and outputs (section 4.3.4) to enable the stakeholder (government) make holistic decision making for capacity utilisation (4.3.7).

The results show where capacity waste can be decreased. Train operating companies that can generate higher passenger-km while avoiding delays are preferred otherwise there will be waste in capacity utilisation. The relative performance of the train operating companies (updated annually) can be used as a criterion for fair judgment of future bidding for running various routes. It also provides incentives to these companies to improve their performance.

Although part of the relative efficiency scores is due to good management or mismanagement of the company, the Tobit regression shows that part of efficiency and inefficiencies are due to characteristics of the route. The efficiency scores are positively correlated with serving London (i.e better routes hence train operating companies that serve London like South West Trains and Virgin Trains have higher efficiency scores). Offering regional services negatively impacts efficiency scores (such as First Scot and Arriva Trains Wales). When these two factors are combined it results in lowest efficiency score (Northern).

5.4 Measuring and Analysing Capacity Utilisation at Stations

Existing studies to improve train operations at stations have focused on: train routing through stations like the works by Zwaneveld et al. (1996), Kroon et al. (1997) and Zwaneveld et al. (2001); robust timetabling and train scheduling to minimise delays at staions such as the work by Carey and Crawford (2007), Yuan and Hansen (2007), Jianxin and Hansen (2007), Jia et al. (2009); combination of train routing and scheduling by Burkolter (2005) and Carey and Carville (2003) or analyses of station capacity utilisation by Lindfeldt (2007), Armstrong et al. (2011a) and Landex (2011). They all fall into one of the categories mentioned in Table 5-2 hence there is still no holistic approach to capacity utilisation analysis at stations especially at tactical levels. As stations are the bottlenecks of the railway network, it is very important to develop appropriate methods of measuring and analysing capacity utilisation at these points.

5.4.1 Intrinsic characteristics of the model

In the following sections the intrinsic characteristics of the model to analyse capacity utilisation at stations are summarised.

5.4.1.1 Stakeholder

Train stations in Great Britain are either run by a train operating company or Network Rail. Network Rail runs 17 stations which are the biggest and busiest ones. Direct stakeholder of capacity utilisation at stations is the station operator. However, as government pays franchise payments to the train operating companies and direct grants to Network Rail, in the big picture the stakeholder is the government. Improving capacity utilisation at stations would benefit them all but the inputs and outputs should be chosen from the eyes of overall stakeholder (government).

5.4.1.2 Controllable inputs

Capacity utilisation at stations is twofold: at macro level of trains (technical efficiency) and at micro level of passengers (service effectiveness). Due to limitation of train movements, the layout of stations has a great impact on technical efficiency. The main parameters of the station layout are the number of platforms, number of through/ending lines and length of platforms. In the tactical planning horizon, layout of stations can be changed if needed. For this end and to find the optimum values for these parameters, the input oriented option should be chosen when solving the model by the software. However, as changing the layout of stations is costly, the main aim analysis of capacity utilisation at stations is getting more outputs with the same inputs (output oriented model). The number of staff working at a station affects capacity utilisation but as the model is built from the perspective of the overall stakeholder, this parameter is not critical because the station operator adjusts staff according to the demand. Moreover data is not available on the number of staff at various stations (and the length of platforms). Number of train stops at each station is the main controllable input at stations for its service effectiveness.

5.4.1.3 Output priorities

For the technical efficiency of stations the main priority is the number of train stops (i.e. the more train stops that can be accommodated at a stations the better). At micro level of capacity utilisation the main priority is the number of passengers (service effectiveness).

5.4.2 Benchmarking from ports and airports

Different modes of transportation face capacity constraints at nodes (Table 5-8). The underlying concepts of nodal capacity constraints are rather similar for ports, airports and train stations as they are where different vehicles and routes merge and diverge.

Table 5-8 - Capacity Constraints for Different Modes of Transportation (Khadem Sameni et al., 2010a)

Mode of Transportation	Main infrastructure	Degrees of freedom for movement	Capacity of main infrastructure	Bottlenecks
Air	Air	3	Abundant	Airports
Marine	Water	2	Abundant	Ports/Locks
Road	Road	2	Limited	Junctions
Railway	Tracks	1	Limited	Stations / Junctions

Generalising the problem of operational efficiency at stations leads to operational efficiency at transportation nodes. There have been comprehensive studies on operational efficiency at ports and airports which makes it possible to benchmark and develop a methodology for station capacity analysis from them (Figure 5-13). The data envelopment analysis which has been extensively used for ports and airports can be adopted for railway stations.

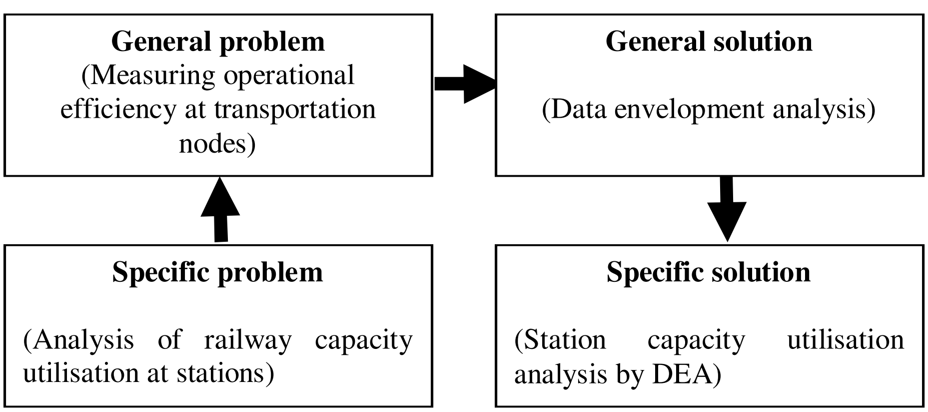


Figure 5-13- Innovative problem solving methodology applied to station capacity utilisation . Based on Rantanen and Domb (2007)

Table 5-9 - Major inputs and outputs for port efficiency analysis

	Range of inputs	**Range of outputs**
Port efficiency analysis by DEA as surveyed by Lozano et al. (2011)	• Number of workers • Book value of assets • Operating costs • Capital invested • Quay length • Terminal area • Number of (quay/yard) gantry cranes • Number of straddle carriers • Total berth length • Stocking area • Number of deep water piers • Number of tugs • Delay time • Annual expenditure on equipment	• Ship calls • Movement of freight • Total cargo/containers handled • Liquid bulk • Dry bulk • Number of ships • Number of passengers • Sales • Movement of containers/hour/ship

Table 5-10 - Major inputs and outputs for airport efficiency analysis

	Range of inputs	**Range of outputs**	**Type of efficiency**		
Airport efficiency studies as surveyed by Pestana Barros and Dieke (2007)	• Number of runways • Number of gates • Terminal area • Number of baggage collection belts • Number of public parking spots • Number of employees	• Number of passengers • Pounds[1] of cargo	Terminal		
	• Airport area • Number of runways • Runway area • Number of employees	• Air carrier movements • Commuter • movements		Movement	
	• Number of employees • Capital input estimated as an • Annual rental based on rate of return • Other inputs defined as the residual of total operating costs • Accumulated capital stock proxied by amortision • Intermediate expenses	• Turnover • Number of passengers • Cargo and mail business			General

[1] Weight unit

5.4.3 First Stage Model: Technical Efficiency of Stations

The main functions of railway stations as stated by Zemp et al. (2011) are linking catchment area and transport network, supporting transfer between modes of transport, facilitating commercial use of real estate, providing public space and contributing to the identity of the surrounding area. To this list 'facilitating railway operations' should be added. In the first stage model we want to analyse how well the existing capacity of the infrastructure is utilised at stations and how efficiently it is transformed into outputs of train stops. In the "definition of capacity" step (as presented in Figure 1-6), we define station capacity as "the ability of station infrastructure to accommodate necessary train services". This is in line with the definition of "macro capacity utilisation: Quantity of discrete steps to use railway capacity" as defined by Khadem Sameni et al (2011b). Hence, in the manner that Yu (2008) characterised "technical efficiency for railway companies", we define technical efficiency for stations as how efficiently infrastructure resources are utilised to accommodate train services.

The main infrastructure resource at stations (for passenger operation) is the number of and length of platforms. It is the equivalent of the number of quays for port efficiency analysis and the number of runways for airport efficiency analysis (Table 5-10). The number of platforms is usually less than or equal to the number of lines at the station. As a platform is needed for passenger trains to stop and for passengers to get on and off trains, we choose the number of platforms. As trains have one degree of freedom for movement along the track, the layout of stations is also very important for capacity utilisation. This concept does not exist for ports and airports as ships have two and planes have three degrees of freedom for movement resulting in more flexible operation. Through lines are more efficient for operation of trains than terminating lines. To represent the layout of infrastructure in the inputs, we suggest using the percentage of through lines which is calculated as:

$$Percentage\ of\ through\ \ lines\ at\ station = \frac{Total\ number\ of\ through\ lines}{Total\ number\ of\ through\ lines + Total\ number\ of\ ending\ lines}$$

The length of platforms (translating to quay length and runway area) can be added to the model as an input, but data on this item was not accessible for this case study. The number of staff at the station is an alternative input when the general technical efficiency of stations is to be assessed (and not the purely physical infrastructure capacity utilisation).

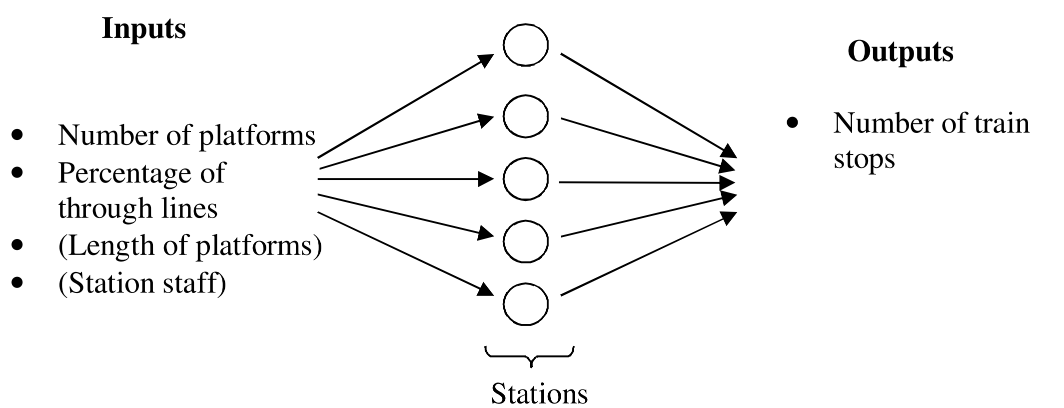


Figure 5-14 - Stage 1: Schematic representation of the technical efficiency model for train stations

We suggest that the output for the technical efficiency of the stations is the number of train stops at the station. These data can be extracted from the working timetable.

5.4.4 Second Stage Model: Service Effectiveness of the Stations

Service effectiveness as suggested by Yu (2008) for railway companies tries to estimate how effectively produced intermediate outputs are consumed. This is in line with "micro capacity utilisation" as defined by Khadem Sameni et al. (2011b). Stations receive different inputs. The second stage service effectiveness model takes the output of the first stage model (the number of train stops at the station) as one of its inputs. One of the main inputs is the number of trains that stop at that station (because clearly the trains that just pass through the station have no role in injecting passengers to the railway system from that station). There is an analogy between the 'number of trains that stop at a station' and 'the number of cranes' in the port efficiency analysis (Table 5-9). The more trains stop at a station, the more passengers can be 'lifted' from that station to increase the throughput of that station. No doubt, there is a logical limitation for the number of trains that stop at the station to be operationally efficient. Determining the optimum number of stops for a station is feasible with data envelopment analysis.

Another input for the stations is potential demand in the local population. It is not just the number of trains stopping at the station that affects the station's throughput: there should be passengers to get on the train or in other words the potential demand in the catchment area of the station. One of the best indicators of this as used in demand studies is the catchment area population and the number of jobs available in that area. To this end, the size of the population and the number of jobs available in the catchment area were chosen as inputs for the data envelopment analysis model. Such data can be extracted from the Geographic Information System (GIS).

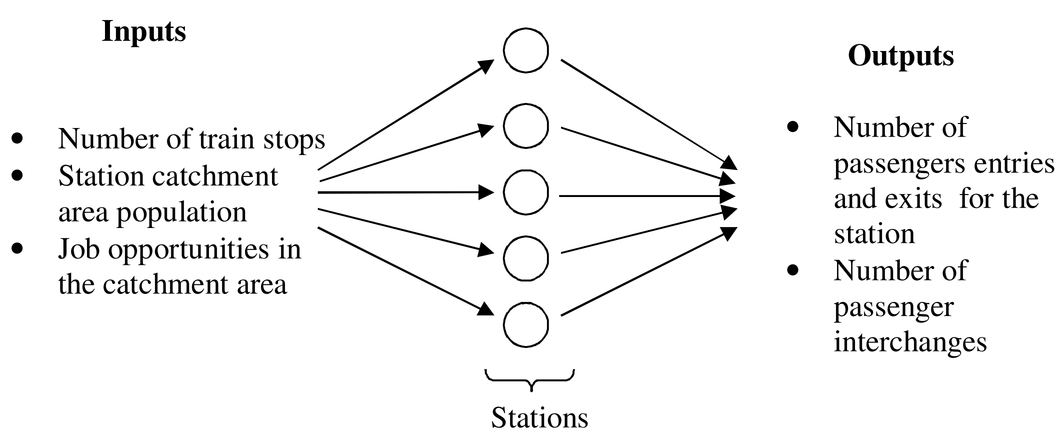


Figure 5-15 - Stage 2: Schematic representation of the service effectiveness of train stations

The outputs suggested are the number of passenger entries exits to the station (to reflect those passengers that start and end their journeys at that station) and another output as the number of passenger interchanges. These data are available in the comprehensive "station usage reports" produced by the Office of Rail Regulation in the UK or can be estimated through ticket sale statistics.

5.4.5 Case study: Busiest Train Stations In Great Britain

Train utilisation (passenger-km divided by train-km) in Great Britain is very low compared to other European railways and infrastructure capacity utilisation is below average (Civity Management Consultants, 2011). As Figure 5-16 shows, Great Britain's average train utilisation is lower than that of France, Sweden, the Netherlands and Switzerland. Average train utilisation is equivalent to 'micro capacity utilisation' and train frequency is equivalent to 'macro capacity utilisation' as defined in section 5.1.2 - Discrete nature of railway capacity. This indicates that micro capacity utilisation of trains are very low and there are some redundant or too frequent services.

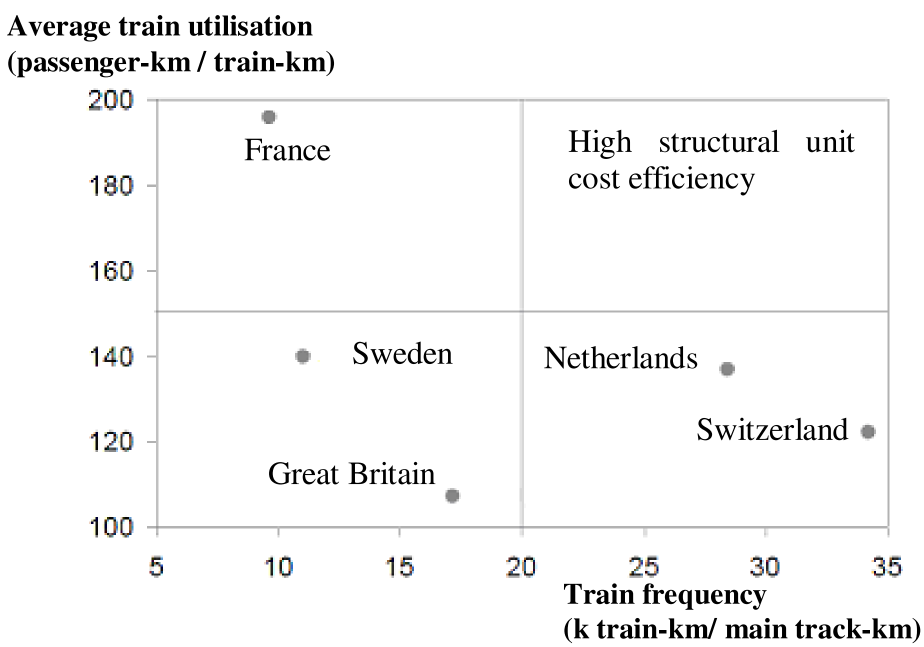


Figure 5-16 - Train utilisation versus infrastructure utilisation in five European railways (Civity Management Consultants, 2011)

The 120 busiest train stations in the UK were initially chosen in terms of passenger entries and exits to the station according to the station usage report (Office of Rail Regulation, 2008). Train frequency for the stations of this case study were extracted by using Perl scripts developed by Armstrong et al. (2009) from Common Interface Format (CIF) timetable files. The year chosen for the data sets was 2007.

For this study, the catchment area was considered to be within a 4-minute drive to the station. The data for catchment area population and job opportunities were extracted from a previous PhD thesis conducted at the University of Southampton (Blainey, 2009, Blainey, 2010) using GIS. The population and jobs figures in the catchment area of three stations (St Pancras, Blackfriars and Stansted Airport) are zero and these three stations were excluded from the analysis. These sets of job and population figures are based on output area zones – and the spatial sizes of these are in turn based on population density, so that all output areas have populations of approximately the same magnitude. Population density around each of these stations is low (and in the London cases station density is also extremely high), meaning that the output areas are large and the population-weighted centroid of the output area in which the stations are located is closer to a neighbouring station. In this regard these three stations effectively have no catchment when all-or-nothing allocation of output areas to stations is used. This is a general problem with defining catchments in this way in areas with a high employment density but low population density (Blainey, 2009). Excluding stations located in Scotland where job opportunities data was not available and also the stations that had zero catchment population at the centre of the output area, the total number of stations in the case study was narrowed down into 96.

The percentage of through lines were calculated manually by studying the station layouts available at http://www.nationalrail.co.uk/stations/. Descriptive statistics of the case study are presented in Table 5-11.

Table 5-11 Descriptive analysis of the station case study data

	Percentage of through lines	**Number of platforms**	**Number of trains with scheduled stop**	**Population**	**Jobs**	**Total entries and exits 07-08**	**Interchanges 07-08**
Average	0.70	6.20	382.56	20536.98	22231.18	10051630.54	1048559.68
SD	0.39	4.13	236.50	17104.52	22197.36	14042197.96	2091589.60
Min	0	2	113	108	13	2502752	116
Max	1	19	1357	98731	128595	91452130	17863239

5.4.6 Analysis of Results

The data envelopment analysis models were solved by PIM DEA-V3.0 software (Emrouznejad and Thanassoulis, 2011). Detailed efficiency scores and ranks are presented in Appendix 3.

The technical efficiency model was solved by variable return to scale assumption and by output orientation. The 7 efficient stations (i.e. efficiency score of 1) for technical efficiency or macro capacity utilisation are: London Waterloo, London Bridge, East Croydon, Clapham Junction, Moorgate, Liverpool Central and Balham. London Waterloo, the busiest station in Great Britain according to passenger entries and exits, is a terminal station (percentage of through lines is 0) and it handles a significant number of trains (1357 per day) with its 19 platforms. The efficiency score for London Victoria, London's second terminal station, which handles 1308 trains with its 19 platforms is 0.964 . Clapham Junction is Great Britain's busiest train station according to the number of trains. It handles 2039 train stops with its 16 platforms but all its lines are through lines. The lowest technical efficiency score (0.198) belongs to Newcastle train station which has 10 platforms, the percentage of through lines is 0.50 and it handles just 264 train stops. In other words, Newcastle has plenty of capacity to accommodate more train stops if necessary.

Out of the 96 stations, 15 stations are efficient in the service effectiveness model or micro capacity utilisation. By attracting passengers from potential demand in the catchment area, they efficiently transform train stops to passenger journeys represented by total entries and exits to the station and passenger interchanges between trains at that station. These stations are: London Waterloo, London Bridge, London Charing Cross, London Euston, London Kings Cross, East Croydon, London Cannon Street, Clapham Junction, Birmingham New Street, Moorgate, City Thames Link, Herne Hill, West Hampstead Thameslink, Southend Victoria and Tunbridge Wells.

The input-oriented service effectiveness model is helpful when it is intended to minimise input of train stops while keeping output levels. The output-oriented service effectiveness model is useful when the goal is maximising outputs of passenger entries and exits and passenger interchanges with the existing levels of inputs. Top and bottom stations of the service efficiency and service effectiveness models are presented in Table 5-12.

Table 5-12 Top and bottom stations of the service efficiency and service effectiveness models

Type of efficiency	Criteria	Name of the station	Score
Technical Efficiency	Top stations	London Waterloo	1.000
		London Bridge	1.000
		East Croydon	1.000
		Clapham Junction	1.000
		Moorgate	1.000
		Liverpool Central	1.000
		Balham	1.000
	Bottom	Newcastle	0.198

Type of efficiency	**Criteria**	**Name of the station**	**Score**
	stations	Milton Keynes Central	0.223
		Hither Green	0.237
		Southend Victoria	0.237
		York	0.250
Service effectiveness (Output oriented)	Top stations	London Waterloo	1.000
		London Bridge	1.000
		London Charing Cross	1.000
		London Euston	1.000
		London Kings Cross	1.000
		East Croydon	1.000
		London Cannon Street	1.000
		Clapham Junction	1.000
		Birmingham New Street	1.000
		Moorgate	1.000
		City Thameslink	1.000
		Herne Hill	1.000
		West Hampstead Thameslink	1.000
		Southend Victoria	1.000
		Tunbridge Wells	1.000
	Bottom stations	Manchester Victoria	0.152
		Barking	0.162
		Raynes Park	0.166
		Bedford	0.190
		Tottenham Hale	0.194
		Luton	0.195
Service effectiveness (Input oriented)	Top stations	Birmingham New Street	1.000
		City Thameslink	1.000
		Clapham Junction	1.000
		East Croydon	1.000
		Herne Hill	1.000
		London Bridge	1.000
		London Cannon Street	1.000
		London Charing Cross	1.000
		London Euston	1.000
		London Kings Cross	1.000
		London Waterloo	1.000
	Bottom stations	Richmond	0.316
		London Waterloo (East)	0.327
		Liverpool Central	0.349
		Cardiff Central	0.351
		Leeds	0.353

The mean for service effectiveness (output-oriented) scores is 0.662 and the mean for its input-oriented variant is 0.467. Many stations have unnecessary train stops which is reflected in a lower mean for the service effectiveness input-oriented model.

This means that many stations are not doing well for maximising outputs for the amount of inputs they receive. This is a waste of capacity. In other words, more passengers can be transported (i.e. passenger entries and exits and passenger interchanges) with the existing level of inputs. If it is intended to minimise inputs (the number of train stops at stations) without decreasing the output levels, the input-oriented service effectiveness should be used. The input-oriented technical efficiency model is not presented in the table as changing the number of platforms and percentage of through lines are not feasible in the tactical planning horizon.

Table 5-13 - Descriptive statistics of efficiency scores

	Mean	**SD**	**Min**	**Max**
Technical efficiency scores (output-oriented)	0.504	0.222	0.198	1.000
Service effectiveness scores (input-oriented)	0.467	0.277	0.152	1.000
Service effectiveness scores (output-oriented)	0.662	0.201	0.316	1.000

For technical efficiency, of 10 most efficient stations, six are in Central London and three in suburban London. For service effectiveness, 15 stations are located on the input oriented frontier, of these eight are in Central London and for in Suburban London. This might suggest that the results are largely due to railway geography and aside from findings on policy ownership, implication may be limited.

An important exception is Birmingham New Street at least in terms of service effectiveness. This might indicate that regional hubs should be considered elsewhere for example in Manchester.

5.4.7 Tobit Regression

Tobit regression was done for the service effectiveness model by SPSS V.19, by adding R and Python plug-ins and 'SPSSINC_TOBIT_REGR' application (IBM, 2011). Independent variables of interests are London location and operation by Network Rail (publicly operated). London area stations are identified according to London Travel Card zone. However, between Network rail operation and the binary variable of London stations there is Pearson Correlation of 0.622, significant at 0.01 levels. The results under Gaussian (normal) assumption are presented in Table 5-14. The results show that there is strong correlation between service effectiveness score and being located in London area.

VRS service effectiveness score= 0.396 + London location * 0.341

Table 5-14 Tobit regression results

	Coefficient	Std. Error	z Value	Sig.
(Intercept)	.396	.034	11.661	.000
LONDON	.341	.067	5.052	.000

5.4.8 Systems engineering and real world implications

In addition to some of the points mentioned in section 5.3.7 for the DEA model for passenger train operating companies, a holistic approach towards capacity utilisation at stations is needed. This is especially important as stations are bottlenecks of railway traffic flow hence the weakest link of the chain in the system (section 4.3.5). Due to the complexities involved, efficiency of services at stations is difficult to analyse by current methods such as operations research or simulation. As discussed in section 4.3.6, in these situations creative problem solving can help to improve the system performance. Hence by using innovative problem solving, a benchmarking was done from the DEA approaches taken for ports and airports to develop an appropriate model for railway stations.

Using DEA can help with the fragmentation of knowledge (section 4.3.4) in railways and for analysing capacity utilisation at stations as it accommodate the concerns of civil engineers (layout of stations), economists (number of passengers) and operation researchers (number of train stops) in one model.

The technical efficiency model showed that out of 96 stations in Great Britain, just 7 stations operated at their full relative macro capacity utilisation. Hence there is enough track capacity to hold more train stops at these 89 stations if needed.

The input oriented service effectiveness model can decrease unnecessary train stops (reduce macro waste) while keeping the same level of passengers. In this model stations that can accommodate more relative train stops with their available infrastructure are more efficient. The output oriented service effectiveness model can maximise the number of passengers (reduce micro waste) that can be handled while keeping the number of train stops constant. In this model the most efficient stations are the ones that are better at attracting potential demand from the catchment area. The Tobit regression results shows that being located in London greatly helps to improve the station's performance in attracting passengers.

5.5 Summary and Conclusions

This chapter focuses on developing a methodology for defining, measuring and analysing capacity utilisation in the passenger sector. The first step towards efficient management of railway capacity utilisation is defining it appropriately. In particular, the discrete nature of capacity utilisation is emphasised. This leads to considering two aspects of capacity utilisation. Macro capacity utilisation which is the *quantity* of discrete steps to use railway capacity (e.g. the number of trains) and micro capacity utilisation which is the *quality* of discrete steps to use railway capacity (e.g. load factor).

Based on the lean thinking concepts, lean capacity utilisation is defined as "The ability of infrastructure to generate added value by enabling passengers to reach their destination as planned". Therefore lean capacity utilisation is a function of micro capacity utilisation multiplied by macro capacity utilisation. It is a function of the number of passengers transported (micro capacity utilisation) multiplied by the distance they are transported in the unit of time (macro capacity utilisation that is determined by the speed of the train). This idea is the foundation of two novel methodologies developed for measuring and analysing capacity utilisation in the passenger sector for passenger operators and stations.

To choose an appropriate tool for measuring and analysing the concept of capacity utilisation, current approaches towards it and their strengths and weaknesses are compared. Data envelopment analysis which is predominantly used by economists to analyse value for money, efficiency and productivity of railways is for the first time bridged with engineering concerns to analyse efficiency in railway capacity utilisation. As the International Union of Railways (2004) has stated: "Capacity as such does not exist. Railway infrastructure capacity depends on the way it is utilised." Therefore instead of trying to directly measure capacity utilisation, in this thesis we suggest measuring the relative efficiency of units in capacity utilisation. This concept is illustrated in two case studies for passenger operators in Great Britain and for the 96 busiest stations. The major strength of DEA is that it can encompass the multidisciplinary and complex nature of railway capacity by having various inputs and outputs with different units without knowing their exact relationship. It is fast to solve due to its non-parametric and deterministic nature and can be a good tool for tactical planning of railways as is intended in this thesis. It can provide insights on the relative operational efficiency of units in transforming inputs to outputs.

To analyse how well different passenger-operating companies use railway capacity and provide added value, a DEA model was developed to assess their efficiency in transforming externally obtained inputs to valuable services. Franchise payment from the government was chosen as an input to reflect the amount of public resources that a train operating company is allocated. Timetabled train-kilometres were chosen as the other input to indicate how much capacity of infrastructure (as a public resource) the train operating company is using. Passenger-kilometres was chosen as one of the outputs as it is the best indicator of lean capacity utilisation. To consider the quality of services provided by the train operating company, delay-minutes was chosen as another output. It should be noted that delay-minutes is an undesirable effect therefore it cannot be used directly in the model. There are various techniques to handle this situation in DEA including using the inverse of such variables. A follow-up Tobit regression showed that efficiency scores are positively correlated with serving London and negatively correlated with average length of haul being less than 40 miles.

For analysing capacity utilisation at stations, a two-stage model was developed. The first stage model analyses the operational efficiency of stations to accommodate train stops (output) by the inputs they have received (the number of platforms and percentage of through lines). This covers the aspect of macro capacity utilisation. The second stage model analyses service effectiveness of stations to assess how effectively these train stops are transformed to passenger journeys (macro capacity utilisation). The inputs of the second stage model are the number of train stops, the catchment area population and job opportunities in the catchment area. The outputs are passenger entries and exits for that station and the number of passenger interchanges at that station. A follow-up Tobit regression model showed a strong positive correlation between service effectiveness scores and whether the stations are located in London.

6 Defining, Measuring and Analysing Capacity Utilisation in the Freight Sector

Transporting freight by railway improves sustainable transportation, reduces congestion on the roads, has lower CO_2 emissions, less impact on the environment and generates macro-economic advantages for societies (UIC, 2011a). Managing capacity utilisation in the freight sector is as important as in the passenger sector. In this chapter we try to develop a methodology for defining, measuring and analysing capacity utilisation in the DMAIC cycle for the freight sector. The first freight case study (section 6.4) was conducted during one month research visit to the Railroad Engineering Program, School of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign in 2010 and was presented in a paper by Khadem Sameni et al. (2011a).

6.1 Introduction

Freight transportation is an important functionality of railways and many industries are dependent on it. Although unlike road, railway cannot provide door-to-door transportation, it is energy- and cost-efficient for transporting heavy and bulky commodities as well as container in mass amounts and quantities. Table 6-1 shows the volume of freight traffic in different parts of the world.

Table 6-1- Freight Traffic Volume in the World: Tonne-kilometre (billion) (UIC, 2011c)

	2006	2007	2008	2009	2010	Change (2010/2009)%
Europe[1]	2,646.6	2,813.6	3,103.0	2,411.4	2,454.4	1.8%
Africa	138.4	135.4	134.6	137.1	139.2	1.6%
America	3,519.5	3,540.2	3,513.8	2,973.2	3,076.1	3.5%
Asia and Oceania	2,872.6	3,095.9	3,452.7	3,466.2	3,435.6	-0.9%
World (estimates)	9,177.1	9,585.1	10,204.1	8,987.9	9,105.4	1.3%

Efficient capacity utilisation of the infrastructure by freight trains is critical and it is necessary to develop an appropriate methodology for managing it.

6.2 Defining Profit-Generating Capacity

Lean capacity utilisation for freight operation can be defined in a similar way to lean capacity utilisation in the passenger sector (section 5.1.3). However, the concept of freight transportation is different from passenger transportation. Passenger transportation by railway is barely profitable and has similarities with service industries such as

[1] This includes Turkey and the Russian Federation.

healthcare and education which are necessary for the socio-economic welfare of societies and need governments' financial support. As discussed in section 5.2, the proposed approaches for capacity management in the passenger sector focus on producing more value for capacity utilisation. Freight transportation is profitable hence its most important aspect of capacity utilisation is the profit generated. Therefore, value can be measured and analysed by the amount of profit that is generated. This value and the profit is a function of tonnage and commodity type (micro capacity utilisation) and the distance it is transported in the unit of time (macro capacity utilisation which is dependent on train speed) as Figure 6-1 and the following formulae summarise:

$$Lean\ capacity\ utilisation = f(macro\ capacity\ utilisation \times micro\ capacity\ utilisation)$$

$$Lean\ capacity\ utilisation = f(t \times c \times d \times \tan(\alpha)) = f(t.c.S)$$

based on (Khadem Sameni et al., 2011b)

d: distance
t: tonnage of goods
c: commodity type

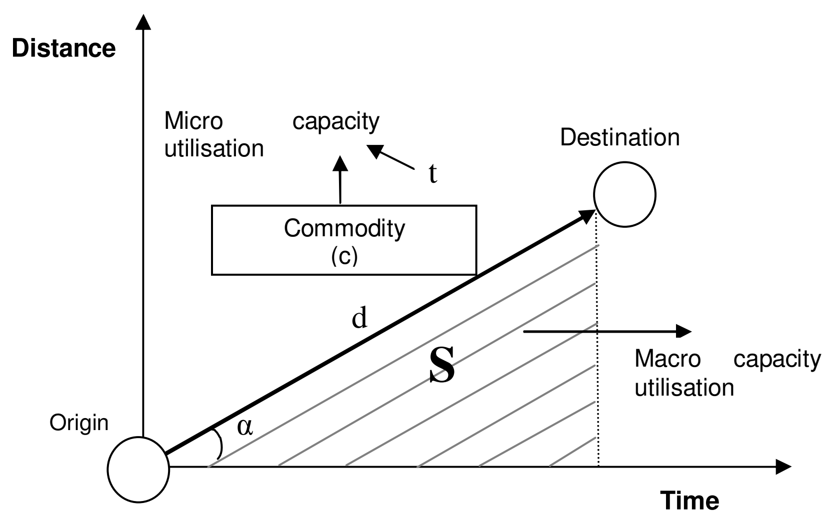


Figure 6-1 - Lean capacity utilisation for the freight sector based on (Khadem Sameni et al., 2011b)

It should be noted that the type of commodity is important in freight transportation as different commodities have different requirements (type of wagons needed, ease of handling, loading and unloading requirements, time sensitivity, etc.).

Profit-generating capacity utilisation is defined as "the ability of infrastructure to generate profit by enabling freight to move toward its destination". Contrary to other metrics of throughput (Table 4-1), this metric uses a currency unit to measure capacity utilisation.

6.3 Measuring and Analysing Profit-Generating Capacity Utilisation

The basic idea behind introducing the concept of profit-generating capacity utilisation is quantifying how well different scenarios of traffic utilise capacity by calculating the profit generated. Two approaches are introduced for measuring and analysing profit-generating capacity utilisation: direct (as in case study 1, section 6.4) and indirect (as in case study 2, section 6.5).

As the first approach measures the profit directly, it is necessary to estimate costs and revenues for each scenario of traffic. For each scenario, simulation software extracts congestion delays and total running time. Rail Traffic Controller (RTC), developed by Berkeley Simulation Software, is the primary simulation package used in class I railways[1] in North America. Based on this information, total rolling stock, crew and fuel costs can be calculated. After total costs have been deducted from total revenue, net profit can be estimated in each scenario and the scenario which makes the best utilisation of capacity can be identified. These steps can be summarised as:

1. Simulating traffic at different levels of traffic (number of trains) and heterogeneity (train commodity type)
2. Calculating total costs for different scenarios
3. Calculating total revenue for different scenarios
4. Calculating total profit for different scenarios
5. Choosing the optimum traffic combination

The second approach indirectly assesses the profit-generating capacity utilisation. Data envelopment analysis is used to compare the profit generated by different types of commodities based on their tonnage and the number of wagons loaded.

6.4 Case study 1[2]: bulk versus intermodal[3] traffic

In North America, freight railways (railroads[4]) own the infrastructure and can usually choose which trains to run on it. Different train types incur different costs and revenues. Currently there exists no appropriate methodology to advise railway authorities which type of freight train provides the maximum value for utilising the track capacity.

---

[1] "Class I Railroads are line haul freight railroads with 2009 operating revenue of $378.8 million or more". ASSOCIATION OF AMERICAN RAILROADS. 2011. *Class I Railroad Statistics* [Online]. Washington, DC. Available: www.aar.org/ [Accessed 06/09/2011.

[2] The raw data for this case study was provided by DINGLER, M. 2010. *Understanding the Impact of Operations and New Technologies on Railroad Capacity*. MSc., University of Illinois at Urbana-Champaign.

3 Intermodal is the common term in North America for container trains.

[4] In North America, the term 'railroad' is commonly used instead of railways and more specifically 'railroads' refer to freight railway companies. To preserve the consistency, throughout the present thesis 'railway' is used.

A wide range of commodities is transported by rail in North America, although coal is by far the most important (Figure 6-2). It is important for the railway owners to choose the combination of train types that offers the maximum value and profit. For example bulk trains are on average longer and slower but the costs of delays for intermodal trains are more than twice as much as bulk trains (Dingler et al., 2009a). The effects of different levels of heterogeneity and combinations of bulk and intermodal trains on delays has recently been studied by Dingler et al. (2009a). However considering delays alone is not enough, and total revenue, costs and profit should also be considered to assist the railway authorities to choose the best combination of traffic.

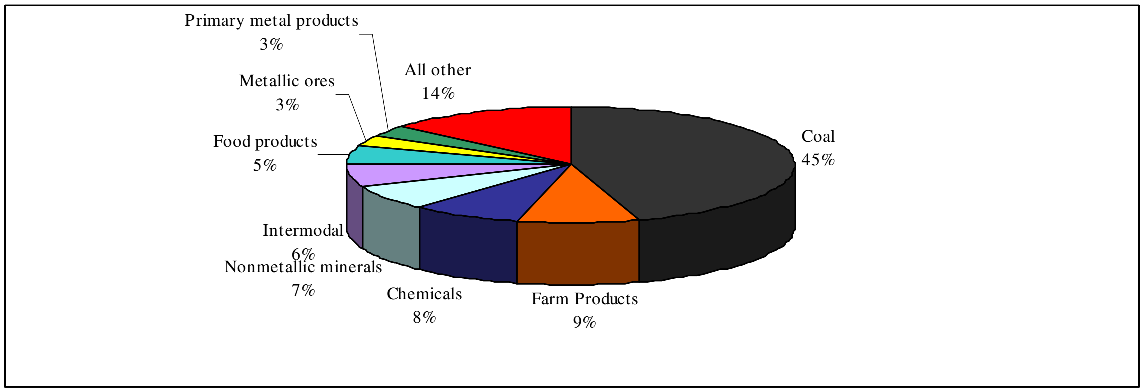


Figure 6-2 - Railway commodity types in the US based on tons[1] originated (Association of American Railroads, 2010)

Bulk and intermodal traffic account for roughly 60% of the American railroad's revenue, 75% of the tonnage and 80% of the wagon (car in American railway terminology) load (Association of American Railroads 2008). However, they utilise the track capacity in different ways and incur different costs and revenue. Existing metrics of capacity utilisation and their strength and weaknesses were discussed previously in Table 4-1. In this case study, the concept of profit-generating capacity utilisation is used to identify which combination of bulk and intermodal trains generates the maximum profit for the railway.Profit provides a better metric for capacity utilisation analysis than current metrics by:

- Using currency as the unit for capacity utilisation, which is in line with the operational goal of freight railway companies
- Considering different types of trains and their values
- Seeing the big picture of using the infrastructure
- Capturing the complex nature of railway capacity utilisation more
- Enabling more efficient decision-making for getting the maximum value

Simulations data was obtained using Rail Traffic Controller (RTC) developed by Berkeley Simulation Software at different levels of traffic from 8 to 48 trains as well as

[1] In American railroads, 'ton' is used as the unit for weight measurements. It equals 2000 pounds or 907.2 kilograms. Tonne, metric tonne and metric ton are all the same and equal 1000 kilograms. 'Tonne' is used in the statistics presented by the International Union of Railways.

for 0%, 12.5%, 25%, 50%, 75%, 870..5% and 100% of each train type. Train composition characteristics are described in Table 6-2. They were equally distributed over a 24 hour period.

Table 6-2- Train composition characteristics in the simulation (Khadem Sameni et al., 2011a)

	Intermodal	Bulk
Wagon (car) combination	163-pack spine cars[1] 95-pack well cars[2]	115 loaded hopper cars
Length of train	5,659 ft	6,325 ft
Tonnage	5,900 tons	16,445 tons
Horse Power per Trailing Ton (HPTT)	3.64	0.78
Locomotives	5*4,300 HP	3*4,300 HP
Maximum Speed	70 mph	50 mph

The track chosen was a single-track mainline subdivision with the following attributes:

- 262 miles long
- 10 miles between siding centres
- 8,700 ft signalled sidings with 24 powered turnouts
- 2.75 mile signal spacing
- 2-block, 3-aspect signalling
- 0% grade and curvature

6.4.1 Calculating total costs

The general workflow of calculating total costs is shown in Figure 6-3.

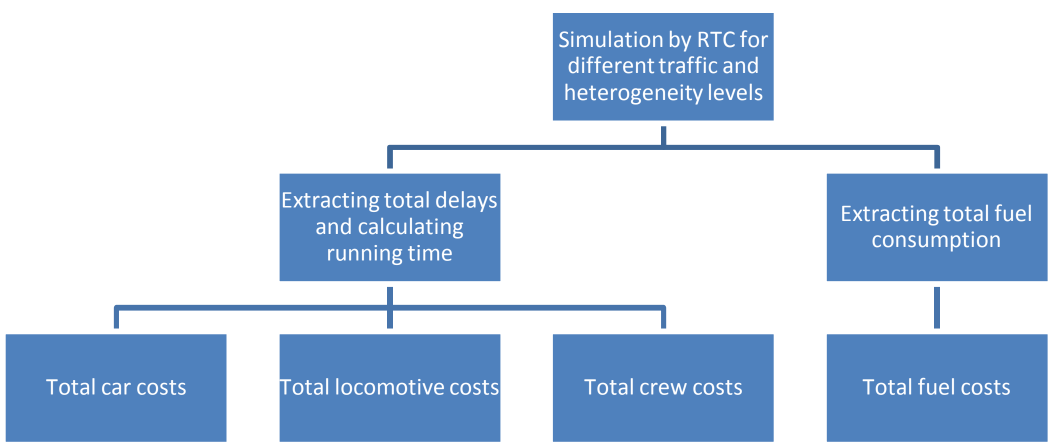


Figure 6-3 - General workflow of calculating total costs (Khadem Sameni et al., 2011a)

1 Pack spine cars hold 1 trailer.
2 Pack well cars hold 2 containers.

Total costs of running freight traffic on a line can be simplified as:

$$TC = a * VM + b * VH + c * V + d * RM$$

TC: total costs
VM: vehicle-mile
VH: vehicle-hour
V: vehicle
RM: route-mile
a,b,c,d : parameters

In this study we considered major costs including fuel, locomotive, car and crew costs. The values used in the study are calculated as estimated by Dingler (2010) in Table 6-3.

Table 6-3 - Major direct costs of running freight trains (Dingler, 2010)

		Intermodal	Bulk
	Avg. Cost per Car Hour	$1.00	$0.58
	Avg. Cars per Train	84.9	99.2
Car Cost per Train-Hour		$84.90	$57.54
		Intermodal	Bulk
	Cost for new locomotive	$1,750,000	$1,750,000
	Economic Life	25	25
	Discount Rate	10%	10%
	Salvage Value	$200,000	$200,000
	Units per Train	5	3
Locomotive Cost per Train-Hour		$111.20	$66.72
		Intermodal	Bulk
	Idling Fuel Consumption/Hr	3.5	3.5
	Avg. Fuel Cost/Gallon	$3.13	$3.13
	Avg. Units per Train	5	3
Fuel Cost per Train-Hour		$54.78	$32.87
		Intermodal	Bulk
	Crew Members per train	2	2
	Average Hourly Pay	$24.68	$24.68
	Fringe Benefits	35%	35%
Crew Cost per Train-Hour		$66.64	$66.64

6.4.2 Calculating revenue

With the tonnage of trains, revenue can be simply calculated. However, two important aspects of revenue should be considered. One is checking the elasticity of revenue to train frequency, i.e. to check if revenue changes as the frequency of trains increase (section 6.4.2.1). The second issue for the profit-generating capacity utilisation is to consider empty return ration (section 6.4.2.2) to have a fair comparison between bulk and intermodal trains.

6.4.2.1 Elasticity of revenue to train frequency

Total revenue should be calculated for each level of traffic and heterogeneity. Due to the different nature and sources of traffic, bulk and intermodal trains are considered to be independent from each other, hence revenue is independent of heterogeneity level. But it should be investigated whether or not the revenue is elastic with regard to train frequency[1]. The author could not find any references about elasticity of revenue to train frequency in the literature. Therefore it is deduced by price elasticity from the following formula:

Frequency elasticity = Price Elasticity × Value of frequency × (Price/Frequency)

$$Value of\ frequency = \frac{\frac{\partial U}{\partial F}}{\frac{\partial U}{\partial P}} = \frac{\partial P}{\partial F}$$

$$\frac{\partial Q}{\partial F}\frac{F}{Q} = \frac{\partial Q}{\partial P}\frac{P}{Q} \times \frac{\partial P}{\partial F}\frac{F}{P}$$

F: number of trains (intermodal/bulk) per day
Q: demand (number of containers)
P: price (dollar per container/car)
U: utility

Rail price elasticity for intermodal trains (nondurable manufactures) is assumed to be (Friedlaender and Spady, 1981). Studies such as Zhong (2007) have been unable to find statistically significant frequency parameters; based on his work value of frequency for intermodal trains ($\frac{\partial P}{\partial F}$) is calculated as 0.157 per container per departure per day. The exchange rate of pound to dollar is taken as 2 for 2007 (the year of that study). The average revenue per container and the average distance for intermodal trains (extracted from Class 1 Railroad statistics (Association of American Railroads 2008)), yields the average freight rate (P) for a container for 100 miles as \$112 per container

At traffic level of 28 intermodal trains per day, elasticity of demand with respect to train frequency is:

[1] Elasticity can be used only for small changes.

$$\frac{\partial Q}{\partial F}\frac{F}{Q} = -0.6*(0.157*1/2)*(28/112) = -0.0118 \approx 0$$

In this regard, intermodal demand is considered to be relatively inelastic to train frequency. It was not possible to calculate elasticity of bulk trains with regard to train frequency as no reference in the literature was found regarding value of frequency for bulk trains. However, bulk (coal) trains are less time-sensitive compared to intermodal trains, therefore their elasticity to train frequency must be less than intermodal trains. Hence, it can be inferred that revenue from bulk trains is inelastic to train frequency as well.

6.4.2.2 Considering Empty Return Ratio

The average revenue per loaded train for bulk trains is more than for intermodal trains. However it should be noted that bulk trains have a higher Empty Return Ratio[1]. This fact should be considered when calculating total revenue for different combinations of trains to reflect their real value. This is because the track capacity is wasted (muda as discussed in section 5.1.4 - Sources of practical capacity waste) when trains are hauled empty. By considering the Empty Return Ratio and average length of haul, revenue is adjusted for each type of train for 100 miles as in Table 6-4.

Table 6-4 - Adjusted average revenue per type of train (Khadem Sameni et al., 2011a)

	Bulk Trains	**Intermodal Trains**	**Source**
Average revenue per train	$203,182	$128,246	(Association of American Railroads 2008)
Average length of haul (mile)	707	828	(Bureau of Transportation Statistics, 2007)
Empty Return Ratio	2.03	1.13	(Cambridge Systematics, 2007)
Adjusted revenue per train (100 miles)	$14,157	$13,707	

Based on the simulation results from RTC software, total costs, revenue and profit were calculated for 77 different scenarios of traffic level (total number of trains) and heterogeneity (percentage of bulk/intermodal trains). Figure 6-4 shows how total profit varies against different traffic and heterogeneity levels.

---

[1] Empty Return Ratio is defined as total miles divided by loaded miles.2.Cambridge Systematics, *National rail freight infrastructure capacity and investment study*. 2007, Association of American Railroads: Cambridge , USA.

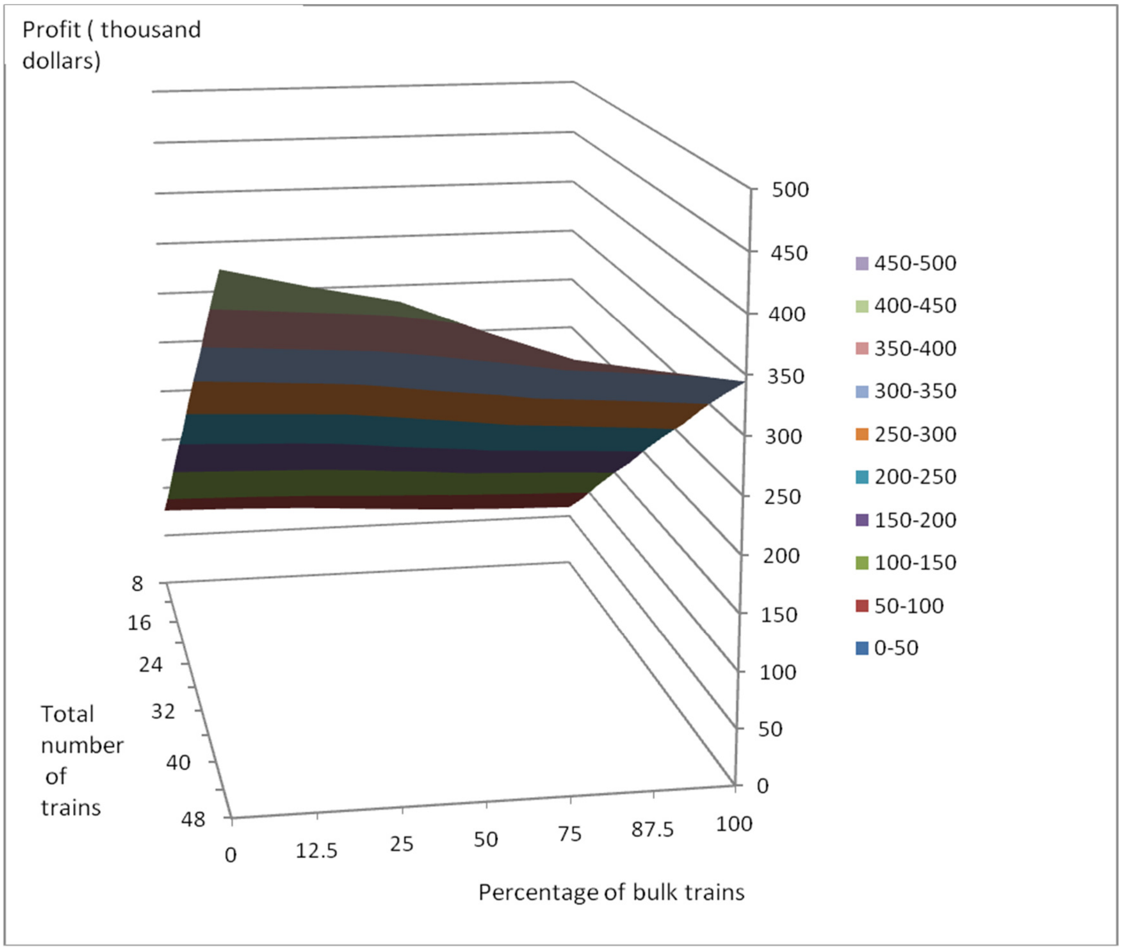


Figure 6-4 - Total profit against different heterogeneity and traffic levels

6.4.3 Analysis of results

The following results can be concluded from the profit-generating case study:

- When revenue is adjusted by the Empty Return Ratio and average length of haul, revenue per bulk train and per intermodal train are very similar (bulk train yielding slightly more revenue). Within the same level of traffic, total profit (total revenue minus total costs) increases as the percentage of intermodal trains increases and the percentage of bulk train decreases. This leads to a better utilisation of track capacity.
- There is a significant increase in total delay and total costs between 25% and 75% of heterogeneity.
- Static costs of delay (eg. $1,392 per hour for intermodal trains and $586 for bulk trains (Dingler et al., 2009b, Dingler, 2010)) are negligible compared to revenues from extra trains. Therefore total profit increases as the number of train increases. Dynamic costs of delays that vary according to the level of traffic better reflect the consequences of adding extra trains.

- Within the same level of traffic, total profit generated can be a good indication of how well the infrastructure is utilised hence it can be used for finding the optimum heterogeneity level. However, further research is needed for finding the optimum level of traffic (total trains per day) as currently in the literature costs of delays are mainly considered as static which falls far below the added revenue of an extra train. For tackling this issue, maximum allowable delay, inventory costs, yard time and added cycle time should be considered to calculate dynamic costs of delay which vary according to the traffic level.

The concept of revenue-generating capacity utilisation can be developed further by conducting a sensitivity analysis to discover whether costs of delays are based on "the value of time per hour per ton of shipment" (De Jong, 2000) which may result in higher costs of delays.

Other ways of extending the methodology include:

- Considering yard times in total costs and profits and how different levels of traffic affect yard times and inventory costs.
- Research on dynamic costs of delays according to the level of traffic.
- Calculating the value of frequency for bulk and intermodal trains in a freight railroad.
- Considering maximum allowable delay and dedicating infinity costs to the delays more than the set limit.

The concept of profit-generating capacity utilisation is very practical for railways and parts of the above-mentioned work and suggestions (as appeared in the paper by Khadem Sameni et al. (2011a)) are being incorporated in practice in a project to optimise freight traffic on the Portuguese network for freight trains (CP Carga) entitled "A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network"(Marinov and Viegas).

6.5 Case Study 2: Identifying the most profitable Commodities

In the previous case study, profit-generating capacity utilisation was directly calculated. Another approach to this concept is using data envelopment analysis. DEA does not need the exact relationship between the outputs and inputs (i.e. cost function, etc.) A DEA model can be used to analyse how efficiently transporting different types of commodities turn inputs such as tons originated and wagon loaded into revenue. Characteristics of a suggested model is summarised below.

6.5.1 Intrinsic characteristics of the model

To analyse profit-generating capacity utilisation for a freight railroad, the stakeholder, controllable inputs and output priorities are summarised below.

6.5.1.1 Stakeholder

As Class I railroads in the US are privately owned and operated the stakeholder for analysing freight capacity utilisation is the railroad company itself.

6.5.1.2 Controllable inputs

Type and volume of each commodity to be transported are controllable inputs for the railroad company. Depending on the circumstances and existing demand, it is railroad's decision to choose which commodities to accept and how much. This affects the costs incurred and the revenue generated.

6.5.1.3 Output priorities

For Class I railroads which are privately owned companies, the revenue generated is the main priority.

6.5.2 Choosing inputs and outputs for the model

The number of wagons loaded represents the wagon costs as well as labour costs involved with the train operations (loading, unloading, train formation, shunting at yards, etc.). Fuel consumption is also proportional to the tons originated. Using the DEA model to consider these inputs for different types of commodities and analyse the revenue generated provides insights about which commodities are more profitable for the freight railway to transport.

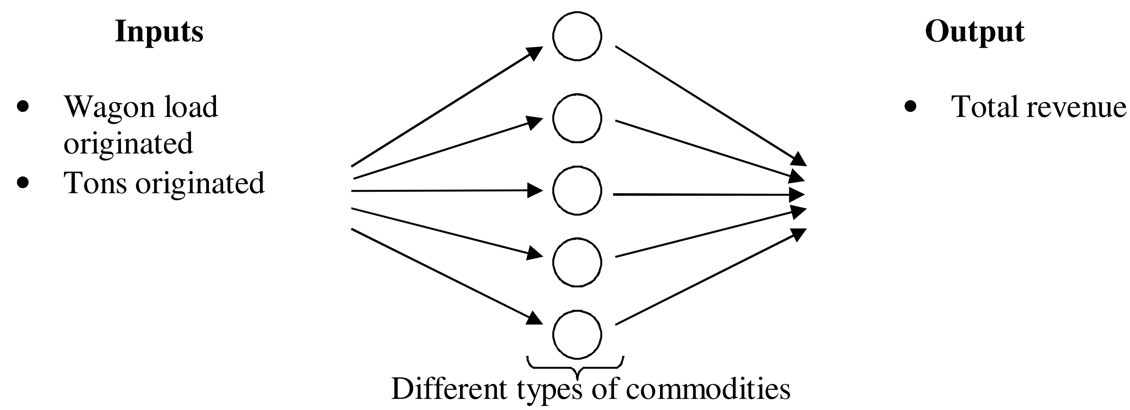


Figure 6-5- Schematic DEA model for the profit generating freight capacity

Data used for this case study is extracted from the "Analysis of Class I Railroads" published by Association of American Railroads (2008) which provides detailed data on every aspect of railway operation by class I railways. The model was solved using PIM DEA-V3.0 software (Emrouznejad and Thanassoulis, 2011). DEA efficiency scores are presented in Table 6-5.

Table 6-5 - Efficiency scores for different types of commodities

Name (Association of American Railroads, 2008)	**Output oriented VRS Efficiency**	**Rank**	**Input oriented VRS Efficiency**	**Rank**
Grain (Including Soybeans)	0.85	10	0.83	10
Other Farm Products	1.00	1	1.00	1
Metallic Ores	0.19	20	0.24	19
Coal	1.00	1	1.00	1
Crushed Stone, Gravel and Sand	0.33	17	0.23	20
Non-Metallic Minerals	0.21	19	0.46	17
Grain Mill Products	0.64	15	0.48	16
Food and Kindred Products	0.95	9	0.93	9
Primary Forest Products	1.00	1	1.00	1
Lumber and Wood Products	1.00	1	1.00	1
Pulp, Paper and Allied Products	0.74	12	0.61	14
Chemicals and Allied Products	1.00	1	1.00	1
Petroleum Products	0.72	13	0.64	13
Stone, Clay and Glass Products	0.67	14	0.52	15
Coke	0.31	18	0.74	12
Metals and Products	0.82	11	0.75	11
Motor Vehicles and Equipment	1.00	1	1.00	1
Waste and Scrap Materials	0.40	16	0.30	18
Forwarder and Shipper Association	1.00	1	1.00	1
All Other[1]	1.00	1	1.00	1

From the results it can be inferred that out of 20 categories of commodities, 8 are the most efficient ones to be transported by (American class I) railways: farm products (other than grains), coal, primary forest products, lumber and wood products, chemicals and allied products, motor vehicles and equipment, forwarding and shipping and container traffic.

This information can be used for railway authorities for allocating the track capacity and trains to commodities that generate the most revenue and profit. Hence "lean capacity utilisation" (section 5.1.3) and avoiding "capacity waste" (section 5.1.4).

[1] 95 percent of the "all other" category is intermodal (container) traffic. ASSOCIATION OF AMERICAN RAILROADS 2008. Analysis of Class I Railroads Washington, DC.

6.5.3 Systems engineering and real world implications

As explained in section 4.3.1, a holistic measure is needed for analysing capacity utilisation. For the case of privately owned freight railroads in the US, profit can be a good indicator of how well capacity is utilised. This is especially useful as these companies are vertically merged (operation and infrastructure management is merged as explained in section 3.7.1. In these companies there is less segmentation in the system (section 4.3.2) and all major concerns of the railroad company in terms of rolling stock, operations and track utilisation all can be reflected by costs and profit. However, the direct approach of estimating costs, revenue and profit (model proposed in section 6.4) posed some difficulty. Hence creative problem solving is needed in the system (section 4.3.6) and DEA can be used as a new way of analysing capacity utilisation for the freight sector. It is more powerful than the previously mentioned model as it enables holistic decision making (section 4.3.7) by considering all commodities and seeing the bigger picture of profit. Using this model helps the railroad to increase their efficiency in a holistic manner (section 4.3.1) by finding the commodities that are less profitable (weakest links of the chain as discussed in section 4.3.5). Annual analysis can confirm whether profitability ranking for different commodities remains the same or changes over time.

6.6 Summary and conclusions

In this chapter the concept of profit-generating capacity utilisation is suggested for better capacity management in the freight sector. Therefore the profit generated can indicate how well the track capacity is utilised. The concept is illustrated in two case studies based on Class I railways in America (which are vertically seperated railways where infrastructure and train operations are integrated). The first case study uses the concept of profit generating capacity utilisation directly by estimating cost and revenue for each scenario of traffic. It is applied to a case study of heterogeneous traffic of bulk and intermodal trains. By means of simulation data, total running time and delays are extracted. Costs associated with wagons, locomotives, fuel and crew are estimated. Revenue is calculated and adjusted by considering the average revenue per train, the average length of haul and the empty return ratio of wagons.

The second case study uses the concept of profit-generating capacity utilisation indirectly by using data envelopment analysis. DEA can be a helpful tool as it does not need the explicit mathematical relationship between the outputs and inputs (cost and revenue functions in this case). For the DEA model, different commodities are considered as different decision-making units (DMUs). It is suggested that wagon load originated and tons originated are used as inputs and revenue as the output for each type of commodity. Solving the DEA model can suggest which types of commodities are the most efficient ones for capacity utilisation. Based on the case study for the 20 types of commodities transported by Class I railways in America, the most efficient ones (efficiency score 1) are: farm products (other than grains), coal, primary forest products, lumber and wood products, chemicals and allied products, motor vehicles and equipment, forwarder and shipping and container traffic.

The concept of profit-generating capacity utilisation can provide helpful insights for railway authorities to use the track capacity more efficiently and avoid waste of capacity as far as possible.

7 Improving and Controlling Capacity Utilisation

This chapter addresses the last two steps in the DMAIC cycle, improving (I) and controlling (C). To *control* railway capacity utilisation two methods that are used in Six Sigma practices are adopted: one is variation reduction, and the other is failure mode and effect analysis (FMEA). To *improve* capacity utilisation, various methods are used to identify the weakest line section, trains and stations. These methods are illustrated in a real world case study.

7.1 Controlling Capacity Utilisation

In the following sections, two of the tools and methodologies used in Six Sigma for improving service quality are explained and discussed for the case of railway capacity utilisation. These are variation reduction and failure mode and effect analysis.

7.1.1 Variation reduction

Variance reduction or reducing faulty parts or services is the main aim of Sigma as discussed in section 4.5.2. The problem with the Public Performance Measure (PPM) used in the UK is that it does not consider variance in services and punctuality as suggested by Six Sigma. It just measures the arithmetic mean of punctuality (and reliability). Punctuality is the first priority of passengers (Figure 7-1), hence a major goal of railways and government (Public Performance Measure). It is one of the four major factors in capacity utilisation as seen in the capacity utilisation balance (Figure 2-1) and discussed in section 3.1.2 - Stability and reliability.

Passenger priorities - Factors	All	Long distance	Regional	London and SE	Commuter	Business	Leisure
	rank	rank	rank	rank	rank	rank	rank
Punctuality/reliability	1	1	1	1	1	1	1
Value for money for the price of ticket	2	2	2	2	2	2	2
Frequency of trains on route	3	3	3	3	3	3	6
Provision of information about train times/platforms	4	4	4	5	4	5	4

Figure 7-1- Passenger Priority Research conducted by the former Strategic Railway Authority in the UK (Network Rail, 2006b)

Figure 7-2 shows a train operator with a low variance in the Public Performance Measure whereas Figure 7-3 shows a train operator with a high variance. A low variance is desirable: if trains run on time, capacity utilisation is positively affected, where as late trains cause a domino effect on the network, disrupt the operation of other trains and thus waste capacity. High variance in punctuality adversely affects capacity utilisation. It is worth mentioning again that punctuality is train running on time (lateness) and reliability is whether trains are running at all (cancellation). The Public Performance Measure does not consider how late the trains are.

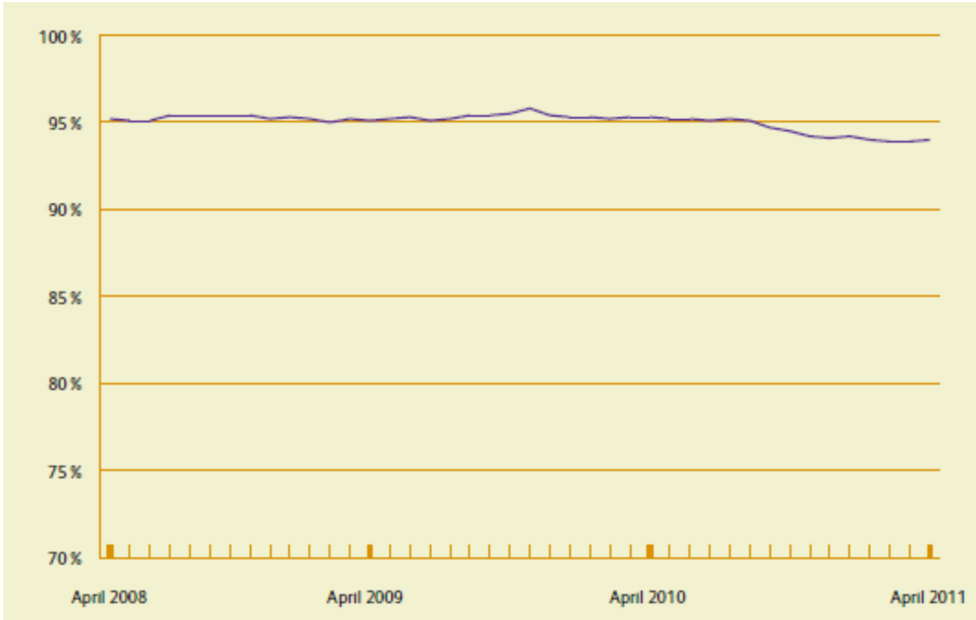


Figure 7-2 - All-day public performance measure of Chiltern Railways (low variance) (Network Rail, 2011d)

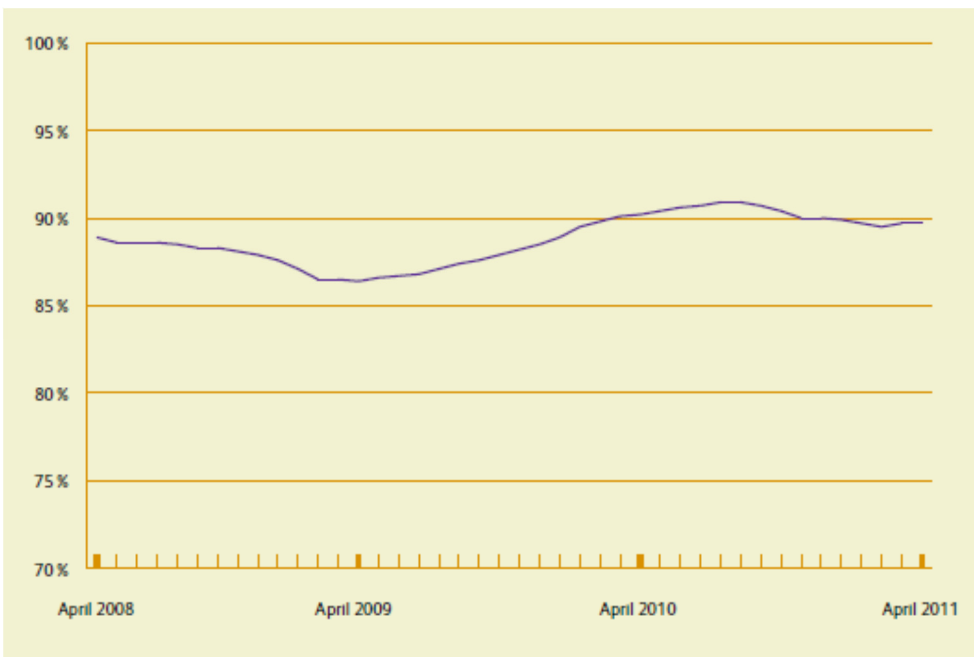


Figure 7-3 - All-day public performance measure of London Midland Railway (high variance) (Network Rail, 2011d)

In order to improve the control of capacity utilisation, the author suggests adding an element of controlling variation of the Public Performance Measure and including it in the franchise contracts. Such an index can be called the Public Performance Variation" or PPV and calculated as the following:

$$PPV = \sqrt{\frac{\sum_{for\,day=1\,to\,n} (PPM_{year} - PPM_{day})^2}{n}}$$

Therefore, the higher the PPM and the lower the PPV, the more desirable is the performance of the passenger operator and the more efficiently the track capacity is utilised. This measure can also be used for different routes.

7.1.2 Failure mode and effect analysis (FMEA)

Reliability or stability is one of the pillars of the capacity utilisation balance (Figure 2-1). Therefore whatever jeopardises the reliability and stability of train services must be controlled as far as possible. Failure mode and effect analysis (FMEA) identifies and prioritises risks in a systematic way to eliminate and reduce potential risks of system failures (Stamatis, 2003). FMEA is one of the major tools used in the control step of the DMAIC cycle (Pyzdek, 2003). It can be a useful tool for controlling the delays which have a severe impact on railway capacity utilisation.

FMEA estimates risk priority numbers (RPNs) by considering occurrence (O), severity (S) and detection (D). A ranking system suited to the particular circumstances of the industry or organisation is developed, with a scale of 1-10 for occurrence, severity and detection categories with 1 being the lowest and 10 the highest. As the formula below shows, every risk receives a risk priority number that is the product of those three numbers which can vary between 1 to 1000. The higher the risk priority number, the more attention and control that risk needs (Stamatis, 2003).

$$RPN = O \times S \times D$$
$$O : Occurance\ rating$$
$$S : Severity\ rating$$
$$D : Detection\ rating$$

The Office of Rail Regulation (2010b) lists 19 causes of infrastructure failure in the railway network in the UK, with their occurrence rates and severity. These failures - track faults, signal failures, telecommunication failures, etc.- occur at different rates, with different severity and consequences. Some, such as bridge strikes are relatively easy to prevent and detect whereas others, such as signal failures, are more difficult to prevent. Risks can be prioritised by FMEA to provide insights for managing the infrastructure more efficiently.

Based on the causes of infrastructure failure in the railway network in the UK (Office of Rail Regulation, 2010b), their occurrence rates and severity, FMEA tables are developed as seen in Table 7-1 for occurrence evaluation, Table 7-2 for severity evaluation and Table 7-3 for detection evaluation.

The occurrence rate varies for the different risks affecting infrastructure failures, ranging from gauge corner cracking with the lowest number of incidents (66 in 2010-2011) to point failures with the highest number of incidents (5,802 in 2010-2011) (Office of Rail Regulation, 2011b). Table 7-1 develops an occurrence table for categorising infrastructure failures according to FMEA.

Table 7-1 - FMEA occurrence evaluation of infrastructure failure. Based on (Chin et al., 2009, Xu et al., 2002)

		Likely failures rate (Average per day in the Great Britain's network)	Ranking
Very high	Persistent failures	40 or more	10
		20	9
High	Frequent failures	10	8
		5	7
Moderate	Occasional failures	2	6
		1	5
Low	Relatively few failures	0.5	4
		0.2	3
Remote	Failure is unlikely	0.1	2
		0.05	1

The severity of infrastructure failures also varies. Some failures can cause enormous disruption to the network while some others cause minor delays. The shortest average delay per incident is for track patrols and related possessions (14.6 minutes in 2010-2011) and the longest delay per incident is for cable faults (270.0 minutes in 2010-2011) (Office of Rail Regulation, 2011b). A severity evaluation table is developed as in Table 7-2.

Table 7-2 - FMEA Severity evaluation of infrastructure failure (Chin et al., 2009, Xu et al., 2002)

Severity evaluation criteria	Disruption to network	Average delay minutes per incident	Ranking
Extremely disrupting	More than 160 minutes		10
Very disrupting	Up to 160 minutes		9
Very high	Up to 80 minutes		8
High	Up to 40 minutes		7

Moderate	Up to 20 minutes	6
Low	Up to 10 minutes	5
Very low	Up to 5 minutes	4
Minor	Up to 2 minutes	3
Very minor	Up to 1 minute	2
None	No discernible effect on the network	1

Detection evaluation in FMEA is a rating indicating the likelihood that the system controls the root cause of a failure mode and prevents it before it affects the customer (Stamatis, 2003). Table 7-3 develops a detection rating for railway infrastructure failures based on the FMEA concepts.

Table 7-3 - FMEA detection evaluation of infrastructure failure (Chin et al., 2009, Xu et al., 2002)

Detection	Description	Ranking
Not detectable	The risk is not detectable by existing control mechanisms in the system.	10
Almost undetectable	The risk is almost undetectable by existing control mechanisms in the system.	9
Very low	There is a very low chance that the risk is detected by existing control mechanisms in the system.	8
low	There is low chance that the risk is detected by existing control mechanisms in the system.	7
Moderately low	There is moderately low chance that the risk is detected by existing control mechanisms in the system.	6
moderate	There is 50-50 chance that the risk is detected by existing control mechanisms in the system.	5
Moderately high	There is moderately high chance that the risk is detected by existing control mechanisms in the system.	4

Detection	Description	Ranking
High	There is high chance that the risk is detected by existing control mechanisms in the system.	3
Very high	There is very high chance that the risk is detected by existing control mechanisms in the system.	2
Definitely detectable	The risk is definitely detectable by existing control mechanisms in the system.	1

Based on the available data on different causes of infrastructure failures (Office of Rail Regulation, 2011b) and the above tables, risk priority numbers are calculated as in Table 7-4.

Table 7-4 - Risk priority numbers for infrastructure failures causes

Risk Item (Office of Rail Regulation, 2011b)	Occurrence	Severity	Detection	RPN
Temporary speed reduction due to condition of track	5	8	3	120
Track fault (including broken rails)	7	9	3	189
Gauge corner cracking	1	10	4	40
Points failures	7	8	3	168
Level Crossing failures	6	7	3	126
Overhead line equipments/third rail faults	5	10	4	200
Signal failures	7	7	4	196
Track circuit failure	7	9	4	252
Axle counter failure	4	8	4	128
Signalling system and power supply failures	7	8	4	224
Other Signal equipment failures	5	7	4	140
Telecoms failures	5	7	4	140
Cable faults (signalling and communication)	4	10	3	120
Civil Engineering structures, earthworks and buildings	1	10	2	20
Other infrastructures	6	7	2	84
Track patrols and related possessions	6	5	1	30
Mishap-infrastructure causes	5	8	3	120
Fires starting on network rail infrastructure	3	9	2	54
Bridge strikes	5	9	3	135

According to FMEA, the following risks must be addressed as the highest priority:

- The highest RPN
- The highest occurrence
- The highest severity (Stamatis, 2003)

Table 7-4 shows that track circuit failures score the highest RPN, while the highest occurrence and severity belong to point failures and cable faults respectively. Therefore these are the three areas which the infrastructure authorities must prioritise in order to control and reduce the risk of failures which interrupt the system, and thus to control capacity utilisation and manage it more efficiently. Some of the general recommendations of FMEA for reducing risks of failures are "adding built-in detection systems", "providing alternatives" and "adding a redundant subsystem". If failure in the infrastructure is reduced, the reliability of the network will increase proportionally and thus enhance capacity utilisation.

7.2 Improving Capacity Utilisation : the South West Main Line Case Study [1]

This section tries to improve the current level of traffic at the South West Main Line without making fundamental changes (which due to the interdependencies cannot be fulfilled at a line or regional level). Throughout the thesis it has been emphasised that a systems wide approach is needed to genuinely improve capacity utilisation. Hence, in section 5.3, a holistic method is suggested to improve capacity utilisation by train operators at the national level and in section 5.4 another holistic method is suggested to improve capacity utilisation at major stations across the country. This section just intends to provide a so called 'painkiller' with existing methods which is doable in a shorter time span.

Capacity utilisation can be improved overall by improving any of the parameters that affect capacity utilisation (as discussed in detail in chapter 3). Past approaches to improving different aspects of railway operations planning and capacity management were reviewed in chapter 2. In this section various methods to improve capacity utilisation are applied in a real world case study. The South West Main Line is one of the major commuter routes to London and also important for freight transportation from the port of Southampton (Network Rail, 2009a). Its geographic location is shown in Figure 7-4. Appendix 4 includes the detailed map of the line.

---

[1] This case study is based on KHADEM SAMENI, M., LANDEX, A. & PRESTON, J. 2011b. Developing the UIC 406 method for capacity analysis. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

Figure 7-4- South West Mainline Region according to Route Utilisation Strategy (Network Rail, 2006b)

As Figure 7-5 shows, current passenger loading levels are near practical capacity and the projected demand shows that it would be over practical capacity in 2030 (Department for Transport, 2007a). Percentage of passengers standing during the morning peak period (7:00 to 9:59) in the trains operated by South West Trains is the second highest in Great Britain (17%) (Network Rail, 2006a).

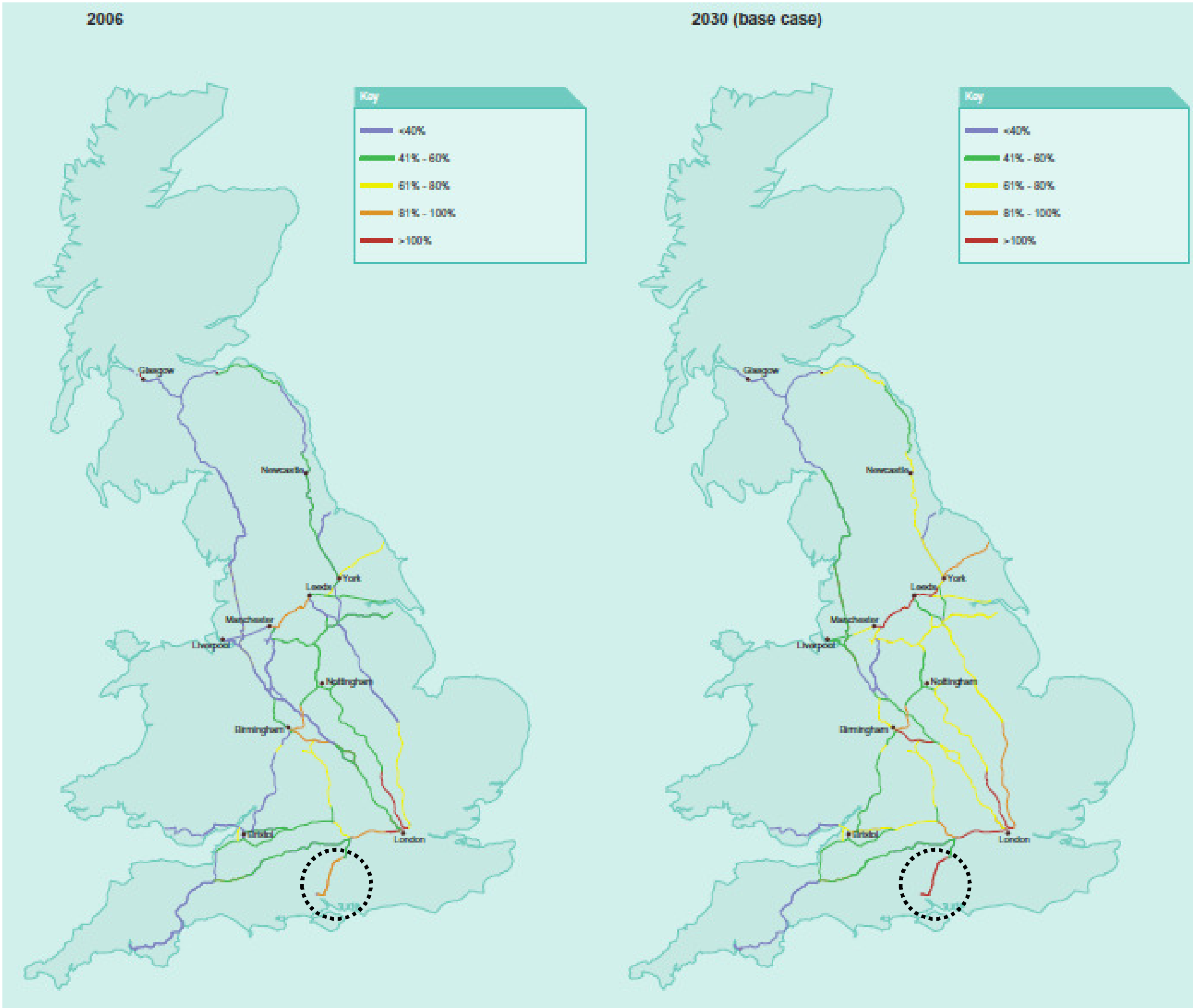
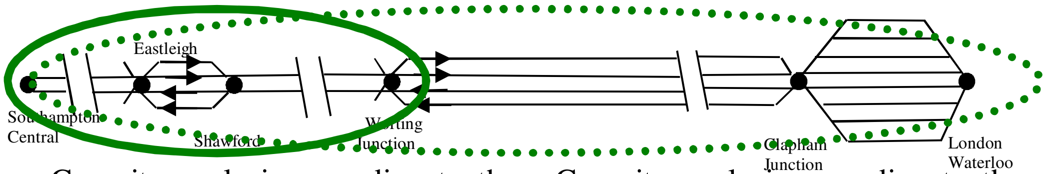


Figure 7-5- Loading levels on inter-urban services at peak hours (Department for Transport, 2007a)

To improve this line section, 'the weakest link of the chain' as discussed in section 4.3.5 should be identified. In the following sections, capacity utilisation analysis is done to find the weakest line section, the weakest station and the weakest train. The UIC 406 method and the CUI methods are used for capacity utilisation analysis of the lines. The DEA methodology for station capacity utilisation analysis in section 5.4 is used for stations. Delays are simulated by the RailSys 6 software to find the top delay causing trains.

7.2.1 Finding and improving the weakest line section

Capacity analysis is undertaken for the line between Southampton Central to London Waterloo stations. The line section between Southampton Central to Worting Junction has 2 tracks. From Worting Junction to Clapham Junction there are 4 tracks which increase to 8 tracks from Clapham Junction to London Waterloo. The line section between Southampton Central to Worting Junction is manually compressed for the CUI method. Capacity utilisation analysis for the whole line is done with the RailSys 6 software and its UIC 406 module. Timetable compression is done for the morning peak hours between 7:00 to 10:00 for travelling towards London.



Capacity analysis according to the CUI method

Capacity analysis according to the UIC 406 method

Figure 7-6- Change in the number of tracks from Southampton Central to London Waterloo

7.2.1.1 Capacity utilisation analysis by the UIC 406 method

The line was broken into sections according to the UIC 406 guidelines at junctions and where there is change in the number of tracks or trains. The results obtained by the RailSys 6 software are shown in Table 7-6.

7.2.1.2 Capacity utilisation analysis by the CUI method

Details of capacity utilisation analysis by the CUI method are not well documented online and are only accessible through Network Rail or Delta Rail. However, with the available information, the compression of timetable was done as shown in Figure 7-7. The CUI method uses headway values, therefore the line was broken between every two stations and the timetable manually compressed in the RailSys software. 'Fast headways' should be used when the preceding train does not stop at the station, otherwise 'slow headway' (which is longer) must be used (Network Rail, 2010f). Appropriate headways are identified for different sections of all routes in Great Britain in the 'Rules of the Plan' and the 'Rules of the Route'. For compressing the timetable, platform reoccupation time was considered where appropriate.

7.2.2 Comparing capacity utilisation analysis by the CUI and the UIC 406 method

The CUI method and the UIC 406 method compress the timetable slightly differently. Table 7-5 compares these two methods.

Table 7-5- Comparing the CUI and the UIC 406 methods

UIC 406	CUI
Considers blocking time at links	Considers either "slow" or "fast" headways for route sections
More detailed	Less detailed
Applied in the continental Europe	Applied in Great Britain
According to the general UIC 406 standard and national railways' specifications	According to the Network Rail's Rules of Plan

Figure 7-7 and Figure 7-8 illustrate the timetable compression results for the line section between Southampton Central to Worting Junction.

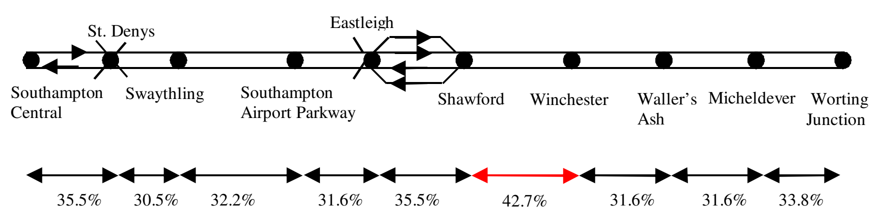


Figure 7-7- Timetable compression according to the CUI method from Southampton Central to Basingstoke (Khadem Sameni et al., 2011b).

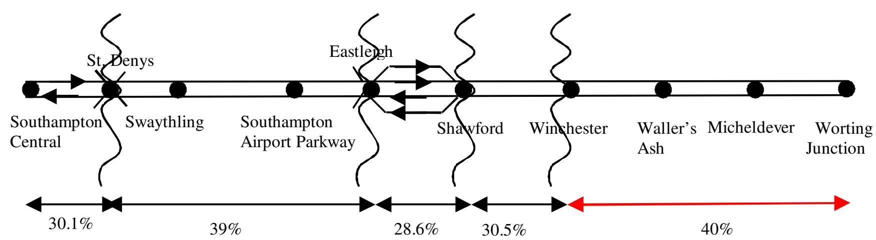


Figure 7-8- Timetable compression according to the UIC 406 method from Southampton Central to Basingstoke (Khadem Sameni et al., 2011b)

The average capacity utilisation by the UIC 406 method is 31.6% and for the CUI method is 33.2%. Shawford-Winchester section has the highest capacity utilisation according to the CUI method (42.7%). The maximum capacity utilisation according to the UIC 406 method belongs to Winchester-Worting Junction section (40%). The CUI method uses headway at nodes whereas the UIC 406 method considers headways at links. Hence part of the capacity constraints at stations is considered by the CUI method. For instance, longer headway times are set at Shawford station where there is a change between quadruple to double tracks which results in higher capacity utilisation by the CUI method. There is also a sudden jump in capacity utilisation from Shawford to Winchester due to the change from quadruple tracks to double tracks (Khadem Sameni et al., 2011b).

Table 7-6- Timetable compression according to the UIC 406 method from Southampton Central to London Waterloo

Line section	Capacity Utilisation	Reason for breaking the line
Southampton Central- St. Denys	30.1%	Junction
St. Denys – Eastleigh	39%	Junction
Eastleigh- Shawford	28.6%	Change in the number of tracks (2 to 4)
Shawford – Winchester	30.5%	Change in the number of tracks (4 to 2)
Winchester- Worting Junction	40%	Change in the number of tracks (2 to 4)
Worting Junction – Brookwood	21.0%	Junction
Brookwood-Woking	15%	Junction
Woking- Weybridge	42.8%	Junction
Weybridge- Hampton Court Junction	31.5%	Junction
Hampton Court Junction- New Malden	56.5%	Junction
New Malden- Raynes Park	30.8%	Junction
Raynes Park- Clapham Junction	87.4%	Change in the number of tracks (4 to 8)
Clapham Junction- Waterloo	86.4%	Terminus

As London Waterloo is a dead-end station, platform availability at this station poses serious capacity problems in the approaching lines. The capacity utilisation index from Raynes Park to Clapham Junction is 87.4% hence 'the weakest link' is this section.

7.2.3 Finding and improving the weakest trains

Two methods are presented to find the weakest train. The first method is used for the weakest line and the second method is used for the whole network.

7.2.3.1 Lowest meso index

If it is intended to remove a train to free up some capacity in this section, the decision cannot be solely made upon macro capacity utilisation (i.e. how much a train blocks the infrastructure as discussed in section 5.1.2- Discrete nature of railway capacity). Therefore the following methodology is suggested:

1. Calculating capacity utilisation for each scenario of omitting a train.
2. Developing a meso capacity utilisation criteria.
3. Calculating the meso capacity utilisation index for each scenario.
4. Choosing the best alternative.

Different meso capacity utilisation indices can be proposed. Hereby we suggest using the following for this case study:

$$Meso\ capacity\ \ index = \frac{Macro\ capacity\ gained\ by\ omitting\ \ the\ train}{Micro\ \ capacity\ lost} \quad (3)$$

$$Meso\ capacity\ index\ to\ free\ \ up\ capacity = \frac{C_b - C_a}{n_{cl}} = \frac{t_{ocu} / t_{total}}{n_{cl}} \quad (4)$$

C_b : Capacity utilisation before omitting the train

C_a : Capacity utilisation after omitting the train

n_{cl} : Number of carriages lost

t_{ocu} : Track occupation time of the train

t_{total} : Total time of analysis period

In this regard, the numerator considers how much micro capacity utilisation would be lost and the denominator considers how much macro capacity utilisation would be freed up by omitting the train. The number of carriages for different trains in Great Britain varies between 2 to 10. As the time period considered was for morning peak hours, the load factor of all trains was considered high and only the number of carriages is used. However, more complicated indices can be developed according to the distribution of loading factor during the peak hours.

Calculating this meso capacity utilisation index for all trains, it can be advised which train is the 'weakest link of the chain': the one with the lowest meso capacity utilisation index. The meso capacity utilisation index can be done for the the busiest line section or for the whole route. It is advisable to analyse the meso capacity utilisation index for the busiest part of the line section. Apart from less complexities being involved, that is where there is shortage of capacity. Enhancing capacity utilisation in any other part of the line will not improve the situation (please refer to section 4.3.5). For this case study meso capacity utilisation index is calculated manually for the trains passing from track 1 between Raynes Park to Clapham Junction (due to change in the layout of tracks it is done for Dursford Road Staff Halt to Clapham Junction). After the timetable is compressed, the time each train occupies the busiest line section is measured in seconds and divided by total time. The time each train occupies the busiest line section is derived from the graphical timetable The number of carriages is extracted from the name of the trains. The results are presented in Table 7-7.

Table 7-7 Meso capacity utilisation index

Train Number	Train Characteristics	Origin Station	Departure time	Arrival time to the Waterloo station	Meso index
2386	220_1x4C Mo-Fr	Portsmouth Harbour	7:45:00	10:08:00	173.49
1293	220_1x4C Mo-Fr	Basingstoke	8:54:00	10:06:00	147.95
1894	220_1x4C Mo-Fr	Alton	8:44:00	9:57:00	151.58
1683	220_1x8C Mo-Fr	Portsmouth Harbour	8:13:00	9:53:03	493.71
1619	442_2x5C Mo-Fr	Weymouth	6:54:00	9:51:00	360
1940	159_2x3C Mo-Fr	Yeovil Junction	7:20:00	9:49:00	216
1749	220_1x4C Mo-Fr	Bedhampton	7:49:00	9:41:00	144
2508	220_1x4C Mo-Fr	Portsmouth and Southsea	7:30:00	9:38:06	242.7
2400	220_1x4C Mo-Fr	Woking	8:47:00	9:38:00	197.26
1779	220_1x4C Mo-Fr	Portsmouth Harbour	7:49:00	9:29:00	324.81
1072	220_1x4C Mo-Fr	Basingstoke	8:24:00	9:27:01	327.27
2541	442_1x5C Mo-Fr	Bournemouth	7:04:00	9:20:00	402.99
1892	220_1x4C Mo-Fr	Alton	8:14:00	9:17:00	322.39
1647	159_2x3C Mo-Fr	Honiton	6:20:00	9:14:58	483.58
1720	220_1x8C Mo-Fr	Portsmouth Harbour	6:55:00	9:11:00	419.42
2110	442_1x5C Mo-Fr	Southampton	7:15:00	9:08:25	306.82
2398	220_1x4C Mo-Fr	Woking	8:17:00	9:04:00	411.43
2561	220_1x4C Mo-Fr	Portsmouth Harbour	6:40:00	9:00:24	322.39

Train Number	Train Characteristics	Origin Station	Departure time	Arrival time to the Waterloo station	Meso index
2446	455_1x8 Mo-Fr	Guildford	8:07:00	8:59:00	445.36
2662	220_1x4C Mo-Fr	Basingstoke	7:52:00	8:57:00	245.45
1682	220_1x8C Mo-Fr	Portsmouth Harbour	7:13:00	8:53:00	502.33
2396	220_1x4C Mo-Fr	West Byfleet	8:16:00	8:53:00	324.81
1865	442_1x5C Mo-Fr	Bournemouth	6:34:00	8:48:00	346.15
1891	220_1x8C Mo-Fr	Alton	7:44:00	8:47:59	644.78
2387	170_1x2C Mo-Fr	Whimple	5:26:00	8:43:38	163.64
2570	455_1x8 Mo-Fr	Woking	8:02:00	8:43:00	600
93	220_1x8C Mo-Fr	Haslemere	7:40:00	8:38:47	490.91
328	220_1x8C Mo-Fr	Basingstoke	7:36:00	8:37:00	488.14
2394	220_1x8C Mo-Fr	Woking	7:47:00	8:33:00	654.55
2491	220_1x8C Mo-Fr	Cosham	6:43:00	8:31:58	557.42
88	220_1x8C Mo-Fr	Portsmouth Harbour	6:44:00	8:29:59	553.85
326	220_1x8C Mo-Fr	Basingstoke	7:24:00	8:27:00	341.5
2392	220_1x8C Mo-Fr	West Byfleet	7:46:00	8:24:00	493.71
1748	220_1x8C Mo-Fr	Hilsea	6:42:00	8:22:00	480
1889	220_1x8C Mo-Fr	Alton	7:14:00	8:18:59	344.22
2390	455_1x8 Mo-Fr	Woking	7:32:00	8:17:00	469.57
2537	442_1x5C Mo-Fr	Bournemouth	6:04:00	8:14:00	270
1646	170_1x2C Mo-Fr	Yeovil Junction	5:50:00	8:12:02	114.89

Train Number	Train Characteristics	Origin Station	Departure time	Arrival time to the Waterloo station	Meso index
209	455_1x8 Mo-Fr	Guildford	7:17:00	8:10:59	553.85
1717	220_1x8C Mo-Fr	Haslemere	7:11:00	8:09:00	576
89	220_1x8C Mo-Fr	Southampton Airport Parkway	6:50:00	8:06:00	644.78
327	220_1x4C Mo-Fr	Basingstoke	7:06:00	8:02:00	204.74
2495	220_1x4C Mo-Fr	Portsmouth Harbour	5:44:59	7:58:33	220.41
265	220_1x4C Mo-Fr	Basingstoke	6:53:00	7:57:00	217.09
2391	220_1x8C Mo-Fr	Woking	7:11:00	7:54:00	502.33
1681	220_1x8C Mo-Fr	Portsmouth Harbour	6:15:00	7:53:00	462.03
1887	220_1x4C Mo-Fr	Alton	6:44:00	7:49:00	223.83
2389	220_1x8C Mo-Fr	Woking	7:02:00	7:47:01	533.33
1781	442_2x5C Mo-Fr	Poole	5:45:00	7:47:00	610.17
1645	159_2x3C Mo-Fr	Yeovil Junction	5:15:00	7:42:00	181.51
1778	220_1x8C Mo-Fr	Portsmouth Harbour	5:50:00	7:42:00	557.42
1716	220_1x4C Mo-Fr	Haslemere	6:32:00	7:35:00	144
2572	220_1x4C Mo-Fr	Basingstoke	6:24:00	7:29:59	120.67
2571	220_1x4C Mo-Fr	Woking	6:41:00	7:26:00	160
1780	442_2x5C Mo-Fr	Poole	5:00:00	7:24:01	465.52
1680	220_1x4C Mo-Fr	Alton	6:14:00	7:20:00	211.76
1643	159_2x3C Mo-Fr	Salisbury	5:45:00	7:16:00	310.05
1714	220_1x4C Mo-Fr	Portsmouth Harbour	5:19:00	7:14:00	205.71

The weakest train is the number 1646 leaving Yeovil Junction at 5:50:00 and arriving to London Waterloo at 8:12:02. It has only two carriages. The next three trains are number 2572 leaving Basingstoke with 4 carriages at 6:24:00 and arriving at London Waterloo at 7:29:59; number 1716 leaving Haslemere at 6:32:00 with 4 carriages and arriving to London Waterloo at 7:35:00 and number 1749 leaving Bedhampton at 7:49 with 4 carriages and arriving to London Waterloo at 9:41:00.

7.2.3.2 Highest delay causing trains

One of the methods suggested to improve capacity utilisation is deleting the top delay causing trains. The RailSys software does not directly provide total delays caused by each train. Hence after simulating delays, all the results are exported to Excel and by making a pivot table, total delays caused by each train are calculated. Top trains that cause delays to other trains are identified as shown in Table 7-8.

4 out of 5 of these trains run towards London Waterloo and in the most crowded time during the peak hours. They are stopping services. For instance, train number 2662 stops at Hook (60 seconds), Winchfield (60 seconds), Fleet (60 Seconds), Farnborough (120 seconds), Brookwood (30 seconds) and Woking (120 seconds). These stops cause considerable delays to other trains. This is also a short train carrying just 4 coaches. Omitting these trains brings down total delays from 527972 seconds to 451235 seconds which is a 14.5 % reduction. These trains are ‘the weakest links of the chain’.

Table 7-8 Top 5 delay causing trains

Train No. (obstructing Train)	**Delay Caused to other trains (sec.)**	**Origin- Destination**	**Journey time**
220_1x4C 2662	13863	Basingstoke- London Waterloo	7:52:00- 8:57:00
220_1x4C 2396	7285	West Byfleet- London Waterloo	8:16:00-8:53:00
220_1x8C 2691	7083	London Waterloo- Northam Junction	8:18:00-9:39:00
220_1x8C 2389	3422	Woking – London Waterloo	7:02:00-7:47:01
220_1x4C 1894	2840	Alton- London Waterloo	8:44:00-9:57:00

After removing these trains, capacity utilisation by the UIC 406 method can be recalculated which are presented in Table 7-9. As a result of this, at many line sections there is reduction in capacity utilisation. However, for the line sections between New Malden to Waterloo stations, turn round time at Waterloo seems to be a constraint.

Table 7-9- Capacity utilisation before and after omitting top delays causing trains

Line section	Capacity Utilisation	Reduction
Southampton Central- St. Denys	26.4%	3.7%
St. Denys – Eastleigh	37.4%	1.6%
Eastleigh- Shawford	26.7%	1.9%
Shawford – Winchester	29.3%	1.2%
Winchester- Worting Junction	37.7%	2.3%
Worting Junction – Brookwood	19.4%	1.6%
Brookwood-Woking	15%	0%
Woking- Weybridge	41.2%	1.6%
Weybridge- Hampton Court Junction	19.4%	12%
Hampton Court Junction- New Malden	48.7%	7.8%
New Malden- Raynes Park	30.8%	0%
Raynes Park- Clapham Junction	87.4%	0%
Clapham Junction- Waterloo	86.4%	0%

7.2.4 Finding and improving the weakest stations

To analyse capacity utilisation at stations and finding the weakest stations, service effectiveness model presented in section 5.4 is used to quantify the relative service effectiveness of stations during a day. London Waterloo and Clapham Junction stations, being busiest stations in the country in terms of total number of passengers and trains, were excluded after running the initial model to provide a homogeneous batch of stations. The results are presented in Table 7-10.

Table 7-10 Service effectiveness for SWML stations

Name	Service Effectiveness
Southampton Central	0.71
St Denys	0.55
Swaythling	0.88
Southampton Airport Parkway	1.00
Eastleigh	0.54
Shawford	0.94
Winchester	0.80
Micheldever	1.00
Basingstoke	0.72
Hook	0.76
Winchfield	0.64
Fleet	0.84
Farnborough (Main)	0.96

Name	Service Effectiveness
Brookwood	0.48
Woking	1.00
West Byfleet	0.52
Byfleet & New Haw	0.63
Weybridge	0.71
Walton-On-Thames	0.87
Hersham	0.73
Esher	0.88
Surbiton	1.00
Berrylands	0.79
New Malden	0.70
Raynes Park	0.50
Wimbledon	1.00
Earlsfield	0.60
Queenstown Road (Battersea)	0.37
Vauxhall	1.00

As previously mentioned, in the service effectiveness model (section 5.4.4), the inputs are total number of train stops, population and job opportunities in the catchment area and the outputs are total passenger entries and exits and passenger interchanges. Stations that have lower service effectiveness scores do not generate enough passenger entries and exists and passenger interchanges for the number of train stops. Hence reducing train stops at these stations are suggested. Based on DEA results, the optimum numbers of train stops are presented in Table 7-11. These results can be used for timetabling at the tactical level.

Table 7-11 Target values for the number of the train stops on a working day

Station	Suggested reduction (percentage)	Current number of train stops	Suggested
Southampton Central	-29	317	225
St Denys	-45	92	51
Swaythling	-12	51	45
Southampton Airport Parkway	0	184	184
Eastleigh	-46	162	87
Shawford	-6	48	45
Winchester	-20	208	166
Micheldever	0	45	45
Basingstoke	-28	371	267
Hook	-24	81	62
Winchfield	-36	81	52

Station	Suggested reduction (percentage)	Current number of train stops	Suggested
Fleet	-16	112	94
Farnborough (Main)	-4	136	130
Brookwood	-52	149	71
Woking	0	499	499
West Byfleet	-48	152	80
Byfleet & New Haw	-37	87	55
Weybridge	-29	221	156
Walton-On-Thames	-13	148	129
Hersham	-27	88	64
Esher	-12	88	78
Surbiton	0	357	357
Berrylands	-21	67	53
New Malden	-30	218	153
Raynes Park	-50	414	208
Wimbledon	0	682	682
Earlsfield	-44	556	311
Queenstown Road (Battersea)	-63	216	80
Vauxhall	0	791	791

It should be noted that removing trains does not always result in reducing capacity utilisation (such as the stretch of the line between Raynes Park and Clapham Junction and Clapham Junction- Waterloo). If there are several tracks, and the omitting train is not using the most congested track, capacity utilisation index would remain the same.

7.3 Systems engineering and real world implications

In section 4.3.1, the importance of a holistic approach towards capacity utilisation was emphasised. This holistic measure can even be holistic consideration of time which was done by the model proposed in section 7.1.1 that analysed punctuality variation over years. For this end a creative (section 4.3.6) index was proposed as Public Performance Variation (PPV). Considering this index can reduce knock-on delays which causes waste in capacity utilisation.

To reduce the risks to infrastructure failures in a holistic manner (section 4.3.1), all the risks should be considered. This is why the Failure Mode and Effect Analysis model was developed. In this way the problem of fragmentation of knowledge can be tackled as well hence various risks to infrastructure failure spanning from signalling to civil engineering can be accommodated in one model. The model is holistic from another point of view as it summarises all aspects of the risk as severity, occurrence and detection in one number

(RPN). The model follows need for innovative problem solving (section 4.3.6) as it uses FMEA in new contexts for capacity utilisation analysis.

In the tough economic situation (section 4.3.8) focusing on higher risk to infrastructure failures (the weakest links as discussed in section 4.3.5) will increase efficiency in capacity utilisation. The results of the model show that "Track circuit failure" and "Signalling system and power supply failures" are top two risks. This may indicate the need to move toward modern signalling systems of ERTMS to improve capacity utilisation (section 3.2).

The importance of finding the 'weakest link of the chain' to improve the system's performance was discussed in section 4.3.5) The model presented in section aimed for finding the weakest link as a line section, train and station in a real world case study of South West Main Line in Great Britain. Various existing methods have been used for this end. Meso index for capacity utilisation was suggested as a creative (section 4.3.6) and holistic (section 4.3.1) measure. Overall this case study shows how efficiency of capacity utilisation can be improved for a specific route.

The results show that the line sections near London are most crowded ones: Raynes Park-Clapham Junction and Clapham Junction-Waterloo. For improving these weakest links of the chain, improving the turnaround time at the terminal station of Waterloo can be helpful. Stopping services that generate excessive delays to other trains can be omitted. These services can be replaced with bus services which is another reason why a holistic measure for capacity utilisation is needed that can preferably even cover all modes of transportation.

Analysing station capacity utilisation showed that the smaller stations are usually not efficient. Small stations can be efficient if they have optimum number of train stops. Excessive train stops causes inefficiency. For instance there are four train stations at the city of Southampton: Southampton Central, Southampton Airport Parkway, St. Denys and Swaythling. The main and mostly used stations are the first two (where there are regular services to London, etc.). However, there are some excessive train stops for St.Denys station which is mainly used by local residents which makes the relative efficiency score of this station low (0.55). If train stops for local stations are at the necessary level, higher efficiency scores is obtained which is the case with Swaythling station (efficiency score of 0.88).

7.4 Summary and Conclusions

The focus of this chapter was on methods to control and improve capacity utilisation in the DMAIC cycle. Two methods for controlling capacity utilisation were suggested. Variation control, a tool used in Six Sigma practices, tries to reduce variance in services. This is especially important for the case of train punctuality and reliability as it is the first priority of passengers and severely affects capacity utilisation. However, the existing measures of punctuality, such as public performance measure in the UK just consider mean of punctuality and reliability not the variance. Hence, a new index is formulated to measure public performance variation for different train operators.

The second tool adopted for controlling capacity utilisation is Failure Mode and Effect Analysis (FMEA). To control reliability and stability of the network which immensely affects capacity utilisation, risk priority numbers (RPNs) were calculated for major risks resulting in infrastructure failure. Occurrence rating, severity rating and detection rating were developed for the risks. The highest RPN belongs to track circuit failure, the highest occurrence to point failure and the highest severity to cable faults. These risks need special attention to control capacity utilisation effectively. More investments needs to be done on these risks.

Improving capacity utilisation is discussed in three categories for line sections, trains and stations. The UIC 406 method and the CUI method can identify the busiest line section. Their results are compared for the first time in the literature by applying them to a line section in the South West Main Line. For finding the weakest trains two methods are presented. The first method suggests developing a meso index to consider both aspects of macro and micro capacity utilisation. To identify which train to remove to free up capacity it should be considered how much track capacity is obtained (macro capacity) and how many passenger- seats are lost (micro capacity). Macro capacity spared can be calculated by subtracting the ‘before’ capacity utilisation index from the ‘after’ capacity utilisation index or dividing total track occupation time by the total time. Number of carriages can be a good proxy variable for the micro capacity that is lost. Meso capacity utilisation index can be calculated for the busiest section of the line or for the whole line. The former involves less complexity and follows the logic of improving the weakest link of the chain. The second method for identifying the weakest trains is simulating train delays and identifying top delay causing trains.

To improve capacity utilisation at stations it is suggested to use the DEA methodology developed in section 5.4. The weakest stations can be identified by the results of the model. Capacity waste at station can be reduced by using DEA target values in tactical timetabling.

8 Conclusions

Growth in rail passenger demand outweighs the increase in the supply side of railway track capacity. Hence many railways worldwide are facing a capacity challenge. Facing this situation, it is very important to measure and manage capacity utilisation appropriately. For highways a comprehensive capacity manual is published by Transportation Research Board (TRB) which is over 1650 pages. However the corresponding capacity leaflet in railways published by the International Union of Railways (UIC) is just 24 pages. This PhD thesis is set out to put the building blocks of such a comprehensive railway capacity manual including defining, measuring, analysing, improving and controlling capacity utilisation.

Returning to research questions stated in section 1.7, the findings of the study suggests using a systems engineering approach towards railway capacity utilisation. The need for adopting such an approach is especially needed due to separation of track ownership and train operation after privatisation, the complex and multidisciplinary nature of railway capacity utilisation and the tough economic situation. The thesis suggests adopting the DMAIC cycle from Six Sigma to deal with defining, measuring, analysing and improving capacity utilisation at the tactical level.

As the first step it is emphasized to differentiate between micro and macro capacity utilisation. In this regard macro capacity utilisation is defined as the quantity of discrete steps to use railway capacity such as the number of trains and train paths whereas micro capacity utilisation is defined as the quality of discrete steps to use railway capacity such as load factor of trains. Lean capacity utilisation is hence defined as “The ability of the infrastructure to generate added value by enabling passengers or freight to reach their destination as planned”. It is a function of both micro and macro capacity utilisation.

With regards to the research question on measuring capacity utilisation, strengths and weaknesses of analytical methods, parametric models, optimisation and simulation are summarised in Table 5-2. Using DEA is suggested as a new method for capacity utilisation analysis. As stated by the International Union of Railways (2004), “Capacity as such does not exist”. The major strength of using the DEA-based methodologies is measuring relative efficiency of units in utilising capacity rather than measuring it directly. It can also encompass the multidisciplinary nature of railway capacity utilisation. In this way two rather isolated trends of analysing railway efficiency by economists and analysing capacity utilisation by engineers are linked together.

To measure relative capacity utilisation by train operators it is suggested to consider franchise payments from the government and timetabled train-kilometres as inputs for the DEA model. Passenger-kilometres and delay-minutes (undesirable effect) are discussed and considered to be good indicators of lean capacity utilisation and quality of service respectively. They were chosen as the outputs of the model.

To measure relative capacity utilisation at stations a two-stage DEA models is suggested. The first model measures relative technical efficiency of stations in accommodating train stops with the available infrastructure. The inputs of this model are the number of platforms and percentage of through lines. The output of the technical effiency model is the number of train stops. For the service effectiveness model catchment area population of the station, job opportunities in the catchment area and the number of train stops are the inputs of the model and the number of passenger entries and exits as well as the number of passenger interchanges is chosen as the outputs.

As the concept of freight railways is revolves around the revenue and the profit, and social impacts are less important than passenger services, the thesis recommends using profit-generating capacity utilisation for them. The developed methodology simulates various scenarios for traffic combinations on the railway network where the number of trains and the percentage of a specific train type (heterogeneity) vary. Total costs, revenues and net profit is calculated for each scenario to choose the optimum traffic combination. Another method developed for freight railways is based on DEA. It suggests using wagon loads originated and tons originated for different commodities as inputs and total revenue as the output. This model can identify the most profitable commodities for the railway.

One of the methods to control capacity utilisation is controlling variance. Hence the new index of public performance variation (PPV) is suggested to complement public performance measure (PPM). To improve and control capacity utilisation the results of this research supports the idea that the weakest link of the chain should be identified and improved. As one of the pillars of railway capacity utilisation is reliability a method needs to be developed to quantify risks affecting it. Failure mode and effect analysis (FMEA) identifies and prioritises such risks by considering the occurrence, severity and detection. The findings of the study shows that track circuit failure and signalling and power system failure pose the highest risk to railway capacity utilisation.

Improving capacity utilisation can be done by improving the weakest line section, trains or stations. Currently there are two methods in the literature to identify the busiest line sections. The CUI method is used in Great Britain and the UIC 406 method is used in the continental Europe. Comparing their results in a case study shows that they provide close results while the CUI results are a bit higher. Simulating train delays and identifying the top delay causing trains can help to pinpoint the weakest trains. Another method was based on developing a meso index that considers the number of carriages as well as capacity utilisation of a train. The former being the numerator and the latter the denominator, trains that have the lowest meso indices are good candidates for being removed to free up some capacity in critical and busiest blocks. To improve capacity utilisation at stations, the DEA methodology developed in chapter 5 is applied. Using target values obtained from the DEA model and feeding it to the tactical timetabling process can remove unnecessary train stops and eliminates capacity waste.

The collection of all the above mentioned methodologies and models provides means to "improve efficiency and reduce costs to taxpayers" and pave the way towards "getting the best out of the rail network" which are the key issues for consultation for the 2013 periodic review as identified by Office of Rail Regulation (Office of Rail Regulation, 2011a)

8.1 Main Contributions of the thesis

The main contributions of the thesis can be summarised as:

- A complete survey of past approaches to railway capacity (Chapter 2)
- Review of various factors affecting capacity utilisation (Chapter 3)
- Developing a framework for a railway capacity manual (Chapter 4-7)
- Adopting a systems engineering approach towards measuring and managing railway capacity utilisation (Chapter 4)
- Defining lean, micro and macro railway capacity utilisation (Chapter 4 and 5)
- Developing two novel methodologies based on DEA method for measuring and analysing railway capacity utilisation in the passenger sector by for train operators and for stations (Chapter 5)
- Developing revenue generating capacity utilisation as a novel methodology for measuring and managing capacity utilisation in the freight sector (Chapter 6)
- Improving and controlling capacity utilisation by applying variation reduction and failure mode and effect analysis method (Chapter 7)
- Developing methods to find the weakest line section, station and train to improve capacity utilisation (Chapter 7)

These contributions are schematically shown in Figure 8-1.

8.2 Implications for practice and policy

This thesis suggests several courses of action for railway practitioners and policy makers. There is great need in the industry to develop an international comprehensive railway capacity manual that encompasses various aspects of defining, measuring, improving and controlling capacity utilisation.

Another important practical implication is reducing capacity waste at stations and by trains. There are some unnecessary train stops at stations that make train journeys longer, increase track occupation time while not attracting enough passengers. The DEA methodology for station capacity utilisation analysis can be applied at a national level to rank the relative operational efficiency and service effectiveness of stations. Target values obtained from the model can be fed to the tactical timetabling process and enhance train operations.

It is also important to assess operational efficiency of passenger train operators based on their efficiency in using the allocated track capacity and franchise payments by the quality and quantity of the service they provide. For the freight sector, the best indicator of efficient capacity utilisation is the revenue and profit generated.

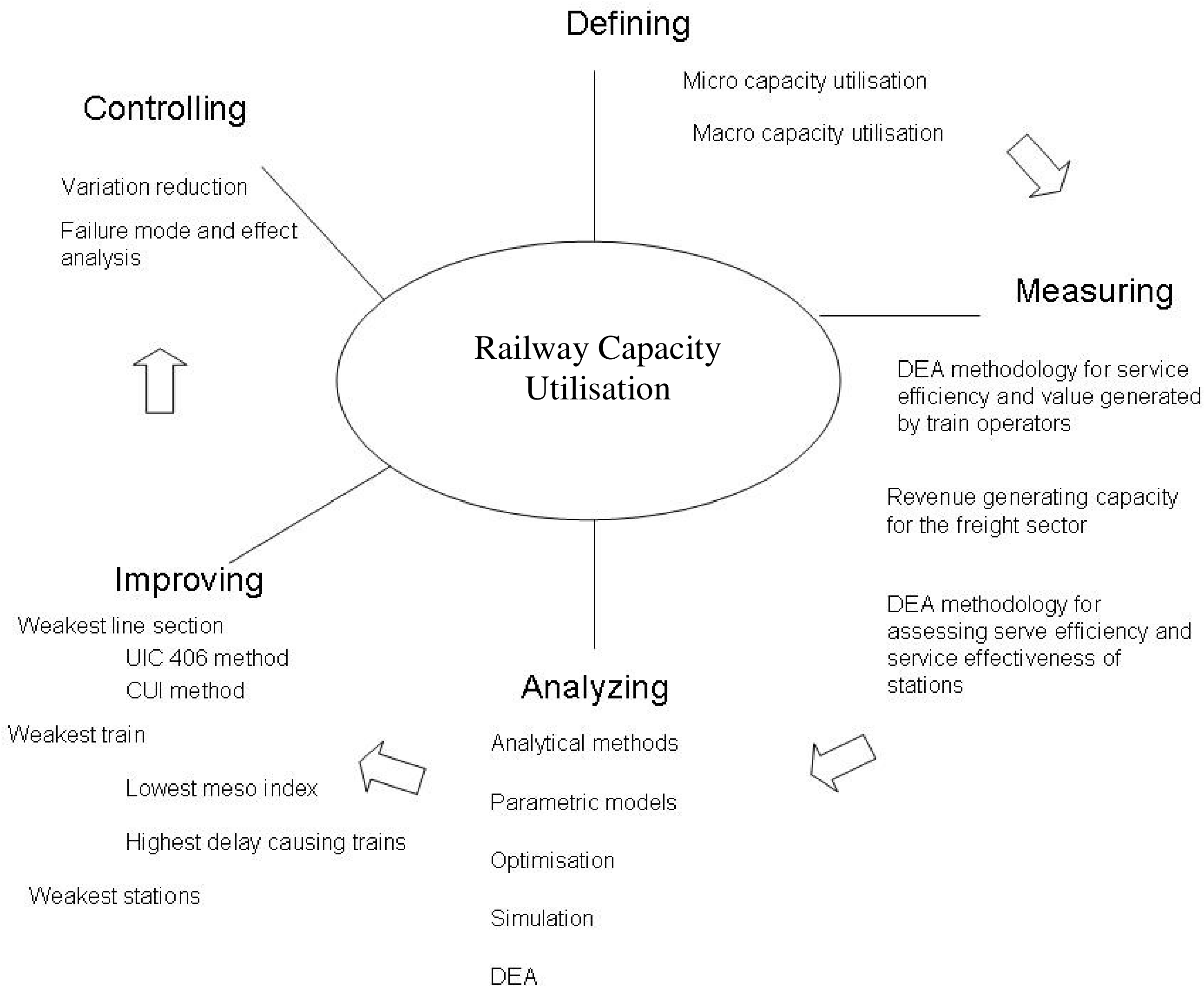


Figure 8-1 Main contributions of the thesis

The new index of Public Performance Variation (PPV) that intends to reflect the daily variation in Public Performance Measure (PPM) can be a good complement to it.

8.3 Limitations of the current study

The current research was limited by a number of constraints. The available timetable used for the SWML case study belongs to year 2005. However there are not much significant changes in the timetable since then that affects the results.

Data for freight trains was not available in the UK hence the freight case study was done in the US context. Exact capacity utilisation indices as used by Network Rail were not accessible to compare with the results of the thesis.

8.4 Recommendations for further research

The current study has studied passenger and freight sector separately. Further research can assess capacity utilisation in a mixed traffic. Revenue and profit generating capacity utilisation is suggested to be extended to passenger trains as well by considering the costs of delay and the value of time. Reductions in CO_2 emissions can also be monetised and be included in the profit generating capacity utilisation. For the DEA models developed for stations and train operators, it would be interesting to do an international study to evaluate major stations of various countries.

9 References

ABRIL, M., BARBER, F., INGOLOTTI, L., SALIDO, M. A., TORMOS, P. & LOVA, A. 2008. An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review,* 44**,** 774-806.

AFFUSO, L. 2003. Auctions of rail capacity? *Utilities Policy,* 11**,** 43-46.

ALBRECHT, A. R., PANTON, D. M. & LEE, D. H. in press. Rescheduling rail networks with maintenance disruptions using Problem Space Search. *Computers & Operations Research*.

AMEMIYA, T. 1985. *Advanced econometrics,* Cambridge, Mass., Harvard University Press.

AMERICAN PUBLIC TRANSPORTATION ASSOCIATION 2006. The U.S rail capacity shortage Washington.

ANTONY, J. 2006. Six Sigma for service processes. *Business Process Management Journal,* 12**,** 234-248.

ARMSTRONG, J. 1998. The Railroad: What It Is, What It Does. *Simmons-Boardman Books, Inc*.

ARMSTRONG, J., BLAINEY, S. & PRESTON, J. 2011a. Developing a CUI-based Approach to Network Capacity Assessment. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

ARMSTRONG, J., HOOD, I. & PRESTON, J. Automating the Production of Train Graphs and the Calculation of CUI Values 3rd International Seminar on Railway Operations Modelling and Analysis, 2009 Zurich, Switzerland.

ARMSTRONG, J., PRESTON, J., KHADEM SAMENI, M., POTTS, C., KHOSRAVI, B., BEKTAS, T. & BENNELL, J. 2011b. Overcoming the capacity constraints imposed by nodes on railway networks. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

ARUP 2011. Rail Value for money study :Rolling stock whole life costs. London: Commissioned by Department for Transport and Office of Rail Regulation. Available online at www.dft.gov.uk.

ASSAD, A. A. 1980. Models for rail transportation. *Transportation Research Part A: General,* 14**,** 205-220.

ASSOCIATION OF AMERICAN RAILROADS 2008. Analysis of Class I Railroads Washington, DC.

ASSOCIATION OF AMERICAN RAILROADS 2008. *Class I Railroad Statistics* [Online]. Washington, DC. Available: www.aar.org/pubcommon/documents/abouttheindustry/statistics.pdf [Accessed 10/04/2010.

ASSOCIATION OF AMERICAN RAILROADS. 2011. *Class I Railroad Statistics* [Online]. Washington, DC. Available: www.aar.org/ [Accessed 06/09/2011.

ASSOCIATION OF TRAIN OPERATING COMPANIES 2007. Ten-year European Rail Growth Trends. London.

ATKINS 2011. Rail Value for Money Study :Asset Management and Supply Chain Management Assessment of GB Rail. London: Commissioned by Department for Transport and Office of Rail Regulation. Available online at www.dft.gov.uk.

BAKER, K. R. & TRIETSCH, D. 2009. *Principles of Sequencing and Scheduling,* Hoboken, New Jersey, Wiley.

BANKER, R. D., CHARNES, A. & COOPER, W. W. 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science,* 30**,** 1078-1092.

BARBER, F., ABRIL, M., SALIDO, M. A., INGOLOTTI, L., TORMOS, P. & LOVA, A. 2007. Survey of automated systems for railway management Valencia: Department of Computer Systems and Computation, Technical University of Valencia.

BASU, R. 2004. *Implementing quality : a practical guide to tools and techniques : enabling the power of operational excellence,* London, Thomson Learning.

BERTALANFFY, L. V. 1968. *General system theory:Foundations, development, applications,* New York, George Braziller, Inc.

BESTERFIELD, D., BESTERFIELD-MICHNA, C., BESTERFIELD, G., BESTERFIELD-SACRE, M., URDHWARESHE, H. & URDHWARESHE, R. 2011. *Total quality management,* Delhi, Pearson.

BLAINEY, S. 2009. *Forecasting the Use of New Local Railway Stations and Services Using GIS.* PhD, University of Southampton.

BLAINEY, S. 2010. Trip end models of local rail demand in England and Wales. *Journal of Transport Geography,* 18**,** 153-165.

BLANCHARD, B. S. 1991. *System Engineering Management,* New York, Wiley.

BOJOVIC, N. J. 2002. A general system theory approach to rail freight car fleet sizing. *European Journal of Operational Research,* 136**,** 136-172.

BONNETT, C. F. 2005. *Practical railway engineering,* London, Imperial College Press.

BOOZ & COMPANY 2011. Rail Value for Money Study: Research on Value for Money Assessment London: Commissioned by Department for Transport and Office of Rail Regulation. Available online at www.dft.gov.uk.

BOULDING, K. 1956. General systems theory-the skeleton of science. *Management science,* 2**,** 197-208.

BUCHMÜLLER, S., WEIDMANN, U. & NASH, A. 2008. Development of a Dwell Time Calculation Model for Timetable Planning. *Computers in Railways XI***,** 525–534.

BUREAU OF TRANSPORTATION STATISTICS 2007. Commodity Flow Survey. Washington, DC: Research and Innovation Technology Administration.

BURKOLTER, D. M. 2005. *Capacity of Railways in Station Areas using Petri Nets.* Doctor of Sciences, Swiss Federal Institute Of Technology Zurich.

BUSH, R. Does every trip need to be on time? Multimodal Scheduling Performance Parameters with an application to Amtrak Service in North Carolina. 86th Annual Meeting at Transport Research Board, 2007 Washington D.C., USA.

CABRERA, D., COLOSI, L. & LOBDELL, C. 2008. Systems thinking. *Evaluation and Program Planning,* 31**,** 299-310.

CAMBRIDGE SYSTEMATICS 2007. National rail freight infrastructure capacity and investment study. Cambridge , USA: Association of American Railroads.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics,* Cambridge; New York, Cambridge University Press.

CANTOS, P. & MAUDOS, J. 2001. Regulation and efficiency: the case of European railways. *Transportation Research Part A: Policy and Practice,* 35**,** 459-472.

CANTOS, P., PASTOR, J. M. & SERRANO, L. 2002. *Cost and revenue inefficiencies in the European railways*, Instituto Valenciano de Investigaciones Económicas.

CANTOS, P., PASTOR, J. M. & SERRANO, L. 2010. Vertical and horizontal separation in the European Railway Sector and its effects on Productivity. *Journal of Transport Economics and Policy (JTEP),* 44**,** 139-160.

CAREY, M. & CARVILLE, S. 2003. Scheduling and platforming trains at busy complex stations. *Transportation Research Part A: Policy and Practice,* 37**,** 195-224.

CAREY, M. & CRAWFORD, I. 2007. Scheduling trains on a network of busy complex stations. *Transportation Research Part B: Methodological,* 41**,** 159-178.

CASTI, J. 1986. On system complexity: Identification, measurement and management. *Complexity, Language and Life: Mathematical Approaches,* 16**,** 146–173.

CETIN, C. & CERIT, A. 2010. Organizational Effectiveness at Seaports: A Systems Approach. *Maritime Policy & Management,* 37.

CHARANTIMATH, P. M. 2011. *Total Quality Management,* New Delhi, Dorling Kindersley Pvt. Ltd.

CHARNES, A., COOPER, W. W. & RHODES, E. 1978a. Measuring the efficiency of decision making units. *European Journal of Operations Research,* 2**,** 429-44.

CHARNES, A., COOPER, W. W. & RHODES, E. 1978b. Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2**,** 429-444.

CHECKLAND, P. 1999. *Soft Systems Methodology: A 30-year Retrospective,* Chichester, Wiley.

CHIN, K.-S., WANG, Y.-M., POON, G. K. K. & YANG, J.-B. 2009. Failure mode and effects analysis by data envelopment analysis. *Decision Support Systems,* 48**,** 246-256.

CIVITY MANAGEMENT CONSULTANTS 2011. International whole industry including train operating cost benchmarking Hamburg: Commissioned by Department for Transport and Office of Rail Regulation. Available online at www.dft.gov.uk.

CLIMENT, A. 2009. *RE: Interoperability.*

COELLI, T. & PERELMAN, S. 1999. A comparison of parametric and non-parametric distance functions: With application to European railways. *European Journal of Operational Research,* 117**,** 326-339.

COELLI, T. & PERELMAN, S. 2000. Technical efficiency of European railways: a distance function approach. *Applied Economics,* 32**,** 1967 - 1976.

CONFESSORE, G., CICINI, P., DE LUCA, P., LIOTTA, G. & RONDINONE, F. A simulation-based approach for estimating the commercial capacity of railways *In:* ROSSETTI, M. D., HILL, R. R., JOHANSSON, B., DUNKIN, A. & INGALLS, R. G., eds. Winter Simulation Conference 2009 Austin, TX, USA.

COOK, W. & ZHU, J. 2008. *Data envelopment analysis: Modeling operational processes and measuring productivity*, Create Space.

COOPER, W. W., SEIFORD, L. M. & TONE, K. 2006. *Data envelopment analysis : a comprehensive text with models, applications, references and DEA-solver software,* New York, Springer.

CORDEAU, J. F., TOTH, P. & VIGO, D. 1998. A survey of optimization models for train routing and scheduling. *Transportation Science,* 32**,** 380-420.

COWIE, J. Technical efficiency versus technical change -the British passenger train operators. *In:* HENSHER, D., ed. International Conference on Competition Ownership in Land Passenger Transport, 2005 2005 Amsterdam. Amsterdam: Elsevier, 739-764.

CROCKER, O. L., CHIU, J. S. L. & CHARNEY, C. 1984. *Quality circles : a guide to participation and productivity,* Toronto; New York, Methuen.

DAAMEN, W., GOVERDE, R. & HANSEN, I. 2009. Non-Discriminatory Automatic Registration of Knock-On Train Delays. *Networks and Spatial Economics,* 9**,** 47-61.

DARE, A. Running a smarter railway. Growth and the capacity challenge-an international perspective, 2009 London, UK.

DE JONG, G. 2000. Value of freight travel-time savings. *Handbook of Transport Modelling, Elsevier***,** 553-564.

DEPARTMENT FOR TRANSPORT 2004. The Future of Rail London, UK.

DEPARTMENT FOR TRANSPORT 2007a. Delivering a sustainable railway. London.

DEPARTMENT FOR TRANSPORT 2007b. Network Modelling Framework London.

DEPARTMENT FOR TRANSPORT. 2008. *High Level Output Specification* [Online]. London,UK. Available: http://www.dft.gov.uk/pgr/rail/strategyfinance/strategy/hlos/ [Accessed January 20, 2010.

DEPARTMENT FOR TRANSPORT AND OFFICE OF RAIL REGULATION 2011. Realising the Potential of GB Rail - Final Independent Report of the Rail Value for Money Study London: Available online at http://www.dft.gov.uk

DEPARTMENT FOR TRANSPORT AND THE OFFICE OF RAIL REGULATION 2010. Increasing passenger rail capacity. London.

DICEMBRE, A. & RICCI, S. 2011. Railway traffic on high density urban corridors: capacity, signalling and timetable. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

DINGLER, M. 2010. *Understanding the Impact of Operations and New Technologies on Railroad Capacity.* MSc., University of Illinois at Urbana-Champaign.

DINGLER, M., LAI, Y. C. & BARKAN, C. P. L. 2009a. Impact of operating heterogeneity on railway capacity. *88th Transportation Research Board annual meeting.* Washington, USA.

DINGLER, M. H., LAI, Y. C. & C.P.L., B. 2009b. Impact of Operating Heterogeneity on Railway Capacity. *Transportation Research Record - Journal of Transportation Research Board,* 2117**,** 41-49.

DIRNBERGER, J. & BARKAN, C. 2007. Lean Railroading for Improving Railroad Classification Terminal Performance: Bottleneck Management Methods. *Transportation Research Record: Journal of the Transportation Research Board,* 1995**,** 52-61.

DRIESSEN, G., LIJESEN, M., MULDER, M. & PLANBUREAU, N. C. 2006. *The impact of competition on productive efficiency in European railways*, CPB Netherlands Bureau for Economic Policy Analysis.

EDDINGTON, S. R. 2006. The Eddington Transport Study. London, UK: HM Treasury and Department for Transport

EMROUZNEJAD, A. & THANASSOULIS, E. 2011. *Performance Improvement Management Software* [Online]. Coventry. Available: http://www.deasoftware.co.uk [Accessed 29/05/2011 2011].

EROL, B., KLEMENZ, M., SCHLECHTE, T., SCHULTZ, S. & TANNER, A. TTPlib 2008 – a library for train timetabling problems. *In:* ALLAN, J., ARIAS, E., BREBBIA, C. A., GOODMAN, C. J., RUMSEY, A. F., SCIUTTO, G. & TOMII, N., eds. Comprail 2008, 2008 Toledo, Spain. Wessex Institute of Technology Press, 605-614.

FABER MAUNSELL. 2007. *Recalculating the Capacity Charge Tariff for PR2008* [Online]. Network Rail. Available: http://www.networkrail.co.uk/browse%20documents/strategicbusinessplan/charges/faber%20maunsell%20capacity%20chargetariff%20report%20%20october%202007%20.pdf [Accessed 12/08/2010.

FARRELL, M. J. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General),* 120**,** 253-290.

FLOOD, R. L. & JACKSON, M. C. 1991. *Creative Problem Solving:Total Systems Intervention,* Chichester Wiley.

FRANKLIN + ANDREWS LTD. 2004. *Spon's Railways Construction Price Book,* London, Routledge.

FRIEDLAENDER, A. & SPADY, R. 1981. *Freight transport regulation: Equity, efficiency, and competition in the rail and trucking industries,* Cambridge, MIT press.

FRISCH, R. 1936. Annual survey of general economic theory: The problem of index numbers. *Econometrica: Journal of the Econometric Society***,** 1-38.

GASPARIK, J. & ZITRICKY, V. 2011. Aspects of railway capacity and new idea of occupation time estimation. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

GATHON, H. J. & PESTIEAU, P. 1995. Decomposing efficiency into its managerial and its regulatory components: The case of European railways. *European Journal of Operational Research,* 80**,** 500-507.

GEORGE, M. L. 2002. *Lean Six Sigma : Combining Six Sigma Quality with Lean Production Speed,* Blacklick, OH, USA, McGraw-Hill Professional Publishing.

GEORGE, M. L. 2003. *Lean Six Sigma for Service,* Blacklick, OH, USA, McGraw-Hill Professional Publishing.

GIBSON, S., COOPER, G. & BALL, B. 2002. Developments in Transport Policy: The Evolution of Capacity Charges on the UK Rail Network. *Journal of Transport Economics and Policy,* 36**,** 341-354.

GILLE, A., KLEMENZ, M. & SIEFER, T. Applying multiscaling analysis to detect capacity resources in railway networks. *In:* ALLAN, J., ARIAS, E., BREBBIA, C. A., GOODMAN, C. J., RUMSEY, A. F., SCIUTTO, G. & TOMII, N., eds. Comprail 2008, 2008 Toledo, Spain. Wessex Institute of Technology Press, 595-604.

GILLE, A. & SIEFER, T. 2011. Capacity simulation. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

GOVERDE, R. M. P. 2007. Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological,* 41**,** 179-201.

GROWITSCH, C. & WETZEL, H. 2009. Testing for economies of scope in European railways: An efficiency analysis. *Journal of Transport Economics and Policy (JTEP),* 43**,** 1-24.

HAINES, S. G. 2000. *The Complete guide to systems thinking and learning,* Amherst, MA, HRD Press.

HALL, A. & FAGEN, R. 1969. Definition of system. *Organizations: structure and behavior***,** 31.

HALL, R. W. 2003. *Handbook of transportation science,* Boston, Kluwer Academic Publishers.

HANSEN, A., PACHL, J., ALBRECHT, T., BRÜNGER, O., GOVERDE, R. M. P. & HUISMAN, D. 2008. *Railway Timetable and Traffic*, EUrail Press.

HANSEN, I. A. & PACHL, J. 2008a. Conclusions. *In:* HANSEN, I. A. & PACHL, J. (eds.) *Railway timetable and traffic.* Hamburg: Eurailpress.

HANSEN, I. A. & PACHL, J. (eds.) 2008b. *Railway timetable and traffic,* Hamburg: Eurailpress.

HARROD, S. 2007. *Railway capacity management and planning.* PhD, University of Cincinnati.

HARROD, S. 2009. Capacity factors of a mixed speed railway network. *Transportation Research Part E: Logistics and Transportation Review,* 45**,** 830-841.

HATANO, L. 2005. (January-February 2005) The international benchmarking project. *Railway Strategies*

HERBST, J. 2006. *The history of transportation,* Minneapolis, Twenty-First Century Books.

HERTEL, G. 1992. Die maximale Verkehrsleistung und die minimale Fahrplanempfindlichkeit auf Eisenbahnstrecken. *ETR,* 41**,** 665–671.

HOFF, A. 2007. Second stage DEA: Comparison of approaches for modelling the DEA score. *European Journal of Operational Research,* 181**,** 425-435.

HÖLLMÜLLER, J. & KLAHN, V. 2005. Implementation of the UIC 406 capacity calculation at Austrian railways. *In:* HANSEN, I. A., DEKKING, F. M., GOVERDE, R. M. P., B., H. & MEESTER, L. E. (eds.) *1st International Seminar on Railway Operations Modelling and Analysis.* Delft, The Netherlands.

HYLEN, B. 1998. An examination of rail infrastructure charges. *Final Report for the European Commission, DG VII.* London: NERA.

IBM. 2011. *IBM developers website* [Online]. IBM. Available: http://www.ibm.com/developerworks/ [Accessed 2011/06/08/.

INTERNATIONAL COUNCIL ON SYSTEMS ENGINEERING 2004 *Systems Engineering Handbook.*

INTERNATIONAL UNION OF RAILWAYS 2009. *Compendium on ERTMS: European Rail Traffic Management System*, Eurail Press.

JIA, W., MAO, B., HO, T., LIU, H. & YANG, B. 2009. Bottlenecks Detection of Track Allocation Schemes at Rail Stations by Petri Nets. *Journal of Transportation Systems Engineering and Information Technology,* 9**,** 136-141.

JIANXIN, Y. & HANSEN, I. A. 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transportation Research, Part B (Methodological),* 41**,** 202-217.

JOHNSON, D. & NASH, C. 2008. Charging for Scarce Rail Capacity in Britain: A Case Study. *Review of Network Economics,* 7**,** 53-76.

JOHNSON, P. L. 2000. *ISO 9000 : the year 2000 and beyond,* New York, McGraw-Hill.

KAAS, A. H. 1998. *Methods to Calculate Capacity of Railways (Metoder til beregning af jernbanekapacitet).* PhD, Technical University of Denmark.

KAHAN, A. M. 1979. Railway capacity analysis and related methodology Ottawa/Hull, Canada: Canadian transport commission.

General Systems Theory: Applications for Organization and Management, 1972. Directed by KAST, F. E. & ROSENZWEIG, J. E.: Academy of Management.

KELLER, P. A. 2001. *Six Sigma Deployment: A Guide for Implementing Six Sigma in Your Organization*, QA Pub., LLC.

KELLER, P. A. 2011. *Six sigma demystified*, McGraw-Hill

KHADEM SAMENI, M., DINGLER, M., PRESTON, J. & BARKAN, C. Profit-Generating Capacity for a Freight Railroad. 90th Annual Meeting of the Transportation Research Board, 2011a Washington, D.C., USA.

KHADEM SAMENI, M., LANDEX, A. & PRESTON, J. 2010a. Revising the UIC 406 method: revenue generating capacity. *American Society of Mechanical Engineers Joint Rail Conference.* Urbana, Illinois, U.S.A.

KHADEM SAMENI, M., LANDEX, A. & PRESTON, J. 2011b. Developing the UIC 406 method for capacity analysis. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

KHADEM SAMENI, M., LANDEX, A. & PRESTON, J. Application of Failure Mode and Effects Analysis for Controlling Railway Capacity Utilisation. *92nd Annual Meeting of the Transportation Research Board*, 2013 Washington, D.C., USA.

KHADEM SAMENI, M. & PRESTON, J. 2012a. A Novel Approach to Rail Station Capacity Analysis (under review). *Transportation Research Part A: Policy and Practice*.

KHADEM SAMENI, M. & PRESTON, J. Value for Rail Capacity: Assessing operators' efficiency in the United Kingdom. *91st Annual Meeting of the Transportation Research Board*, 2012b Washington, D.C., USA.

KHADEM SAMENI, M. & PRESTON, J. 2012c. Value for Rail Capacity: Assessing operators' efficiency in the United Kingdom. *Transportation Research Record (under publication).*

KHADEM SAMENI, M., PRESTON, J. & ARMSTRONG, J. 2010b. Railway capacity challenge: measuring and managing. *42nd Annual Conference of the Universities' Transport Study Group.* Plymouth, UK.

KHADEM SAMENI, M., PRESTON, J. & ARMSTRONG, J. 2010c. Railway Capacity Challenge: Measuring and Managing in Britain. *American Society of Mechanical Engineers Joint Rail Conference.* Urbana, Illinois, U.S.A.

KIERAN, M. 2001. Methods and practices in pricing railway track access. Kieran Management Advisory Services Ltd.

KONTAXI, E. K. & RICCI, S. 2011. Calculation of railway network capacity: comparing methodologies for lines and nodes. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

KRAAIJEVELD 2009. Making room on the rails. *Growth and capacity challenge-an international perspective.* London, UK.

KROON, L. G., EDWIN ROMEIJN, H. & ZWANEVELD, P. J. 1997. Routing trains through railway stations: complexity issues. *European Journal of Operational Research,* 98**,** 485-498.

KRUEGER, H. Parametric modelling in rail capacity planning. 1999 winter simulation conference, 1999. 1194-2000.

KUCKELBERG, A., WENDLER, E. & GRÖGER, T. 2011. A UIC 406-compliant, practically relevant capacity-consumption evaluation algorithm. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

LAI, Y. C. 2008. *Increasing railway efficiency and capacity through improved operations, control and planning.* PhD, University of Illinois at Urbana-Champaign.

LAI, Y. C. & BARKAN, C. P. L. 2009a. Enhanced Parametric Railway Capacity Evaluation Tool. *Transportation Research Record - Journal of Transportation Research Board,* 2117 33-40.

LAI, Y. C. & BARKAN, C. P. L. 2009b. An enhanced parametric railway capacity evaluation tool (RCET). *88th Transportation Research Board annual meeting.* Washington, USA.

LANDEX, A. Capacity statement for railways. *Annual Transport Conference at Aalborg University,*Trafikdage, 2007 Aalborg, Denmark.

LANDEX, A. 2008. *Methods to estimate railway capacity and passenger delays.* PhD, Technical university of Denmark.

LANDEX, A. 2011. Station capacity. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

LANDEX, A., KAAS, A. H., SCHITTENHELM, B. & SCHNEIDER TILLI, J. Evaluation of railway capacity. *In:* HANSEN, L. G., NIELSEN, L. D., NIELSEN, O. A. & HELS, T., eds. Annual Transport Conference at Aalborg University, 2006 Aalborg, Denmark. 232.

LANDEX, A., SCHITTENHELM, B., KAAS, A. H. & SCHNEIDER TILLI, J. Capacity Measurement with the UIC 406 capacity method. *In:* ALLAN, J., ARIAS, E. & BREBBIA, C. A., eds. 11th International Conference on Computers in Railways, 2008. WIT Press, Great Britain.

LARSON, A. 2003. *Demystifying six sigma: a company-wide approach to continuous improvement*, AMACOM/American Management Association.

LARSSON, P., DEKKER, S. W. A. & TINGVALL, C. 2010. The need for a systems theory approach to road safety. *Safety Science,* In Press, Corrected Proof.

LINDFELDT, O. 2007. *Quality on single-track railway lines with passenger traffic: analytical model for evaluation of crossing stations and partial double-tracks.* KTH.

LINDFELDT, O. Analysis of capacity on double-track railway lines. 3rd International Seminar on Railway Operations Modelling and Analysis, 2009 Zurich, Switzerland.

LINDFELDT, O. 2010. *Railway operation analysis.* PhD, KTH.
LINDNER, T. 2011. Applicability of the analytical compression method for evaluating node capacity. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.
LINDNER, T. & PACHL, J. 2010. Recommendations for Enhancing UIC Code 406 Method to Evaluate Railroad Infrastructure Capacity. *89th Transportation Research Board annual meeting.* Washington, USA.
LITTERER, J. A. 1969. *Organization: Systems, control and adaption,* New York, Wiley.
LOIZIDES, J. & TSIONAS, E. G. 2004. Dynamic distributions of productivity growth in European Railways. *Journal of Transport Economics and Policy (JTEP),* 38**,** 45-75.
LOVELL, K., BOUCH, C., SMITH, A., NASH, C., ROBERTS, C., WHEAT, P., GRIFFITHS, C. & SMITH, R. 2011. Introducing new technology to the railway industry: system-wide incentives and impacts. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 225**,** 192-201.
LOWE, D. 2005. *Intermodal freight transport* [Online]. Amsterdam: Elsevier Butterworth-Heinemann. Available: http://www.sciencedirect.com/science/book/9780750659352.
LOYD, N., JENNINGS, L., SINIARD, J., SPAYD, M., HOLDEN, A. & RITTENHOUSE, G. 2009. Application of Lean Enterprise to Improve Seaport Operations. *Transportation Research Record: Journal of the Transportation Research Board,* 2100**,** 29-37.
LOZANO, S., VILLA, G. & CANCA, D. 2011. Application of centralised DEA approach to capital budgeting in Spanish ports. *Computers & Industrial Engineering,* 60**,** 455-465.
LUSBY, R., LARSEN, J., EHRGOTT, M. & RYAN, D. 2009. Railway track allocation: models and methods. *OR Spectrum*.
MACBETH, D. K. & DE OPACUA, A. I. 2010. Review of Services Science and possible application in rail maintenance. *European Management Journal,* 28**,** 1-13.
MACDERMOTT, R. E., BEAUREGARD, M. R., MCDERMOTT, R. E. & MIKULAK, R. J. 2000. *The basics of FMEA,* Portland, Or., Productivity.
MANHEIM, M. L. 1979. *Fundamentals of transportation systems analysis,* Cambridge, Mass., MIT Press.
MANSON, S. M. 2001. Simplifying complexity: a review of complexity theory. *Geoforum,* 32**,** 405-414.
MARINOV, M., PACHL, J., LAUTALA, P., MACÁRIO, R., REIS, V. & EDWARDS, J. R. 2011. Policy-Oriented Measures for Tuning and Intensifying Rail Higher Education on both Sides of the Atlantic *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.
MARINOV, M. & VIEGAS, J. A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network. *Simulation Modelling Practice and Theory,* 19**,** 516-539.
MARLOW, P. B. & PAIXÃO CASACA, A. C. 2003. Measuring lean ports performance. *International Journal of Transport Management,* 1**,** 189-202.

MATTSON, L. 2004. Train service reliability: A survey of methods for deriving relationship for train delays. *Research Report, Unit for Transport and Location Analysis, Royal Institute of Technology, Stockholm, Sweden*.

MCMULLEN, T. B. 1998. *Introduction to the theory of constraints management system,* Boca Raton, FL, The St. Lucie Press.

MEADOWS, D. H. & WRIGHT, D. 2008. *Thinking in systems : a primer,* White River Junction, Vt., Chelsea Green Pub.

MERKERT, R., SMITH, A. S. J. & NASH, C. A. 2010. Benchmarking of train operating firms–a transaction cost efficiency analysis. *Transportation Planning and Technology,* 33**,** 35-53.

MILINKOVIC, S., MARKOVIC, M., VESKOVIC, S., IVIC, M. & PAVLOVIC, N. 2011. A petri net based simulation model of a railway junction system. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

MODERN RAILWAYS. 2011. *Golden Spanners award* [Online]. Sittingbourne. Available: http://www.modern-railways.com [Accessed 30/11/2011.

MOORE, R. 2007. *Selecting the right manufacturing improvement tools: what tool? when?,* Burlington, MA, Elsevier Butterworth-Heineman.

MUNDREY, J. S. 2000. *Railway track engineering,* New Delhi, Tata McGraw-Hill Publication.

NASH, C. 1982. *Economics of public transport,* London, Longman.

NASH, C., COULTHARD, S. & MATTHEWS, B. 2004. Rail track charges in Great Britain--the issue of charging for capacity. *Transport Policy,* 11**,** 315-327.

NASH, C. & JOHNSON, D. 2005. Issues in defining and measuring railway capacity. *Scoping study for scarcity charges.* Leeds: Institute for Transport Studies, University of Leeds.

NASH, C. & PRESTON, J. 1994. Railway performance—how does Britain compare? *Public Money & Management,* 14**,** 47-53.

NASH, C. A. & SHIRES, J. D. 2000. Benchmarking of European railways : an assessment of current data and recommended indicators. *European Conference of Ministers of Transport and European Commission.Transport benchmarking. Methodologies, applications and data needs.* Paris: OECD.

NETWORK RAIL 2006a. Network Rail Initial Strategic Business Plan - Control Period 4 London.

NETWORK RAIL 2006b. South West Main Line Route Utilisation Strategy. London,UK.

NETWORK RAIL 2009a. Route Plans - South West Main Line. London.

NETWORK RAIL. 2009b. *Route Utilisation Strategies* [Online]. London,UK. Available: http://www.networkrail.co.uk/aspx/4449.aspx [Accessed January 20, 2010.

NETWORK RAIL 2009c. Route Utilisation Strategies : Technical Guide London,UK.

NETWORK RAIL. 2010a. London. Available: http://www.networkrail.co.uk/aspx/1034.aspx [Accessed 16/08/2010.

NETWORK RAIL 2010b. Annual Return 2010 Report. *Available at http://www.networkrail.co.uk/. (accessed 20/04/2011).*

NETWORK RAIL. 2010c. *General facts* [Online]. London. Available: http://www.networkrail.co.uk/ [Accessed 13/09/2010.

NETWORK RAIL. 2010d. *Last period performance* [Online]. London. Available: http://www.networkrail.co.uk/aspx/742.aspx.

NETWORK RAIL 2010e. Prioritising investment to support our economy: A new approach to appraisal methodology. London.

NETWORK RAIL 2010f. Rules of the plan 2010. *Former Southern region.* London.

NETWORK RAIL 2011a. Britain Relies on Rail. London.

NETWORK RAIL. 2011b. *'How do we measure up?'* [Online]. London. Available: http://www.networkrail.co.uk/aspx/699.aspx [Accessed 22/06/2011.

NETWORK RAIL. 2011c. *How we measure up* [Online]. London. Available: http://www.networkrail.co.uk/aspx/699.aspx [Accessed 25/07/2011.

NETWORK RAIL 2011d. London and South East Route Utilisation Strategy. London,UK.

NETWORK RAIL. 2011e. *Train Operating Companies* [Online]. London. Available: http://www.nationalrail.co.uk/tocs_maps/tocs/ [Accessed 24/06/2011.

NETWORK RAIL MEDIA CENTRE. 2006. *Facts and Figures: Dealing with the weather* [Online]. London. Available: http://www.networkrailmediacentre.co.uk/.

NILSSON, J.-E. 2002. Towards a welfare enhancing process to manage railway infrastructure access. *Transportation Research Part A: Policy and Practice,* 36**,** 419-436.

OCCASION CONSORTIUM 2011. Overcoming Capacity Constraints - A Simulation Integrated with Optimisation for Nodes (OCCASION) - Literature Review Southampton: University of Southampton.

OFFICE OF RAIL REGULATION. 2008. *Station usage report and data for 2007-08* [Online]. London. Available: http://www.rail-reg.gov.uk/ [Accessed 29/05/2011.

OFFICE OF RAIL REGULATION. 2010a. *National rail trends 2009-2010* [Online]. London, UK. Available: http://www.rail-reg.gov.uk [Accessed 10/05/2011.

OFFICE OF RAIL REGULATION 2010b. Network Rail Monitor :Quarter 4 of year I of CP4, 10 January 2010-31 March 2010 and annual assessment 2009-10. London, UK.

OFFICE OF RAIL REGULATION. 2011a. *Key issues for consultation: 2013 periodic review* [Online]. London. Available: http://www.rail-reg.gov.uk/ [Accessed 31/05/2011.

OFFICE OF RAIL REGULATION 2011b. Network Rail Monitor :Quarter 4 of year 2 of CP4, 9 January 2011-31 March 2011 and annual assessment 2010-11. London, UK.

OUM, T. H. & YU, C. 1994. Economic efficiency of railways and implications for public policy: a comparative study of the OECD countries' railways. *Journal of Transport Economics and Policy,* 28**,** 121-138.

PACHL, J. 2008. Timetable design principles. *In:* HANSEN, I. A. & PACHL, J. (eds.) *Railway timetable and traffic.* Hamburg: Eurailpress.

PACHL, J. 2009. *Railway Operation and Control,* Mountlake Terrace (USA), VTD Rail Publishing.

PARNELL, G. S., DRISCOLL, P. J. & HENDERSON, D. L. 2008. *Decision making in systems engineering and management,* Hoboken, N.J., Wiley-Interscience.

PEAT MARWICK AND PARTNERS 1977. Passenger service costing, the impact of freight/passenger interference on capacity and cost. *Prepared for transport Canada.* Montreal.

PESTANA BARROS, C. & DIEKE, P. U. C. 2007. Performance evaluation of Italian airports: A data envelopment analysis. *Journal of Air Transport Management,* 13**,** 184-191.

POOLE, E. C. 1962. *Costs—A Tool for Railroad Management,* New York, Simmons Boardman.

PROFILLIDIS, V. A. 2006. *Railway Management and Engineering,* Aldershot, UK, Ashgate Publishing.

PROKOPY, J. C. & RUBIN, R. B. 1975. Parametric Analysis of Railway Line Capacity. *DOT-FR-5014-2, Federal Railroad Association, U.S. Department of Transportation.* Washington, DC.

PUDNEY, P., HOWLETT, P., MACKENZIE, S., HARRIS, D. & WARDROP, A. 2010. Corridor Capacity Analysis. Brisbane: CRC for Rail Innovation.

PYZDEK, T. 2003. *The Six Sigma project planner a step-by-step guide to leading a Six Sigma project through DMAIC* [Online]. New York: McGraw-Hill.

RADTKE, A. 2008. Infrastructure Modelling. *In:* HANSEN, I. A. & PACHL, J. (eds.) *Railway timetable and traffic.* Hamburg: Eurailpress.

RAILWAY SAFETY AND STANDARD BOARD 2002. ERTMS - Towards a better safer rail system. London: Railway Safety and Standard Board,.

RAMANATHAN, R. 2003. *An introduction to data envelopment analysis: a tool for performance measurement*, Sage Publications Pvt. Ltd.

RAND 2008. The State of U.S. Railroads: A Review of Capacity and Performance Data. Santa Monica, US.

RANGWALA, S. 1998. *Principles of railway engineering*, Charotar publishing house.

RANTANEN, K. & DOMB, E. 2007. *Simplified TRIZ: New Problem Solving Applications for Engineers and Manufacturing Professionals*, Auerbach Publications.

RAY, S. C. 2004. *Data envelopment analysis : theory and techniques for economics and operations research,* Cambridge, Cambridge university press.

REIGELUTH, C. M. 1983. Meaningfulness and instruction: Relating what is being learned to what a student knows. *Instructional Science,* 12**,** 197-218.

REVELLE, J. B., MORAN, J. W. & COX, C. A. 1998. *The QFD handbook,* New York; Toronto, Wiley.

RIVERA-TRUJILLO, C. 2004. *Measuring The Productivity and Efficiency of Railways (an international comparison).* unpublished thesis, University of Leeds.

RMCON. 2009. *RailSys information brochure* [Online]. Available: http://www.rmcon.de.

ROBERTS, C., SCHMID, F., CONNOR, P., RAMDAS, V. & SHARPE, J. 2010. A new railway system capacity model. Transport Research Laboratory

ROSS, P. J. 1996. *Taguchi techniques for quality engineering : loss function, orthogonal experiments, parameter and tolerance design,* New York, N.Y., McGraw-Hill.

SASSER, W. E., OLSEN, R. P., WYCKOFF, D. D. & ADMINISTRATION, H. U. G. S. O. B. 1978. *Management of service operations: Text, cases, and readings*, Allyn and Bacon Boston.

SAVE 2007. Value standard and body of knowledge. USA: SAVE International.

SCHEINKOPF, L. 1999. *Thinking for a Change: Putting the TOC thinking processes to use,* Boca Raton, USA, CRC Press.

SCHLECHTE, T. 2011. Railway Track Allocation - Simulation and Optimization. *4th International Seminar on Railway Operations Modelling and Analysis* Rome, Italy.

SCHOBEL, A. 2001. A model for the delay management problem based on mixed-integer-programming. *Electronic Notes in Theoretical Computer Science,* 50**,** 1-10.

SCHWANHÄUßER, W. 1974. *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn.* PhD, Instituts RWTH Aachen.

SEDDON, J. 2005. *Freedom from command & control rethinking management for lean service,* New York, Productivity Press.

SEIFORD, L. M. & ZHU, J. 2002. Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research,* 142**,** 16-20.

SKYTTNER, L. 2001. *General systems theory : ideas & applications,* Singapore; River Edge, N.J., World Scientific.

SMITH, A. & WHEAT, P. A Quantitative study of Train Operating Companies cost and efficiency trends 1996 to 2006: Lessons for future franchising policy. 2007.

SMITH, D. 2009. Timetabling is the culprit in 'empty seats' problem. *Rail Professional.* Cambridge Rail Professional Ltd.

STAMATIS, D. H. 2003. *Failure mode and effect analysis : FMEA from theory to execution,* Milwaukee, Wisc, ASQ Quality Press.

SWAMIDASS, P. M. 2000. *Encyclopedia of production and manufacturing management,* Boston, Kluwer Academic.

THANASSOULIS, E. 2001. *Introduction to the theory and application of data envelopment analysis : a foundation text with integrated software,* Norwell, Mass., Kluwer Academic Publ.

TONGZON, J. 2001. Efficiency measurement of selected Australian and other international ports using data envelopment analysis. *Transportation Research Part A: Policy and Practice,* 35**,** 107-122.

TORNQUIST, J. 2005. Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms. *5th Workshop on Algorithmic Methods and Models for Optimization of Railways.*

TÖRNQUIST, J. 2007. Railway traffic disturbance management--An experimental analysis of disturbance complexity, management objectives and limitations in planning horizon. *Transportation Research Part A: Policy and Practice,* 41**,** 249-266.

TRAINS: THE MAGAZINE OF RAILROADING. 2012. *Railroading Glossary* [Online]. [Accessed 25/06/2012.

TRANSPORTATION RESEARCH BOARD 2003. Rail Transit Capacity. *Transit capacity and quality of service manual.* Washington, US.

TRUSCOTT, W. 2003. *Six sigma: continual improvement for business: a practical guide*, Butterworth-Heinemann.

UIC 2004. Capacity (UIC code 406). Paris, France: International Union of Railways (UIC).

UIC. 2008. *Influence of ETCS on Line Capacity* [Online]. Paris: International Union of Railways. Available: http://www.uic.org/etf/publication/publication-detail.php?code_pub=510 [Accessed 12/09/2012.

UIC 2009. Capacity survery result. *http://www2.uic.asso.fr/spip.php?article2043*. Paris: International Union of Railways (UIC).

UIC 2011a. Rail and Sustainable Environment. Paris: International Union of Railways (UIC).

UIC 2011b. Rail and sustainable transport. Paris: International Union of Railways (UIC).

UIC. 2011c. *Railway Statistics - Synopsis Year 2010* [Online]. Paris: International Union of Railways (UIC). Available: http://www.uic.org [Accessed 7/07/2011.

UNIFE. 2009a. *ERTMS LEVELS* [Online]. Brussels: Association of the European Rail Industry. Available: http://www.ertms.com/2007v2/factsheets.html [Accessed 02/08/2010.

UNIFE. 2009b. *INCREASING INFRASTRUCTURE CAPACITY :HOW ERTMS IMPROVES RAILWAY PERFORMANCE* [Online]. Brussels: Association of the European Rail Industry. Available: http://www.ertms.com/2007v2/factsheets.html [Accessed 02/08/2010.

VAN DE VELDE, D. M. 1999. Organisational forms and entrepreneurship in public transport: classifying organisational forms. *Transport Policy,* 6**,** 147-157.

VROMANS, M. J., DEKKER, R. & KROON, L. 2004. Reliability and Heterogeneity of Railway Services. *European Journal of opeational research,* 172**,** 647-665.

VROMANS, M. J. C. M., DEKKER, R. & KROON, L. G. 2006. Reliability and heterogeneity of railway services. *European Journal of Operational Research,* 172**,** 647-665.

VUCHIC, V. R. & NEWELL, G. F. 1968. Rapid transit interstation spacings for minimum travel time. *Transportation Science,* 2**,** 303-339.

WAHLBORG, M. 2005. Applicationof the UIC Capacity leaflet at Banverket *In:* HANSEN, I. A., DEKKING, F. M., GOVERDE, R. M. P., B., H. & MEESTER, L. E. (eds.) *1st International Seminar on Railway Operations Modelling and Analysis.* Delft, The Netherlands.

WAKOB, H. 1985. *bleitung eines generellen Wartemodells zur Ermittlung der planmässigen Wartezeiten im Eisenbahnbetrieb unter besonderer Berücksichtigung der Aspekte der Leistungsfähigkeit und Anlagenbelastung.* PhD, Institutes der RWTH Aachen.

WALKER, J. T. & MADDAN, S. 2009. *Statistics in criminology and criminal justice : analysis and interpretation,* Sudbury, Mass., Jones and Bartlett Publishers.

WANG, J. X. 2010. *Lean manufacturing: business bottom-line based,* Boca Raton, CRC press.

WANG, Q. 2008. Ideology and Practice of Systems Engineering in Multi-Modal Transport Planning. *Journal of Transportation Systems Engineering and Information Technology,* 8**,** 11-16.

WARDMAN, M., SHIRES, J., LYTHGOE, W. & TYLER, J. 2004. Consumer benefits and demand impacts of regular train timetables. *International Journal of Transport Management,* 2**,** 39-49.

WASSON, C. S. 2006. *System analysis, design, and development : concepts, principles, and practices,* Hoboken, NJ, Wiley-Interscience.

WATSON, R. 2008. *Train planning in a fragmented railway.* PhD, Loughborough University.

WCM. 2011. *Process Sigma and DPMO calculator* [Online]. Available: http://world-class-manufacturing.com/Sigma/level.html [Accessed 25/07/2011.

WENDLER, E. 1999. *Analytische Berechnung der planmässigen Wartezeiten bei ansynchroner Fahrplankonstruktion.* Verkehrswiss. Inst. der Rheinisch-Westfälischen Techn. Hochsch.

WENDLER, E. 2007. Influence of ETCS on line capacity. *UIC ERTMS world conference.* Berne.

WENDLER, E. 2008. Queueing. *In:* HANSEN, I. A. & PACHL, J. (eds.) *Railway Timetable and Traffic.* Hamburg: Eurailpress.

WESTWOOD, J. 2009. *The Historical Atlas of World Railways*, Cartographica Press.

WETZEL, H. 2008. *European railway deregulation: the influence of regulatory and environmental conditions on efficiency.* University of Luumlneburg.

WISNER, J. D., TAN, K.-C. & LEONG, G. K. 2009. *Principles of supply chain management : a balanced approach,* Mason, OH, South-Western Cengage Learning.

WOMACK, J. P., JONES, D. T. & ROOS, D. 2007. *The machine that changed the world : the story of lean production,* New York, NY Free Press.

XU, K., TANG, L. C., XIE, M., HO, S. L. & ZHU, M. L. 2002. Fuzzy assessment of FMEA for engine systems. *Reliability Engineering & System Safety,* 75**,** 17-29.

YAN, Z. X. & YAN, X. H. A Revolution in the Field of Systems Thinking-A Review of Checkland's System Thinking. *Systems Research and Behavioral Science,* 27**,** 140-155.

YU, M.-M. 2008. Assessing the technical efficiency, service effectiveness, and technical effectiveness of the world's railways through NDEA analysis. *Transportation Research Part A: Policy and Practice,* 42**,** 1283-1294.

YUAN, J. 2008. Satistical analysis of train delays. *In:* HANSEN, I. A. & PACHL, J. (eds.) *Railway timetable and traffic.* Hamburg: Eurailpress.

YUAN, J. & HANSEN, I. A. 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transportation Research Part B: Methodological,* 41**,** 202-217.

YUAN, J., NETHERLANDS RESEARCH SCHOOL FOR TRANSPORT, I. & LOGISTICS. 2006. *Stochastic modelling of train delays and delay propagation in stations.* Netherlands TRAIL Research School.

ZEMP, S., STAUFFACHER, M., LANG, D. J. & SCHOLZ, R. W. 2011. Generic functions of railway stations--A conceptual basis for the development of common system understanding and assessment criteria. *Transport Policy,* 18**,** 446-455.

ZERAFAT ANGIZ, M., MUSTAFA, A. & EMROUZNEJAD, A. 2010. Ranking efficient decision-making units in data envelopment analysis using fuzzy concept. *Computers & Industrial Engineering,* 59**,** 712-719.

ZHONG, C. 2007. *The potential for moving containers by intermodal transports in the UK.* PhD, University of Southampton.

ZWANEVELD, P., KROON, L., ROMEIJN, H., SALOMON, M., DAUZERE-PERES, S., VAN HOESEL, S. & AMBERGEN, H. 1996. Routing trains through railway stations: Model formulation and algorithms. *Transportation Science,* 30**,** 181-194.

ZWANEVELD, P. J., KROON, L. G. & VAN HOESEL, S. P. M. 2001. Routing trains through a railway station based on a node packing model. *European Journal of Operational Research,* 128**,** 14-33.

Appendix 1-Glossary

Major terms used in the thesis and their meaning are summarised below. More details for each term can be found within the thesis in the relevant section.

Term	Meaning
Available capacity	The difference between used and practical capacity.
Blocking stairway	A graph displaying the blocking time of all block sections that a train passes into a time-distance graph
Capacity	(Track) capacity is the traffic volume a line can handle at a given level of service which depends on the way it is utilised. Four main types of capacity can be defined: theoretical capacity, practical capacity, used capacity and available capacity.
Capacity utilisation	The act of using capacity.
CUI method	Capacity utilisation index which is the main method for analysing capacity utilisation in Great Britain.
Data envelopment analysis (DEA)	A holistic method for evaluating relative performance of units which was initially developed by Farrell in 1975.
Economy	How cheaply inputs are provided
Effectiveness	The extent of delivering desired outcomes by the cost of producing outputs
Efficiency	How much output is produced by using inputs

Term	Meaning
ERTMS	The European Rail Traffic Management System that is an initiative backed by the European Union to enhance cross-border interoperability of railways.
Failure mode and effect analysis (FMEA)	A systematic way of identifying and prioritising risks according to their occurrence, severity and detection
Heterogeneity	Difference between the running time and stopping pattern of trains
Lean thinking	Lean thinking originated from the Toyota company in the early 1990s. It aims to improve manufacturing and service processes by preserving value by reducing 'muda'. .
Macro capacity utilisation	Quantity of discrete steps to use railway capacity
Micro capacity utilisation	Quantity of discrete steps to use railway capacity
Muda	Muda is a Japanese word that in lean production terminology that means waste or any activity for which the customer is not willing to pay.
Practical capacity	Practical limit of traffic for a defined performance level
Public performance measure (PPM)	Public Performance Measure (PPM) is used in Great Britain and is the percentage of passenger trains that arrive at their destination on time (not later than 5 minutes for local services and not later than 10 minutes for inter-urban trains).
punctuality	The percentage of trains which arrive to/depart from/ pass a location with a delay less than a certain time in minutes
reliability	Reliability is the percentage of planned trains that were actually operated
Six Sigma	Six Sigma is a process improvement framework which was launched in the Motorola Company in 1987 and helped this company to make huge cost savings while improving quality. It is comprised from 5 steps of defining, measuring, analysing, improving and controlling.

Term	Meaning
stability	Stability of the timetable is its ability to compensate for delays and returning to the desired state
Stakeholder	Anyone or an organisation having a vested interest in a system and its outcomes
System	A set of objects together with relationships between the objects and between their attributes
Theoretical capacity	Upper bound of capacity
UIC	The French and widely used abbrivieation for the International Union of Railways based in Paris.
UIC 406 method	A method suggested by the International Union of Railways to estimate capacity utiliaiton by compressing the timetable
Used capacity	The share of [track] capacity that is consumed by the traffic volume.
value	Expression of the relationship between function and resources where function is measured by the performance requirements of the customer (such as quality of service) and resources are measured in materials, labour, price, time, etc. required to accomplish that function

Appendix 2– Major Works on Train Timetabling Problem by Operations Research (OCCASION consortium, 2011)

Authors	Problem Details	Periodicity	Mathematical Model Details	Solution Approach	Case Study
Szpigel (1973)	Single track	Acyclic	MILP model, Disjunctive constraints	Branch & Bound	N/A
Jovanovic and Harker (1991)	Tactical scheduling of freight railroad traffic	Acyclic	N/A	Branch & Bound	Major line (24 lines, 100 trains)
Brännlund et al. (1998)	Single track	Acyclic	TTP model, Discretisation of the time horizon	Lagrangian relaxation	Swedish National Railway (26 trains, 17 stations)
Oliveira and Smith (2000)	Single track	N/A	Job Shop model	Constraint Programming	19 real life problems by Higgins (1997)

Authors	**Problem Details**	**Periodicity**	**Mathematical Model Details**	**Solution Approach**	**Case Study**
Caprara et al. (2002)	Single one way track	Acyclic	Graph theoretic formulation	Lagrangian relaxation	Italian railway companies Ferriovie dello Stato SpA, Ansaldo Segnalamento Ferroviario SpA
Dorfman and Medanic (2004)	Single & double track	Acyclic	Discrete event model	Greedy travel advance strategy	Numerical example (36 trains, 31 stations)
Zhou and Zhong (2007)	Single track	Cyclic	RCPSP model	Branch & Bound: Lagrangian relaxation, Exact lower bounds, heuristic upper bounds	Laizhou to Shaowu, Fujian, China. (18 stations, 62 trains, 138 km)
Fischetti et al. (2009)	Single track	Acyclic	PESP	LP stochastic programming, robust optimisation	Italian railway company Trenitalia

Authors	**Problem Details**	**Periodicity**	**Mathematical Model Details**	**Solution Approach**	**Case Study**
Schrijver and Steenbeek (1994)	TTP	Cyclic	PESP	Constraint propagation	Netherlands Railways and ProRail (250 trains)
Nachtigall (1999)	TTP	Cyclic	Improved PESP model, less variables, better LP relaxation	N/A	N/A
Lindner (2000)	TTP	Cyclic	Improved PESP model, less variables, better LP relaxation	Commercial MIP solver	Intercity, Interegio Aggloregio Germany Netherlands
Peeters (2003)	TTP	Cyclic	Improved PESP model, less variables, better LP relaxation	CPLEX 7.5 solver	Dutch Intercity, Netherlands NorthHolland
Odijk (1996)	TTP	Cyclic	PESP model	PESP Cut Generation	Netherlands, Arnhem CS
Kroon and Peeters (2003)	Variable trip times,	Cyclic	PESP with variable trip times,	DSS DONS system, Hooghiemstra et al. (1999)	Dutch railway system (200-250 trains)

Authors	**Problem Details**	**Periodicity**	**Mathematical Model Details**	**Solution Approach**	**Case Study**
Liebchen and Mohring (2004)	Train line bundling, network planning, TTP	Cyclic	Various PESP models	N/A	Various examples from Dutch railway
Kroon et al (2005)	Single track, Minimise train avg delay	Cyclic	Stochastic Optimisation model	CPLEX 9.0 Solver	Dutch operator NS Reizigers
Cacchiani et al. (2008)	TTP	Cyclic	MILP model, variables for the timetables of each train	Column Generation, Branch & Cut & Price, Local search Heuristics	Rete Ferroviaria Italiana, Italian railway IM company
Borndörfer et al. (2005);Borndörfer and Schlechte (2007a, b)	TTP	Cyclic	IP models, LP relaxations can be solved in polynomial time	Column Generation	German railway

Authors	**Problem Details**	**Periodicity**	**Mathematical Model Details**	**Solution Approach**	**Case Study**
Wong et al. (2008)	Minimum passenger waiting times, dwell times, dispatch times, terminal turnaround times, adjustable train run times and headways.	Acyclic	MILP model, binary variables for waiting times	Branch & Bound, CPLEX 9.1 Optimisation-based heuristic for the model.	Rail Mass Transit Hong Kong

TTP: train timetabling problem
PESP: periodic event scheduling problem
MILP: mixed integer linear programming

References

Borndörfer, R., Grötschel, M., Lukac, S., Mitusch, M., Schlechte, T., Schultz, S., Tanner, A. (2005). An auctioning approach to railway slot allocation. Technical Report 05–45, Konrad-Zuse-Zentrum für Informationstechnik Berlin.

Borndörfer, R., Schlechte, T. (2007a). Models for railway track allocation. Technical Report 07–02, Konrad-Zuse-Zentrum für Informationstechnik Berlin.

Borndörfer, R., Schlechte, T. (2007b). Solving railway track allocation problems. Technical Report 07–20, Konrad-Zuse-Zentrum für Informationstechnik Berlin.

Brännlund, U., Lindberg, P.O., Nõu, A., Nilsson, J.E. (1998). Railway timetabling using Lagrangian relaxation. Transp Sci 32(4):358–369.

Cacchiani, V., Caprara, A., Toth, P. (2008). A column generation approach to train timetabling on a corridor. 4OR 6:125–142.

Caprara, A., Monaci, M., Toth, P., Guida, P.L. (2006). A Lagrangian heuristic approach to real-world train timetabling problems. Discrete Applied Mathematics 154, 738–753.

Dorfman, M.J., Medanic, J. (2004) Scheduling trains on a railway network using a discrete event model of railway traffic. Transp Res Part B 38(1):81–98.

Fischetti, M., Salvagnin, D., Zanette, A. (2009). Fast Approaches to Improve the Robustness of a Railway Timetable, Transportation Science 43:321-335

Jovanovic, D., Harker, P.T. (1991) Tactical scheduling of rail operations: The Scan I System. Transportation Science 25(1):46–64.

Kroon, L.G., Peeters, L.W.P. (2003). A variable trip time model for cyclic railway timetabling. Transportation Science 37, 198–212.

Kroon, L.G., Dekker, R., Vromans, M.J.C.M. (2005). Cyclic railway timetabling: A stochastic optimization approach. Technical Report ERS-2005-051-LIS, Erasmus University Rotterdam.

Liebchen, C., Möhring, R. (2004). The modelling power of the periodic event scheduling problem: railway timetables—and beyond. Technical Report 2004/20, Technische Universität Berlin, Institut fur Mathematik.

Lindner, T. (2000). Train schedule optimization in public rail transport. PhD thesis, TU Braunschweig.

Nachtigall, K., Voget, S. (1996). A genetic algorithm approach to periodic railway synchronization. Computers & Operations Research 23, 453–463

Odijk, M.A. (1996.) A constraint generation algorithm for the construction of periodic timetables. Transportation Research B 30(6):455–464.

Oliveira, E., Smith, B. M. (2000). A Job-Shop Scheduling Model for the Single-Track Railway Scheduling Problem (Tech. Rep. No. 21). England: School of Computing, University of Leeds.

Peeters, L. (2003). Cyclic railway timetable optimization. PhD thesis, Erasmus Research Institute of Management.

Schrijver, A., Steenbeek, A. (1994). Timetable construction for Railned (in Dutch). Technical report, C.W.I. Center for Mathematics and Computer Science, Amsterdam.

Szpigel, B. (1973). Optimal train scheduling on a single line railway. In Operational Research 72, M. Ross (ed.) North Holland, Amsterdam, 344–351.

Wong, R. C. W., Yuen, T. W. Y., Fung, K. W., & Leung, J. M. Y. (2008). Optimizing Timetable Synchronization for Rail Mass Transit. Transportation Science, 42 (1), 57–69.

Zhou, X., Zhong, M. (2007). Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds. Transportation Research, Part B, 41, 320–341.

Appendix 3- Efficiency Scores for 96 Busiest Stations in the UK

	Name	Stage 1- Technical Efficiency (Output oriented- VRS): Macro capacity utilisation		Stage 2- Service Effectiveness (output oriented- VRS): Micro capacity utilisation		Stage 2- Service Effectiveness (Input oriented-VRS): Micro capacity utilisation	
		Efficiency score	**Rank**	**Efficiency score**	**Rank**	**Efficiency score**	**Rank**
1	London Waterloo	1.000	1	1.000	1	1.000	1
2	London Victoria	0.964	8	0.803	18	0.800	25
3	London Liverpool Street	0.962	9	0.675	22	0.668	38
4	London Bridge	1.000	1	1.000	1	1.000	1
5	London Charing Cross	0.771	14	1.000	1	1.000	1
6	London Euston	0.406	53	1.000	1	1.000	1
7	London Paddington	0.631	21	0.594	25	0.624	47
8	London Kings Cross	0.374	61	1.000	1	1.000	1
9	East Croydon	1.000	1	1.000	1	1.000	1
10	London Cannon Street	0.344	72	1.000	1	1.000	1
11	Manchester Piccadilly	0.741	15	0.324	59	0.364	91
12	Clapham Junction	1.000	1	1.000	1	1.000	1
13	Leeds	0.536	36	0.370	42	0.353	92
14	Birmingham New Street	0.585	27	1.000	1	1.000	1
15	London Fenchurch Street	0.605	24	0.692	21	0.786	28
16	Wimbledon	0.833	13	0.389	39	0.452	81
17	Reading	0.569	29	0.459	31	0.458	79
18	Brighton	0.632	20	0.426	34	0.451	84

	Name	**Stage 1- Technical Efficiency (Output oriented- VRS): Macro capacity utilisation**		**Stage 2- Service Effectiveness (output oriented- VRS): Micro capacity utilisation**		**Stage 2- Service Effectiveness (Input oriented-VRS): Micro capacity utilisation**	
		Efficiency score	**Rank**	**Efficiency score**	**Rank**	**Efficiency score**	**Rank**
19	Gatwick Airport	0.654	19	0.334	54	0.430	90
20	London Marylebone	0.347	71	0.746	20	0.823	24
21	Stratford (London)	0.480	43	0.346	49	0.434	88
22	Moorgate	1.000	1	1.000	1	1.000	1
23	Cardiff Central	0.626	23	0.281	71	0.351	93
24	Surbiton	0.504	41	0.544	26	0.642	45
25	Liverpool Central	1.000	1	0.217	86	0.349	94
26	Lewisham	0.542	34	0.917	16	0.933	16
27	Guildford	0.391	57	0.301	65	0.434	89
28	Woking	0.552	31	0.382	40	0.474	78
29	Chelmsford	0.514	39	0.643	24	0.795	26
30	Romford	0.450	47	0.337	51	0.518	71
31	Bristol Temple Meads	0.268	88	0.434	33	0.626	46
32	London Waterloo (East)	0.931	10	0.254	77	0.327	95
33	Cambridge	0.443	48	0.312	63	0.511	73
34	City Thameslink	0.601	25	1.000	1	1.000	1
35	Richmond	0.877	12	0.243	79	0.316	96
36	Sutton (Surrey)	0.545	32	0.271	75	0.483	77
37	York	0.250	92	0.346	48	0.549	64
38	Newcastle	0.198	96	0.469	29	0.763	31

	Name	Stage 1- Technical Efficiency (Output oriented- VRS): Macro capacity utilisation		Stage 2- Service Effectiveness (output oriented- VRS): Micro capacity utilisation		Stage 2- Service Effectiveness (Input oriented-VRS): Micro capacity utilisation	
		Efficiency score	**Rank**	**Efficiency score**	**Rank**	**Efficiency score**	**Rank**
39	St Albans	0.422	51	0.359	45	0.610	49
40	Bromley South	0.672	18	0.353	47	0.535	69
41	Nottingham	0.312	80	0.329	57	0.567	58
42	Sheffield	0.366	63	0.283	69	0.451	82
43	Southampton Central	0.394	55	0.311	64	0.544	66
44	Orpington	0.259	90	0.319	61	0.606	50
45	Finsbury Park	0.401	54	0.902	17	0.921	17
46	Ilford	0.313	79	0.435	32	0.696	34
47	Leicester	0.329	76	0.394	38	0.643	44
48	Twickenham	0.437	49	0.282	70	0.588	53
49	Slough	0.355	70	0.525	27	0.662	40
50	Balham	1.000	1	0.235	84	0.498	75
51	Oxford	0.455	46	0.322	60	0.566	59
52	Milton Keynes Central	0.223	95	0.371	41	0.793	27
53	Basingstoke	0.462	44	0.235	83	0.497	76
54	Colchester	0.363	65	0.274	73	0.603	51
55	Bath Spa	0.366	62	0.789	19	0.915	19
56	Watford Junction	0.319	77	0.204	90	0.450	85
57	Liverpool Lime Street	0.288	86	0.464	30	0.570	57
58	Tonbridge	0.389	58	0.335	53	0.654	42

	Name	Stage 1- Technical Efficiency (Output oriented- VRS): Macro capacity utilisation		Stage 2- Service Effectiveness (output oriented- VRS): Micro capacity utilisation		Stage 2- Service Effectiveness (Input oriented-VRS): Micro capacity utilisation	
		Efficiency score	**Rank**	**Efficiency score**	**Rank**	**Efficiency score**	**Rank**
59	Stevenage	0.362	66	0.274	73	0.543	67
60	Peterborough	0.366	64	0.334	55	0.583	54
61	Herne Hill	0.338	74	1.000	1	1.000	1
62	West Hampstead Thameslink	0.257	91	1.000	1	1.000	1
63	Sevenoaks	0.386	59	0.277	72	0.654	43
64	Raynes Park	0.585	26	0.166	94	0.447	87
65	Southend Victoria	0.237	93	1.000	1	1.000	1
66	Elephant & Castle	0.305	82	0.368	43	0.777	29
67	Highbury & Islington	0.361	67	0.246	78	0.563	61
68	Denmark Hill	0.278	87	0.362	44	0.824	23
69	Haywards Heath	0.461	45	0.238	82	0.550	63
70	Tunbridge Wells	0.316	78	1.000	1	1.000	1
71	Manchester Victoria	0.432	50	0.152	96	0.449	86
72	Winchester	0.529	37	0.313	62	0.680	37
73	Epsom	0.413	52	0.211	88	0.557	62
74	Preston	0.261	89	0.425	35	0.602	52
75	Tottenham Hale	0.886	11	0.194	92	0.536	68
76	Redhill	0.513	40	0.496	28	0.663	39
77	Barking	0.481	42	0.162	95	0.451	83
78	Luton	0.358	68	0.195	91	0.512	72

	Name	Stage 1- Technical Efficiency (Output oriented- VRS): Macro capacity utilisation		Stage 2- Service Effectiveness (output oriented- VRS): Micro capacity utilisation		Stage 2- Service Effectiveness (Input oriented-VRS): Micro capacity utilisation	
		Efficiency score	**Rank**	**Efficiency score**	**Rank**	**Efficiency score**	**Rank**
79	Eastbourne	0.379	60	0.650	23	0.918	18
80	Ealing Broadway	0.741	16	0.214	87	0.580	55
81	Norwich	0.289	85	0.255	76	0.577	56
82	Dartford	0.544	33	0.210	89	0.457	80
83	Chatham	0.570	28	0.288	66	0.704	33
84	Bedford	0.356	69	0.190	93	0.564	60
85	New Malden	0.555	30	0.334	56	0.770	30
86	Coventry	0.297	84	0.288	67	0.661	41
87	Shenfield	0.672	17	0.423	36	0.501	74
88	Streatham Common	0.626	22	0.287	68	0.746	32
89	Derby	0.302	83	0.359	46	0.689	36
90	Harpenden	0.342	73	0.241	80	0.691	35
91	Doncaster	0.307	81	0.328	58	0.533	70
92	Hither Green	0.237	94	0.337	52	0.828	22
93	Blackheath	0.542	35	0.345	50	0.830	21
94	Grays	0.332	75	0.414	37	0.871	20
95	Ashford International	0.394	56	0.229	85	0.618	48
96	Peckham Rye	0.528	38	0.239	81	0.545	65

Appendix 4- Detailed map of South West Main Line (Network Rail, 2006b)