

Aspects of Improving Punctuality

From Data to Decision in Railway Maintenance

Birre Nyström

Luleå University of Technology
Department of Civil, Mining and Environmental Engineering
Division of Operation and Maintenance Engineering

Doctoral Thesis

Division of Operation and Maintenance Engineering

Aspects of Improving Punctuality

From Data to Decision in Railway Maintenance

Birre Nyström

Luleå University of Technology

Division of Operation and Maintenance Engineering

“There are three things which make a nation great and prosperous: a fertile soil, busy workshops, easy conveyance for men and goods from place to place.”

–Francis Bacon

Preface

This thesis has its origins in the observation that trains, sometimes, are delayed. Studies into the causes for this and how punctuality might be improved are presented here. In such a large system as the railway, a wide range of methodologies from different disciplines might contribute to improved punctuality.

My thanks go to the supervisors for this work: Professor Uday Kumar with co-supervisor Professor Per Anders Akersten, as well as Professor Erik Höglund.

The author is grateful for the financial support from Banverket (the Swedish Rail Administration) and EU Regional Fund. The author also wants to thank the informants of A-Train, Banverket, CargoNet, EuroMaint, Green Cargo, Jernbaneverket, Luftfartsverket (the Swedish National Aviation Administration), MTAB, The Norwegian Post, Scandinavian Airlines, SJ, SL (Stockholm Local Traffic), SSAB, Strukton Rail and Tågkompaniet for sharing their experiences.

I would like to thank the following for making this thesis possible: Dr Mattias Holmgren for involving me in PhD studies, Dr Ulla Espling for sharing her world of railway knowledge, Professor Jørn Vatn for welcoming me to Trondheim, my co-researcher licentiate Stefan Niska, licentiate Rikard Granström for sharing his knowledge in signalling, Tomoko Okamura for Japanese translations, Stefan Karlsson, Fredrik Johansson and Lina Helgesson for help with questionnaires and the staff at the university library for excellent service. Thanks also go to Lars Sundholm and Torgny Johansson of Banverket for their insights and support. Last, but not least, I am grateful to Professor Bengt Klefsjö and Dr Peter Söderholm for timely guidance in my writing.

And of course I also want to express my gratitude to all my family and friends for being there.

Luleå, February 2008

Birre Nyström

Abstract

The increasing demand for transportation and sustainability makes railways attractive. The ongoing deregulation of state-owned railways means that many new organisations are entering the railway sector. Hence, reducing railway delays is increasingly important to many stakeholders, including passengers, freight customers, train operating companies, railway infrastructure managers and society in general. Therefore, the study of punctuality and its improvement is essential.

The purpose of the research presented in this thesis is to explore and describe information and requirements related to railway punctuality in order to support systematic improvements. The focus is on delay causes related to infrastructure maintenance. To fulfil the stated purpose, punctuality requirements, availability concepts, failure and delay data, as well as maintenance decisions, have been studied via theoretical and empirical approaches. Data was collected through interviews, document studies, archival analysis, observations and experiments.

It is found that punctuality requirements and performance are currently expressed in many, hardly commensurable, ways. Hence, it is difficult to compare punctuality data from different railways. This is further complicated by the fact that delay attribution is inconsistently performed. It is also found that there is a lack of data on train traffic and infrastructure, for example, causes of delays. Although the consistency regarding ranking of decision-making criteria is rather high, the consistency of maintenance decisions is rather low. In addition, there are many interacting causes affecting punctuality, including infrastructure, timetable, rolling stock, weather and personnel. It is also found that even though unpunctuality might be explained by unavailability of some parts of the railway system, the concept of availability is not well-established and agreed upon within the railway sector.

Based on the research findings, it is proposed that punctuality should be treated as the extent to which an event takes place when agreed, for example, the agreement between a passenger and a train operating company concerning the arrival of a train at a certain time. A number of availability measures for railway are also proposed, partly based on analogies to the power industry. Furthermore, the developed and applied methodologies, based on vignettes and the Analytic Hierarchy Process (AHP), are proposed to support punctuality improvements.

To summarise, based on the results of this research, it is possible to improve data collection and recording, select suitable indicators and increase the awareness of the grounds on which decisions are made, all of which contribute to improved punctuality.

Keywords: Punctuality, Maintenance, Improvement, Railway, Delay reporting system, Failure reporting system, Decisions, Availability

Sammanfattning (Summary in Swedish)

Förbättrad punktlighet – från data till beslut i järnvägsunderhåll

Den ökande efterfrågan på transporter, helst med låg miljöbelastning, gör järnvägen till ett attraktivt alternativ. I den pågående omregleringen av statliga järnvägar gör många nya aktörer entré i branschen. Att minska järnvägsförseningarna blir allt viktigare för många intressenter, som passagerare, godskunder, tågoperatörer, infrastrukturförvaltare och samhället i stort. Därför är det mycket viktigt att studera de faktorer som påverkar punktligheten.

Syftet med forskningen som presenteras i denna avhandling är att undersöka och beskriva information och krav relaterade till punktlighet, i syfte att stödja systematiskt förbättringsarbete. Fokus ligger på förseningsorsaker relaterade till infrastrukturunderhåll. För att uppfylla syftet har punktlighetskrav, tillgänglighetskoncept, data för fel och förseningar, liksom underhållsbeslut, studerats teoretiskt och empiriskt. Databesamling har skett med intervjuer, dokumentstudier, arkivmaterial, observationer och experiment.

I avhandlingen visas att punktlighetsprestation och -krav idag uttrycks på många, knappast jämförbara, sätt. Således är det svårt att jämföra punktlighetsdata, eftersom olika järnvägar presenterar punktlighet på skilda sätt. Detta försvåras dessutom av att förseningsrapporteringen utförs inkonsistent. Det visas också att det saknas data om tågtrafik och infrastruktur som skulle kunna underlätta för förbättringar, även beträffande förseningsorsaker. Även om konsistensen för rangordning av kriterierna för beslutsfattande är god, är konsistensen för underhållsbeslut ganska låg. Dessutom påverkar många interagerande faktorer punktligheten, till exempel infrastruktur, tidtabell, rullande materiel, väder och personal. Det visas också att trots att opunktlighet kan förklaras av otillgänglighet hos enheter i järnvägssystemet, så saknar järnvägssektorn ett allmänt accepterat tillgänglighetsbegrepp.

Baserat på forskningsresultaten föreslås att punktlighet definieras som i vilken utsträckning en händelse äger rum när det är överenskommet, till exempel överenskommelsen mellan en passagerare och en tågoperatör beträffande händelsen att ett tåg ankommer vid en viss tidpunkt. Ett antal tillgänglighetsmått för järnvägen föreslås också, delvis baserade på analogier med elkraftsindustrin. De arbetssätt som utvecklats och tillämpats, baserade på scenarier och AHP (Analytic Hierarchy Process), föreslås användas för att stödja punktlighetsförbättringar.

Sammanfattningsvis, baserat på resultaten av denna forskning, är det möjligt att förbättra databesamlingen och registreringen, välja lämpliga indikatorer och öka medvetenheten om använda beslutsgrunder, vilket bidrar till förbättrad punktlighet.

Nyckelord: Punktlighet, Underhåll, Förbättring, Järnväg, System för förseningsrapportering, System för felrapportering, Beslut, Tillgänglighet

Contents

List of Appended Papers	ix
1 Introduction.....	1
1.1 The Swedish Railway System and its Stakeholders	1
1.2 Unpunctuality Problems and Improvement	3
1.3 Purpose, Scope and Delimitations of Research	6
1.4 Thesis Structure	7
2 Theoretical Frame of Reference.....	9
2.1 Time, Punctuality and Associated Terms	9
2.2 How a Delay to One Train Inflicts Delays on Other Trains	14
2.3 Unpunctuality Cost.....	15
2.4 Maintenance.....	18
2.5 Factors Influencing the Unpunctuality Cost	21
2.6 Explanations of Unpunctuality	24
2.7 Comparisons of Punctuality.....	31
2.8 Unpunctuality Cost versus Maintenance Cost.....	32
3 Research Design.....	35
3.1 Research Strategies.....	35
3.2 Data Collection	36
3.3 The Concepts of Reliability and Validity	36
3.4 Methodological Choices in this Thesis.....	37
4 Description of Appended Papers	43
4.1 Description of Paper I.....	44
4.2 Description of Paper II	45
4.3 Description of Paper III	47
4.4 Description of Paper IV	48
4.5 Description of Paper V	49
4.6 Description of Paper VI.....	50
4.7 Description of Paper VII.....	51
4.8 Description of Paper VIII	52
5 Discussion and Conclusions	55
5.1 Discussion and Conclusions Related to Research Question A	55
5.2 Discussion and Conclusions Related to Research Question B	57
5.3 Discussion and Conclusions Related to Research Question C	59
5.4 Discussion and Conclusions Related to Research Question D	60
5.5 Discussion and Conclusions Related to Research Question E	63
5.6 Some Aspects of Improving Punctuality	65
6 Suggestions for Improvements	69
6.1 Performance Measurement	69
6.2 Data Management.....	69
6.3 Decisions	70
7 Suggestions for Further Research	71
7.1 Definitions	71
7.2 Data Collection	71
7.3 Culture and Organisation.....	72
References	73
Appendix A: Basic Railway Technology	85
Appendix B: Databases	93
Appendix C: Delay Attribution Codes	95
Appendix D: Glossary of Terms and Abbreviations	97
Appendix E: Appended Papers	103

List of Appended Papers

Paper I.

Nyström, B. & Kumar, U. (2003). Analysis of train delay information. In: *Proceedings of the World Congress on Railway Research (WCRR 2003)*, Edinburgh, Scotland, 28 September - 1 October, pp. 253-261.

Paper II.

Nyström, B. & Söderholm, P. (2005). Improved railway punctuality by effective maintenance – a case study. In: Kołowrocki, K. (ed.): *Advances in Safety and Reliability*, pp. 1487-1491, Taylor & Francis, London. ISBN 0 415 38340 4. Presented at ESREL, Gdansk, June 2005.

Paper III.

Nyström, B. (2006). Delay analysis of a freight train – an improvement case study from a steel company. Submitted for publication.

Paper IV.

Nyström, B. (2008). The Use of Availability Concepts in the Railway System. Submitted for publication.

Paper V.

Nyström, B. (2008). A Methodology for Measuring the Quality of Deviation Reporting – Applied to Railway Delay Attribution. Submitted for publication.

Paper VI.

Nyström, B. (2007). Consistency of Railroad Delay Attribution in Sweden – Measurement and Analysis. Submitted for publication.

Paper VII.

Niska, S. & Nyström, B. (2008). Investigating Recurring Faults – a Case Study of Oxmyran Signal Box. Submitted for publication.

Paper VIII.

Nyström, B. & Söderholm, P. (2008). Selection of maintenance actions using the Analytic Hierarchy Process (AHP): decision-making in railway infrastructure. Accepted for publication in *Structure and Infrastructure Engineering*.

1 Introduction

In this chapter, the research area is described and the research questions and the limitations are stated. The railway system is also outlined. The relation between the research questions and the appended papers is given and the following chapters are sketched.

Railways are a prime mode of transport for passengers and goods in many countries. The railway is not only known for being a comfortable means of transport, it also constitutes a safe and sustainable mode of transport of passengers and goods. In Sweden, the demand for railway transport, for both passengers and goods, is increasing (Banverket, 2006). The issue of the punctuality of trains, or rather the unpunctuality, becomes a topic of public attention when passengers or cargo suffer many and lengthy delays.

Transport costs constitute 18% of GDP for Sweden, compared to 12% in the EU and 10% in USA (Godstransportdelegationen, 2001). So, transports are relatively more important to Sweden compared to the EU and the USA. Therefore, reducing delays should be relatively more important in Sweden. As an illustration, delays make a regular commuter in the Stockholm area lose 1.7 work-weeks in a year (Transek, 2006). A rough estimation of the costs of passenger delays to commuter trains in the Greater Stockholm area is €150 million/year, including the cost of the extra time passengers add in order to account for possible delays. Other costs caused by train delay, such as train personnel overtime, more rolling stock, congestion on the railway tracks and the increased demand for road traffic, are not included in this figure. No indication of that the cost profiles for air traffic and rail traffic differ has been found. A study of European air traffic indicates that the delay costs for the operators are about as large as for the passengers (Institut du Transport Aérien, 2000). This gives a hint as to the total cost of passenger train delays, just in the Stockholm area.

In Sweden, the government has formulated aims on connectivity, quality of transport, safety, environment, regional considerations and gender equality for Banverket (the Swedish Rail Administration). For instance, measures of quality of transport include the number of train delays (the Swedish Ministry of Enterprise, Energy and Communications, 2005). The Swedish railway is still in a state of transition from one single governmental body responsible for the railway to many bodies; governmental, otherwise publicly funded or private. The increasing traffic and the new business setting call for new ways of improving punctuality.

1.1 The Swedish Railway System and its Stakeholders

A *railway system* consists of vehicles steered by a track on a dedicated area, which are governed by a signalling system. Internationally, the degree to which railways are

distinguished from underground and tramway systems varies (Andersson & Berg, 2001). The term *railway infrastructure* covers all the fixed installations on routes and stations¹ which are required for the running of trains (UIC 405, 1996).

The functions needed to run the trains are shown in Figure 1 alongside the names of the currently responsible bodies in Sweden. The newest governmental body is the Swedish Rail Agency, established in 2004, which manages safety and competition issues. The Swedish railway is representative of some other European railways, in that the government aims to make more private actors enter the railway sector.

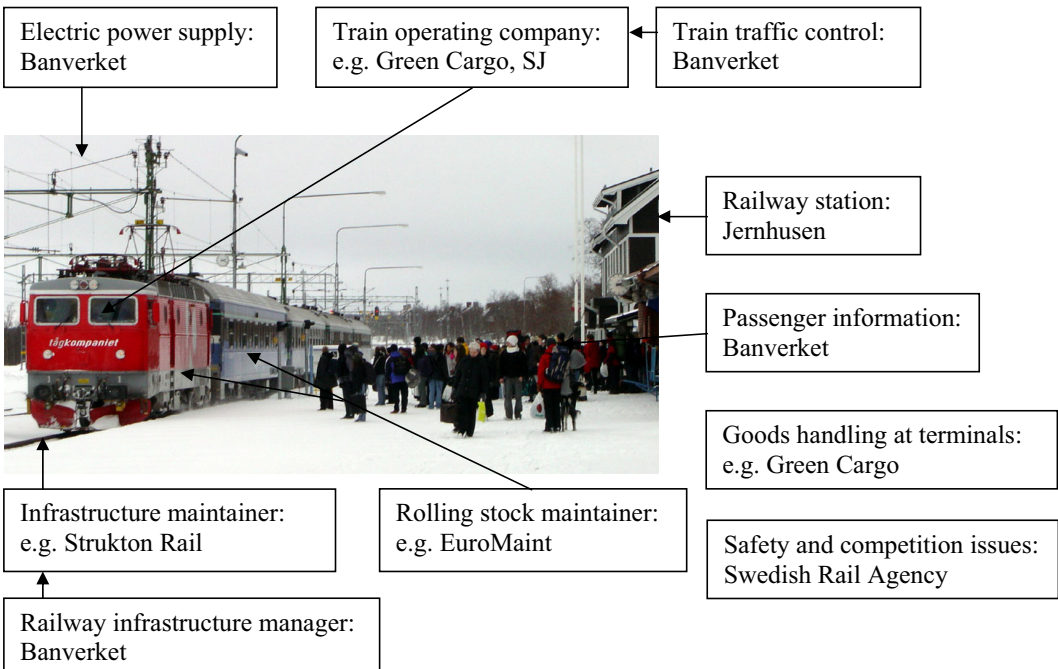


Figure 1. The functions needed to run a train and the names of the responsible bodies in Sweden. Photo: Jan Nilsson.

The trains are run by the train operating companies (TOCs), of which a few are described here, since they will be referred to later in this thesis. Green Cargo (freight) and SJ (passenger) are both offspring of the former state-owned monopoly. Another TOC is MTAB, a subsidiary of the mining company LKAB. MTAB transports ore to Luleå, by the Gulf of Bothnia, and to the port of Narvik by the Norwegian Sea. These transports form the dominant traffic along the iron ore line, although there is also other freight traffic and some passenger traffic (<http://www.banverket.se> (06-02-01); <http://www.lkab.com> (06-02-01)). Another TOC in Sweden is A-Train, which runs the Arlanda Airport train. The Swedish state is leasing the Arlanda line exclusively to A-Train until 2040. Thus, A-Train is, in contrast to most other Swedish train

¹ A station is in thesis understood as a place where trains might meet and overtake. It has two or more tracks. See Appendix D: Glossary of Terms and Abbreviations, p. 100.

companies, both a TOC and an infrastructure manager (IM) (<http://www.arlandaexpress.com> (06-02-01)).

Train traffic control (dispatching) is carried out by Banverket, and each TOC has a control centre of its own to monitor its traffic. Eight centralised train traffic control centres (CTTC) run by Banverket, perform train traffic control in their respective areas; that is, they execute the train plan (timetable), operationally coordinating trains and track works, including maintenance. In connection with train traffic control, delay attribution, that is, to record causes of delays, is performed. The delay attributors report delays into the database TFÖR, while infrastructure failures are entered into Ofelia. Both databases are managed by Banverket. Failures in the railway infrastructure strike, for example, turnouts, contact wire or signalling. More examples of problems are failures on locomotive motors, the freight being improperly secured and thus interfering with other trains, passengers embarking slowly and external factors such as severe weather conditions and sabotage.

Banverket, the principal infrastructure manager, plans the train traffic and the railway infrastructure maintenance. Maintenance execution is tendered out to public and private bodies. Electric power supply and passenger information are also handled by Banverket.

One of the infrastructure maintainers is the maintenance contractor (MC) Strukton Rail, formerly named Svensk Banproduktion, an offspring of Stockholm Local Traffic (SL). Within SL, it performed maintenance of, for example, the Stockholm underground; now it places tenders all over Sweden. The Dutch company Strukton Railinfra has been the sole owner of the company since 2007 (<http://www.strukton.se> (07-11-29)). Regarding rolling stock, that is, locomotives and wagons, maintenance and modifications of varying complexity is carried out in workshops near major railway routes, for example, by EuroMaint.

The bodies illustrated in Figure 1, together with rolling stock leasing companies and public traffic subsidisers, are considered the main stakeholders of the railway. A *stakeholder* is in this thesis defined as “*an interested party having a right, share or claim in the system or in its possession of characteristics that meets that party’s needs and/or expectations*” (ISO/IEC 15288:2002).

1.2 Unpunctuality Problems and Improvement

The unpunctuality in the railway affects its competitiveness relative to other transport modes, that is, road, sea and air transport, due to the inconveniences unpunctuality causes the customers, but also due to the costs imposed on other stakeholders. On densely trafficked railway lines, a small delay to one train might cause further delays to other trains, especially in Sweden, where slow freight and rapid passenger trains often share the same single-track.

The timetable plays an important role in managing the risk of unpunctuality (Riksdagens revisorer, 2002). This is because one might construct a sparser timetable by allowing more travel time for the trains and allow more time on the track for maintenance and investments. However, this allows less traffic to fit in the railway network. It might also lower the ambitions. For example, one scenario is that TOCs might not pay enough attention to maintaining their rolling stock on the trains for which they do not consider punctuality to be important, and these trains are therefore more likely to hinder other trains. The same reasoning goes for infrastructure maintenance, where recurrent faults that take only a short time to repair are considered 'normal'. Another example is that track work time might be reserved 'just in case', but not used, thus hindering trains from being timetabled during scheduled track work time. Conversely, a denser timetable makes it more likely that the delay of one train hinders other trains; that is, the delay multiplies. These phenomena make planning of train traffic and infrastructure maintenance, as well as train traffic control, important areas of research (Bente et al., 2003; Genovesi & Ronzino, 2003; Mücke, 2002; Simson et al., 2000).

Often, figures are presented to demonstrate that punctuality is bad: "the delays cost the passengers €150 million each year" or good: "90% of the trains are punctual". The choice of what is measured partly accounts for the perceived difference. The author therefore points out the need for a clear specification of unpunctuality when one wants to compare unpunctuality of different railways, as well as, for instance when defining the qualifications for unpunctuality penalties.

There are many causes of unpunctuality. The causes might be related to a technical system in the railway or related to organisation. The Swedish railway sector consists of many business areas managed by different independent organisations and companies, as already illustrated. Therefore, the assumption of a single rational stakeholder does not apply to the railway (Zoeteman, 2004). That each stakeholder sees to its best might generate suboptimisations. This creates many problems related to effective management of the railway sector. It also creates difficulties in assuring high punctuality.

The technical causes of unpunctual trains are diverse. Examples are non-functioning of wagon doors, faulty turnouts and contact wire failures (Granström, 2005). Another cause of unpunctuality is overweight trains (as a TOC wants to load as much as it can) leading to trains getting stuck on slopes. The physical interfaces between rolling stock and infrastructure, and hence between TOC and IM, are wheel-rail (makes the propulsion possible), pantograph-contact wire (collects the power needed for propulsion) and signalling (ensures safety), including ATC (Automatic Train Control). Also, the interface to road traffic, level crossings, might be places of interference with train traffic. When problems occur in the interfaces, it might be hard for the involved stakeholders to agree on a suitable action.

Different stakeholders see different parts of the problems related to unpunctuality. The locomotive drivers, employed by the TOC, know the problems with their rolling stock by heart and see only the manifestations of infrastructure failures, for example, a red signal. For the IM or MC personnel, it is the other way around. They know that, by executing maintenance and other work that will make the track unavailable for a while, problems will be avoided later on. From this, it seems important for the stakeholders to agree on a common picture of the problems.

In their effort to reduce unpunctuality, railways might learn from other railways, but as the circumstances vary from country to country and region to region, one should not just copy the solutions without thorough consideration. For example, a wagon with a fault, discovered underway, might be left at a station. If there are only two or a few tracks at a station, the hindering wagon may make it difficult to execute other planned traffic. In Sweden, this might be solved by building more tracks, but not in Japan, where railway stations seldom have space to build additional tracks.

Not only between stakeholders, but also within one stakeholder, sub-optimisations occur. For example, TOC personnel used to take out empty freight wagons from passing trains and put in loaded wagons instead. After all, customers were waiting for loaded wagons, not empty. This also had the advantage of making empty wagons of the right type available in case they were needed. As personnel followed this procedure in many places, there was a lack of wagons, which was solved by buying more wagons. This increased the cost of wagons, but also caused costs for the extra tracks required to store them and the extra time needed to locate and shunt a certain wagon, as well as for administration (Joborn, 2001). As this strategy caused more trains to be run, the availability of track for other trains diminished. A TOC, with trains that do not hinder other trains, for example, by being delayed, might be seen as not unnecessarily lowering track availability for others. Well-maintained rolling stock that does not halt underway or damage the infrastructure is a prerequisite for train punctuality. Not only the rolling stock, but also the infrastructure needs to be available, in order to run trains. This is catered for by the maintenance infrastructure organisations, that is, the IM and the MCs.

Databases on punctuality and maintenance are often used to give performance measures and they comprise a valuable basis of information for improvement work. The work that was formerly performed by local railway personnel (watching, touching, listening and smelling), is now divided among many persons, often not in direct contact with the assets. Since these people might be replaced when contracts end, it is even more important that the data used is suitable and accurate enough to provide a basis for decision-making. This is also stressed by the fact that the life-length of a railway is longer than a human working life. Banverket, for instance, usually makes economic analyses in a time-frame of at least 60 years (BVH 706, 2005).

The choice between different investment alternatives in the railway infrastructure, for example, where to build a new railway line, is often publicly discussed and has to follow regulated steps. For the performance of the railway system, the choice of which maintenance actions to carry out is also important. Today, Banverket allocates a fixed budget for maintenance (Espling, 2007). Track managers have to fit their expenditures in their budgets, trying to maximise the utility of the maintenance actions carried out, although it is unclear how these should be assessed. The selected maintenance strategy has, of course, great influence on track and vehicle performance.

Furthermore, if the decided maintenance actions are not performed properly, it may lead to failure of infrastructure or rolling stock, leading to delays. If maintenance is not executed in a correct manner, or if necessary preventive maintenance is not performed, it may often lead to poor reliability and availability of the railway infrastructure and rolling stock. It may also lead to incidents and accidents with extensive losses (Holmgren, 2006). The incidents and accidents might also cause unpunctuality.

Therefore, to reduce unpunctuality, it is of interest to study the process by which maintenance actions are prioritised.

1.3 Purpose, Scope and Delimitations of Research

The purpose of this research is to explore and describe information and requirements related to railway punctuality in order to support systematic improvements. To fulfil the purpose a main research question has been formulated: *How can punctuality of railways be improved?* From the main research question, the following sub-questions were formulated:

- Research Question A: *How should (un)punctuality be described?*
- Research Question B: *Which problems are encountered when trying to reduce unpunctuality?*
- Research Question C: *How might availability of the railway infrastructure be described?*
- Research Question D: *What is the suitability of the information, from systems measuring punctuality and maintenance, for improvement work?*
- Research Question E: *How is the choice made between different maintenance actions?*

The research is focused on punctuality problems in the Swedish railway due to accessibility of information, and is mostly related to infrastructure maintenance. However, the current deregulation of railways in Sweden is representative of a similar ambition in some other European railways. Therefore, the research might be applicable to more railways than the Swedish.

1.4 Thesis Structure

The relation between the research questions and the appended papers is given in Table 1. After this chapter, the thesis continues with the theoretical frame of reference in Chapter 2. Thereafter follow chapters on research design, descriptions of appended papers and a discussion and conclusions chapter. The thesis concludes with Chapter 6, which gives suggestions for improvements, and Chapter 7, which gives suggestions for further research.

Table 1. Illustration of which of the appended papers (I-VIII) are most relevant in answering the research questions (A-E).

	I	II	III	IV	V	VI	VII	VIII	Main methodology
A	x	x				x			Literature study
B			x				x		Case study
C				x					Literature study
D	x				x	x			Archival, Survey
E								x	Experiment

The railway technology, relevant for understanding this thesis, is described in Appendix A, some important databases in the Swedish railway are described in Appendix B and different sets of delay attribution codes are described in Appendix C. Appendix D contains a glossary. Finally, Appendix E consists of the eight appended papers.

2 Theoretical Frame of Reference

In this chapter, the concept of time and the definitions of punctuality and associated terms are discussed. The lifecycle of the railway system is considered, including the operation and maintenance phase. Factors affecting the unpunctuality cost are presented. Finally, possible methodologies and tools for improving the performance of the railway system are discussed.

2.1 Time, Punctuality and Associated Terms

To give the reader a background to the subsequent discussions about punctuality, some different thoughts about the concept of time are first presented.

2.1.1 Time

The fundamental nature of time has been perceived differently in different civilisations. Today, a linear perception of time is prevalent in the western world². It might be contrasted to a circular perception of time, which has its origin in the recurring seasonal changes. In ancient Egypt, it was the Nile that fertilised the fields, making people perceive time as a recurrence of what had happened (Whitrow, 1988). In some cultures, for example in South America, time is perceived as more defined by the occurring events than by the clock, while it is the other way around in the west. One might say that in event-time cultures, the time is indistinguishable from the events that take place, for example, sleeping. Conversely, in clock-time cultures, time is external to the events that take place inside it. In clock-time cultures, people often see chunks of five minutes as a means of organising their workday (Levine & Norenzayan, 1999). Basu & Weibull (2002) observe that it is profitable for an individual to be punctual only if others are also punctual and conclude that this might be the explanation for the large differences in attitude towards punctuality in different cultures (Levine et al., 1980; Shaw, 2001).

The first time scales were based on celestial motions. Today, time is standardised as an average of atomic clocks (Arias, 2005). However, time as perceived by humans is not necessarily identical with physical time. For example, sound that is slightly lagging the picture in a dubbed movie appears as synchronised with the picture. So, for humans, ‘moments’ in time have a certain extent (Durgin & Sternberg, 2002). Our estimation of the length of a period of time is affected by, for example, mood, body temperature, arousal and whether we estimate the length of a passed or future period of time (Glicksohn, 2001; Kuriyama et al., 2003). So, many factors affect how we

² This is manifested in the way of representing dates and times as relative to a certain point in time (the birth of Jesus Christ), codified in ISO 8601. An example of another way is to indicate the era and the year of the era. For example, the year 2008 according to the Gregorian calendar, is in Japan also called Heisei 20. (<http://www.allcalendars.net/JapaneseYearConverter.php> (07-11-14)).

perceive the duration of a time period. The author stresses that one should be aware of this when asking, for example, how long a journey is perceived to be.

Time perception research dates back to the late 18th century, according to Roeckelein (2000). His bibliography indicates that the research on subjective duration has dealt mostly with short time (seconds) or long time (days), but not so much with time periods of intermediate duration, among which travel times are usually included. Regarding the perceived durations of time intervals of the same objective length, applicable to, for example, a journey involving change of trains, Roeckelein (2000) presents Fraisee's principles:

- A time interval divided in several parts is experienced as longer than a non-divided interval of the same duration.
- More divisions make the interval to be experienced seem longer.
- An evenly divided interval is experienced as longer than one which is irregularly divided.

If travel time is considered a loss, these principles are consistent with Prospect Theory (Kahneman & Tversky, 1979), which states that many small losses are considered worse than one large loss of equal amount. Horowitz (1978) studied the subjective value of time spent travelling as a function of travel mode and purpose. The subjective value of travel time by car to work was found linear to the time, so other subjective travel times might be expressed as car-hours to work.

The extension of the railway is an example of how civilisation has affected time perception. Before the introduction of the railway, each city had its own time. For instance, there was a difference of 24 minutes between Stockholm and Gothenburg. With the advent of railway timetables, this was not feasible. Nevertheless, the introduction of a single Swedish time in 1879 was not realised without protests (Eriksen, 2000).

2.1.2 Punctuality and Related Terms

A term related to punctuality is timeliness³. *Timeliness* is in this thesis defined as the extent to which a journey or transport occurs when desired. Note that timeliness is highly subjective⁴, and varies between stakeholders. For example, it might be

³ No explicit definition of timeliness in the sense used in this thesis has been found in literature. There are other authors that use the word synonymously to punctuality (Biederbick & Suhl, 2007).

⁴ One might measure timeliness, for example as Rosenlind et al. (2001), who ask current train passengers to relate prospective changes to the train departure time to changes in fare. Subsequently, a model is sketched, where the stated preferences of the passengers, regarding for example departure time, on-board service and fare, taken together with the current timetable, are used to predict how passengers will distribute themselves on different trains, dependent on various changes in the timetable. However, because they ask passengers onboard certain trains about the current journey, it is likely, as the authors point out, that these passengers have adapted to the current timetable, and therefore underestimate the value of a change to, for example, departure time.

different to a freight customer and the corresponding TOC, as the freight customer considers its production process, while the TOC considers the demands of the customers of the TOC. Timeliness includes the facet of length of journey time. An example of desire for timeliness is expressed by SJ, which aims at always achieving a journey time of three hours or less for its passengers on the X2000 trains⁵ between Stockholm and Gothenburg.

Different railways have different definitions of *punctual*⁶. For instance, railway authorities sometimes define punctual for departure and arrival separately, defining a tolerance, and *punctuality*⁷ as the share of trains within that tolerance. The tolerance varies between countries and with, for instance, the length of travel. For example, the tolerance is 1 minute (for the Japanese TOC JRK), 2 minutes (Denmark DSB commuting), 3 minutes (The Netherlands NS, Austria ÖBB, Norwegian NSB local, Australia local), 5 minutes (Sweden Banverket, Norway NSB long-distance, Finland VR, Switzerland SBB, GB regional), 10 or 15 minutes (GB), 30 minutes (Australia freight), 10, 15, 20, 25 or 30 minutes (Amtrak) (<http://www.bts.gov> (04-09-27)). A-Train, which operates between Stockholm Central Railway Station and Arlanda Airport, instead considers a train to be punctual if the journey takes at most 22 minutes. Stockholm Local Traffic (SL) means by ‘punctuality’ the share of departures that occur between one minute before timetable and three minutes after timetable (<http://www.sl.se> (04-04-01)).

In this thesis, the definition of *punctual* is based on an everyday meaning of the word: the property of an *event* occurring “on time”. The word *punctual* stems from Latin’s *punctum*, derived from *pungere*, which means “to hit the point”. An *event* is commonly classified as being of one of two sorts: *accomplishments* and *achievements*. Accomplishments are reported by sentences like “She made a sandwich”. This sort of event may or may not take an extended amount of time and may or may not have a culmination. Achievements are reported in sentences like “She won the race”. Achievements are instantaneous (Parsons, 1990). So, punctual refers either to a point in time or an indivisible quantity of time. This discussion motivates the following definitions of punctual and punctuality.

Punctual means in this thesis that one or several events occur when agreed between involved stakeholders. The antonym to punctual is *unpunctual*. Punctual is an

⁵ The traffic with trains of type X2 is sold by SJ under the trade name ‘X2000’.

⁶ Rudnicki (1997), who investigated urban public transport in Poland, takes punctual to mean whether departure (of a bus or tram) is within a certain tolerance (maximum allowed difference from the timetable).

⁷ Rudnicki (1997) does not explicitly define punctuality. However, he defines *factor of unpunctuality inconvenience* as passengers’ extra waiting time due to the difference between departure time and timetabled departure time. As the tolerance might allow departure prior to timetable while the departure is still considered punctual, passengers’ extra waiting time might be longer than the timetabled interval between buses/trams, and hence the *factor of unpunctuality inconvenience* higher, despite the bus/tram departs according to timetable.

attribute variable in the sense that an event is either punctual or not. Note that whether an event is punctual or not is undefined when an agreement does not exist or is unclear. *Punctuality* is the extent to which one or several events occur when agreed. Punctuality is often measured using the concepts of proportion or percentage. For an operational definition of railway punctuality, using generative grammar (Chomsky, 1965), see Appendix D.

Delayed is generally taken to mean happening after the agreed time. For trains, a *delay* in this thesis means that the train lags compared to the timetable. Note that parts of the timetable might be defined by agreements, and thus punctuality of trains is defined for these parts, but is undefined for others. The difference from the timetable is usually measured in minutes. The difference might indicate a delay or a lead, i.e. happening before the time stated in the timetable. The definitions include the further specification of, for instance, which stakeholders have made the agreement.

Railways usually treat cancelled trains separately, that is to say, they are not included in the term punctuality. This is not in line with the definition of punctuality used in this thesis, which requires events, for instance, departure of a train, to “occur when agreed”. The public performance measure (PPM), which merges unpunctuality and cancellations, is used in the UK (<http://www.sra.gov.uk> (04-01-31)). Most of the railways in the BEST (2002) study, which compared train punctuality in fourteen countries, have goals for punctuality, but fewer measure cancelled trains (<http://www.bts.gov> (04-09-27)). In the UK, TOCs that alter the timetable from the printed one are penalised (<http://www.sra.gov.uk> (04-01-31)). So, although these deviations are not included in the term punctuality, they might be seen as divergences from the agreement between a TOC and (potential) passengers, and are measured and subject to incentives in the UK.

It is also interesting to compare definitions related to punctuality in the railway with those in other transport sectors. A flight is considered as departed when the aeroplane starts moving from its parking lot⁸ (Luftfartsverket, 2000:2). So, when the aeroplane actually ascends does not affect the departure time. A flight is considered as arrived when the aeroplane stands still at the arrival gate⁹. Cases when the passengers are not able to leave the plane, such as when the doors do not open, do not affect the arrival time, according to this definition. If the departure and arrival times given on the flight ticket are considered to constitute an agreement, there is a difference between the definition of punctuality in this thesis and the definition of Luftfartsverket (2000:2). The author judges Luftfartsverket’s definition to be reasonable from the perspective of the airport gate, but not from the perspective of the airport or the passenger. A difficulty encountered in air traffic is whether or not the term punctuality also should include access to checked-in luggage. From the perspective of a person travelling

⁸ This is called *off-block time* (Luftfartsverket, 2000:2).

⁹ This is called *on-block time* (Luftfartsverket, 2000:2).

with luggage it seems reasonable that it should, but it is not obvious how this should be done.

The Association of European Airlines, AEA, shows the punctuality (defined as the share of the flights that are punctual) of its members at <http://www.aea.be>. Its statistics differ from those of Eurocontrol¹⁰, as AEA employs a tolerance of 3 minutes to define a flight as punctual, but Eurocontrol has a tolerance of 5 minutes. In the US, a flight that departs or arrives more than 15 minutes after timetable counts as delayed (<http://www.bts.gov> (04-01-31)). So, as in the railway, tolerances differ, but not for each country. In essence, air traffic is more internationalised than railway traffic.

Generally, a journey might have good punctuality (that is, it adheres to agreed departure and arrival times), even if it has bad timeliness, due to, for example, long exchange waiting times because of poorly timetabled connections. It is understood that timeliness is a broader concept than punctuality. The multitude of definitions in use of punctuality and interrelated terms makes it important to be careful, when, for example, comparing railways in different countries and comparing railways to air traffic. Important features of the definitions given in this thesis are that cancelled trains are included in the punctuality definition and that a delayed train might, or might not, be considered unpunctual, while an unpunctual train is always considered delayed.

2.1.3 A Subdivision of the Travel Time of a Train

The *actual running time (journey time)* of a train is in this thesis divided according to Figure 2.

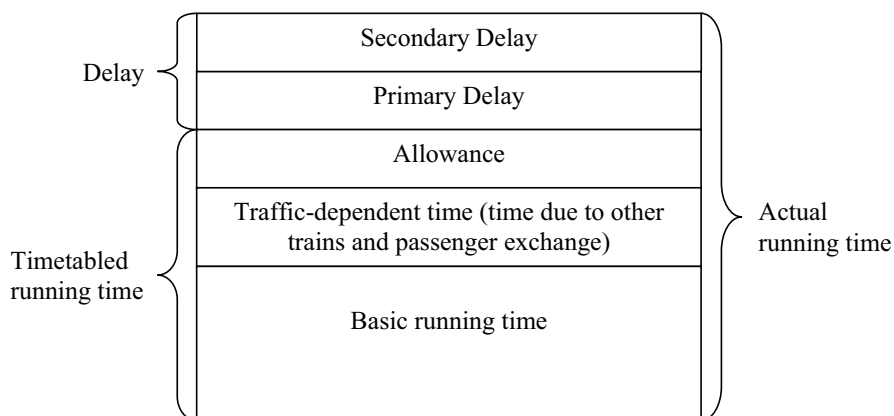


Figure 2. A subdivision of the actual running time. The typical relative size of the basic running time is understated in the figure.

¹⁰ Eurocontrol, the intergovernmental European Organisation for the Safety of Air Navigation, aims to create a uniform air traffic management system in Europe (<http://www.eurocontrol.be> (08-01-25)).

The actual running time consists of the *timetabled running time* and the *delay*. The timetabled running time might be further subdivided into basic running time, traffic-dependent time and allowance. The *basic running time* is defined as the shortest time needed to cover a given section of line on the basis of the technical performance of the rolling stock and according to mean values for the factors driving style, rolling stock performance, adhesion and power supply (UIC 451-1, 2000). The *traffic-dependent time* is in this thesis defined as the time that is added in the timetable to deal with train meetings and overtakes, as well as the time added for passenger exchange¹¹. *Allowance* is defined as the amount of time added to the *basic running time* or to the *traffic-dependent time*¹². Its purpose is to serve as a recovery time and thereby prevent or reduce the delay. Its length might vary between trains. It is understood that the differences from the basic running time, caused by, for example, the factors mentioned, will require there to be allowances in the timetable, in order to avoid delays. In Sweden, the lengths of the allowances are established by instructions and rules of thumb (TF601, 2000). The delay might be further subdivided into primary and secondary delay, as defined in Section 2.2.

Mattsson (2004) defines *time lost*¹³ as the difference between actual running time and basic running time, that is, $time\ lost = actual\ running\ time - basic\ running\ time$. According to this definition, the time lost thus consists of the delays, but also of the allowances and the traffic-dependent time, as illustrated in Figure 2.

By prioritising between trains, timetablers and train traffic controllers try to minimise total consequences when a certain train is delayed. The timetablers anticipate probable deviations beforehand, and adjust the timetable accordingly. The train traffic controllers operationally deal with the delayed trains, for example, by changing the order of trains.

2.2 How a Delay to One Train Inflicts Delays on Other Trains

A railway line might be single-track, with the line merely consisting of one track. To permit trains to meet (or overtake), stations are built, as illustrated in Figure 3. In Figure 3, two trains are moving east, towards a westbound train, which stands waiting at a station. A delay to the leading of the eastbound trains or to the westbound train might inflict a further delay on the second eastbound train. This situation illustrates that a delay of one train can spread to other trains. Therefore, it is beneficial to define two kinds of delays.

¹¹ These two time intervals might intersect, as a train might meet, or be overtaken by, another train when it picks up or drops off passengers.

¹² The traffic-dependent time includes time due to stand-stills as well as due to accelerations and decelerations.

¹³ Mattsson (2004) employs a terminology partly different from the one used in this thesis, as he uses *delay* and *time lost* synonymously. Mattsson (2004) and Rudolph (2003) seem to use *minimum running time* as a term identical or similar to *basic running time*.

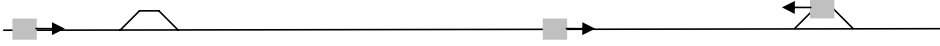


Figure 3. A railway line viewed from above, showing two stations. Two trains are moving east, towards a westbound train standing at a station's deviating track. In the figure, the widths of trains and stations are exaggerated.

A *primary delay* to a train is a delay that directly strikes that particular train. A *secondary delay* is a delay caused by another train's deviation from the timetable (Dahl, 1997). Often, other terms are used synonymously to secondary delay, front mostly *knock-on delay* (for example, by Carey & Kwieciński, 1995) and *cascade delay* (for example, by Nelson & O'Neil, 2000).

So, a primary delay to the train in the middle of Figure 3 might impose secondary delays on both the other trains. Primary delays occurring simultaneously to the eastbound trains (for example, signalling and engine problems, respectively) might cause a secondary delay to the westbound train. However, it is not clear whether to blame the secondary delay on the faulty signalling or the faulty engine. From this, it is understood that it is not true in principle that removing a cause of primary delay nullifies the secondary delay.

A train might inflict damage to the infrastructure. As an example, a wheel flat on a wagon of a train, caused by locked brakes, might break the rail, thereby causing delays to other trains¹⁴. TOC personnel (drivers, conductors) are frequently travelling by train to the place from which they depart with their train. An arrival delay of the arriving train might then impose a delay on the departing train, via the personnel. The chance of this happening might be reduced at the expense of, for example, having standby-personnel.

2.3 Unpunctuality Cost

The *unpunctuality cost* is the cost of the consequences of unpunctuality. In order to value punctuality, the Stated Preference method is often used¹⁵. Using this method, the subject is asked to study hypothetical alternatives, where several parameters vary, and select the preferred alternative. Ackermann (1998) uses costs obtained by the Stated Preference method together with the average number of passengers for different train types in Germany (such as ICE or IR), to compute the cost of one train-delay minute, including the unpunctuality cost that strikes the passengers and the unpunctuality cost that strikes TOCs' operation. Also König & Axhausen (2002) employ the Stated Preference method. In their study of the Swiss railway, they find that the cost to passengers is non-linear; the cost of a delay increases more rapidly up

¹⁴ Even the suspicion of rail break causes a time-consuming inspection of the line, possibly causing train delays.

¹⁵ It has also been used to value, for example, transport frequency, flexibility and the risk of loss and damage, as in Witlox & Vandaele (2005).

to about ten minutes than for longer delays. They estimate the cost for an average delay as about 1.5 times a timetabled journey time prolongation of the same duration, a factor which is slightly lower than SIKA (2002). Others give other factors, for example six, which is used by SL (Stockholm Local Traffic), according to Mohns (2007), and seven, from the costs given by Nelldal et al. (2000). König (2004) models the unpunctuality cost, found by Stated Preference studies, considering the length and probability of delay for public transport. The formula reflects the rapid increase in unpunctuality cost with increased probability of delay¹⁶ and gives the unpunctuality cost (in Swiss francs) as

$$0.074 \cdot \text{Length of delay} \cdot \left(\frac{\text{Probability of delay}}{0.40} \right)^{2.233}$$

Instead of asking for hypothetical choices between trains, one might study revealed preferences, that is, how passengers actually choose between trains (as done by, for example, Slagmolen, 1980). A difficulty associated with studying revealed preferences is that the parameters that one wants to investigate relative to each other, such as ticket price, travel time and delay, often covary. A difficulty associated with the Stated Preference method is that the way the risk of delay is presented to travellers, as well as other differences in presentation, might affect their answers, as discussed by Bates et al. (2001) and Widlert (1998).

The cost of a certain length of journey time, which might be calculated by using, for example, the Stated Preference method, varies with the situation. The time might be timetable or delay, the passenger might be in a vehicle or waiting for change, the purpose of the journey might be business or private¹⁷. Also the mode of transport affects the cost of journey time (SIKA, 2002). Regarding train freight, Banverket values delay time according to what is transported. For example ‘Round timber’, is SEK 0.12 / (tonne·h) and ‘Consumer products’ is SEK 11.24 / (tonne·h) (BVH 706, 2005). So, the cost of delay time varies by about a factor of 100, which means that the transported product must be known (to some extent) in order to value the delay time. For freight and passengers, Banverket values the delay time as twice the timetabled time.

When timetabled travel time and delay time of equal duration are valued differently, the punctuality of a journey or transport might be improved (or worsened) by adjusting the timetable and the total cost is thereby considered lower (Östlund et al., 2001). When timetableers know the probability distribution of train travel time, the

¹⁶ For private transport, the unpunctuality cost according to König (2004) is calculated as $0.215 \cdot \text{Length of delay} \cdot \left(\frac{\text{Probability of delay}}{0.40} \right)^{1.299}$. So, the cost for public transport increases faster than for private transport.

¹⁷ Whether a journey is to be considered as business (high value of delay time) or private (lower value of delay time), is not always clear-cut, for instance when commuters do some work during the journey.

cost of timetabled time and the cost of delay time, they are, in principle, able to minimise the expected total cost, following classical utility theory of von Neumann & Morgenstern (1953). However, people are not perfectly informed and rational, contributing to the difficulties of such an optimisation. For example, according to Rietveld et al. (2001), having a 50% probability of being two minutes delayed is considered worse than 100% probability of being one minute delayed; that is, the passengers are risk-averse. The unpunctuality and its cost affect the demand for train traffic. Unpunctuality perceived as more expensive, for example, being delayed in getting to work the first day after vacation, might have a greater effect on demand for rail commuting. Then, people might be more prone to catch an earlier train or use another travel mode than they would have been otherwise.

Data on each passenger's (or other stakeholder's) valuation, together with the occurred unpunctuality, is needed to calculate the unpunctuality cost. For practical use, approximations might have to be made, both regarding the price tags and how the passengers try to recover from deviations and hence the unpunctuality that strikes passengers. In an analysis, for example, of a commuting network, the unpunctuality that strikes the passengers might be approximated by the following models:

- the passengers act as if they are perfectly informed about the delays that will happen
- the passengers act based on information such as traffic information or own experience of probability of delay
- the passengers act as if there were no delays

These models might be used separately or in combination to estimate the unpunctuality of the passengers (Landex & Nielsen, 2006). The author suggests that this reasoning be applied to freight as well, as the route of a wagon might be chosen from different routes in the network (possibly in different trains).

A different viewpoint of punctuality might be derived from observations indicating that travel time is constant, averaging about one hour per day, across a population. The hypothesis of constant travel time is discussed by, for example, Höjer & Mattsson (2000) and Metz (2004). If travel time is constant, and the effects of early arrivals are neglected, to increase punctuality would then not result in 'saved time', but in increased travelled distance. The view of travel time as constant is also in contrast to roadway practice, which calculates the reduction of average travel time from home to, for example, work or emergency care, when evaluating prospective road investments (Vägverket, 2002).

Yet another view of unpunctuality and the unpunctuality cost might be from the view of failures causing unpunctuality. Failures might happen to both railway infrastructure and rolling stock. Unpunctuality cost due to different kinds of infrastructure failures in the London underground was calculated by Harris &

Ramsey (1994). This calculation, including rolling stock failures, was also performed for the Italian railway, by Di Marco et al. (2000).

To summarise, calculation of unpunctuality cost is sensitive to the different values which are used for timetabled delay and waiting time. By knowing the cost of delay time relative to timetabled time, timetabling might facilitate minimisation of unpunctuality cost. However, actions that might require expenditures, such as increasing availability of railway infrastructure or rolling stock, need to be put in relation to an amount of money.

2.4 Maintenance

One way to improve punctuality is to perform maintenance on different items in the railway system. However, on the other hand, maintenance can also cause delays and in that way increase unpunctuality. This might happen, for instance, if maintenance is performed in an incorrect way or if maintenance of the infrastructure disturbs the traffic. In this thesis, mainly maintenance of railway infrastructure is considered. However, due to the interplay between rolling stock and infrastructure, some examples of rolling stock maintenance are also discussed.

2.4.1 Some Concepts

Here, some concepts of importance when discussing maintenance-related issues are presented. Maintenance and related terms are defined in the SS-EN 13306:2001 standard, which is the basis for the discussion below. Other standards, such as IEC 60300-3-14 and BS 3811, have comparable definitions.

Maintenance is, according to SS-EN 13306, “*the combination of all technical, administrative and managerial actions during the life cycle¹⁸ of an item¹⁹ intended to retain it in, or restore it to, a state in which it can perform the required function*”. Observe that, according to this definition, maintenance is to assure the function of an item. Improving or changing the function of an item is not considered maintenance.

Failure is the termination of the ability of an item to perform a required function. After failure the item is said to have a *fault*, which may be complete or partial. Hence, failure is an event, as distinguished from fault, which is a state (SS-EN 13306). An example of *partial fault* of a railway line is bad track alignment, which necessitates a reduction in speed. A *complete fault* of one of several motors in a locomotive, will cause a partial fault to the locomotive, if the train is still able to travel at reduced speed.

¹⁸ The life cycle is the time interval that commences with the initiation of the concept and terminates with the disposal of the item (SS-EN 13306).

¹⁹ An item is any part, component, device, subsystem, functional unit, equipment or system that can be individually considered. A number of items, example given, a population of items, or a sample, may itself be considered as an item (SS-EN 13306).

Maintenance performed after a detected fault is called *corrective maintenance*, while maintenance performed before a detected fault, is called *preventive maintenance*. Preventive maintenance, in turn, might be based on the condition of the item, so-called *condition-based maintenance*, or be decided in advance, for instance to be performed annually, so-called *predetermined maintenance*. In a railway context, this means, for example, that to measure the thickness of a pantograph's carbon slippers and replace them when they have reached a certain thinness, is condition-based maintenance (Granström, 2005), but to replace the light bulbs in signals once a year is predetermined maintenance. If a light bulb fails, trains are not allowed to pass the signal at full speed. Trains, so delayed, will possibly inflict secondary delays to other trains. Therefore, fast corrective maintenance is needed. The current failure management procedures, from fault detection to registering the fault as repaired, are described in Paper I and Wahlström (2006).

As mentioned, maintenance is performed to assure the function of an item. One measure, of how well an item functions, is its availability. *Availability* is “*the ability of an item to be in a state to perform a required function under given conditions at a given instant of time or during a given time interval, assuming that the required external resources are provided*” (SS-EN 13306). Accordingly, maintenance of a railway infrastructure item, such as a turnout, tries to ensure its availability and thereby makes it possible for trains to be forwarded through it at the desired time, resulting in better possibilities for punctual trains. Roughly, availability is estimated as the proportion of time the item is in a ‘functioning state’. For a deeper discussion on how to estimate availability, see, for instance, SS 4410505 or the railway standard SS-EN 50126.

The availability depends on the combined aspects of reliability, maintainability and maintenance supportability. *Reliability* is here defined as “*the ability of an item to perform a required function under given conditions for a given time interval*” (SS-EN 13306). However, the term *reliability* is also used as a measure of reliability performance and may also be defined as a probability (SS-EN 13306). *Maintainability* is “*the ability of an item under given conditions of use, to be retained in, or restored to, a state in which it can perform a required function, when maintenance is performed under given conditions and using stated procedures and resources*” (SS-EN 13306). Finally, *maintenance supportability* is “*the ability of a maintenance organization of having the right maintenance support at the necessary place to perform the required maintenance activity at a given instant of time or during a given time interval*” (SS-EN 13306). Concerning rolling stock, high maintenance supportability means, for example, that the workshops manage the turning of wheels in a timely manner. So, reliability and maintainability are characteristics of an item, for example, a railway turnout, a signal lamp or an entire railway line, while maintenance supportability is a characteristic of a maintenance organisation.

2.4.2 The Maintenance Process

The activities needed to perform maintenance might be ordered in a process, with respect to their role in assuming the required function of the item. A *process* may here be defined as “a set of interrelated work activities characterised by a set of specific inputs and value-added tasks that make up a procedure for specific outputs” (ASQ, 2005). Another definition is that a process is “a set of interrelated or interacting activities which transforms inputs to outputs” (ISO 9000:2000). The combination of generic maintenance activities or actions that transforms input to output may be seen as a maintenance process (Holmgren, 2003). The purpose of the maintenance process is to sustain the capacity of the system to provide service (ISO/IEC 15288).

The maintenance process can be described with different degrees of resolution. In Figure 4, the maintenance process is illustrated with a rather low resolution (IEC 60300-3-14:2004). See also discussions in Holmgren (2006) and Söderholm et al. (2007).

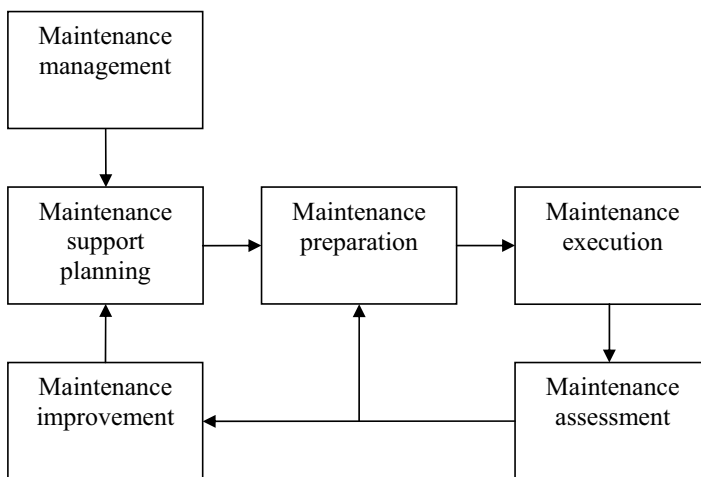


Figure 4. The maintenance process (IEC 60300-3-14:2004).

In Figure 4, *Maintenance management* includes, for instance, policy-making and strategies, budgeting, coordination and supervision of maintenance activities on an overall level. *Maintenance support planning* includes maintenance support definition, task identification and analysis and maintenance support resources. The next step, *maintenance preparation*, includes planning, scheduling and assigning resources for the decided maintenance activities. In a railway context, this includes designating possession of the track for maintenance, with consideration of the risk of disturbing the traffic and causing unpunctuality. *Maintenance execution* deals with the operational and practical performance of maintenance, and the recording of the results. *Maintenance assessment* includes measuring maintenance performance,

analysing results and assessing the actions that are to be done. Finally, *maintenance improvement* is the phase aimed at taking care of the obtained experience to improve maintenance concepts, improve allocation of resources, improve and implement new and more efficient procedures intended to ameliorate the reliability of an item, without changing its required function. It might be a combination of technical, administrative and managerial actions.

2.5 Factors Influencing the Unpunctuality Cost

Figure 5 shows relationships between factors influencing railway punctuality and thereby unpunctuality cost (the author is inspired by Al-Haimi, 1991, who investigated air passenger travel). Some key points are now discussed.

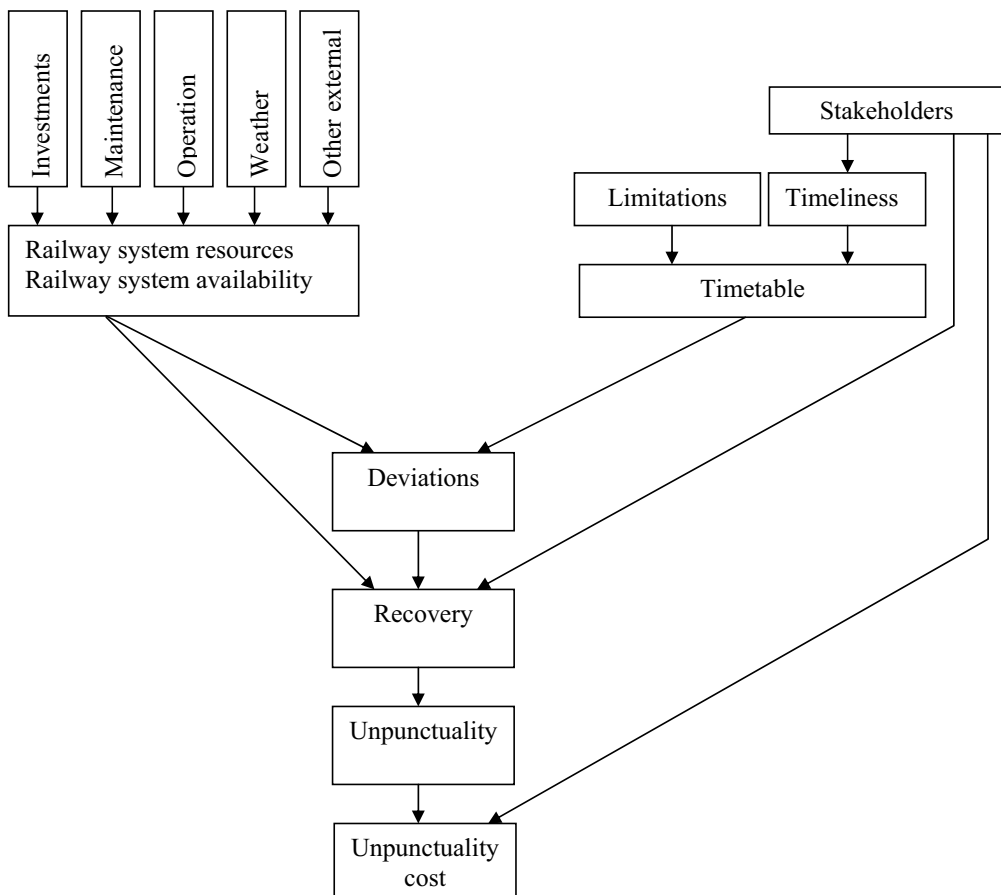


Figure 5. Factors influencing the unpunctuality cost. The resulting unpunctuality cost is found at the bottom. (The figure is inspired by Al-Haimi, 1991.)

2.5.1 Railway System Resources and Railway System Availability

The resources that the railway system has on hand are determined by many factors, spanning from which investments are made in infrastructure and rolling stock to external disturbances such as adverse weather, level crossing accidents and sabotage. Failures to railway infrastructure or rolling stock might lower availability. For example, a wagon with a wheel flat discovered underway, is to be taken out of the train and parked at a station. However, stations with only two tracks, main track and deviating track, then become temporarily unavailable. This is because the function of the station, to allow trains to meet and overtake, is disabled. Also, bad condition of track, for instance, worn rail or due to thawing of the frost in the ground, might force allowed train speed to be reduced, that is, availability to be lowered.

Track works, including investments in new items in the railway infrastructure, as well as preventive and corrective maintenance, might hinder train traffic, as in Figure 6.

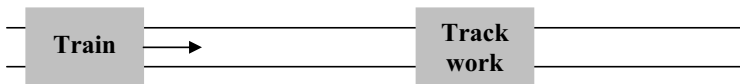


Figure 6. A train approaches a track work. If the track work has begun before the train is to pass, or the track work exceeds the time allotted to it, a train delay might result.

Large railway system resources, for example, plenty of standby locomotives or double-track instead of single-track along long distances, hinder delays from spreading to other trains and make recovery easier after delays.

2.5.2 Timetable

The timetable is constructed based on the limitations of the railway; for example, maximum allowed axle load and maximum allowed length of trains in the case of single track, as trains have to meet at stations of a certain length. Requirements regarding timeliness, that is, the requests of the TOCs and their customers regarding departure and arrival times, together with the requests of IM and MCs for track possessions, are inputs to the timetable construction process.

The timetable might be made less sensitive to small deviations by allowing a train a longer time than is strictly necessary to run a distance. Such allowances might be added to running times of the train en route or added to times at planned stops (Figure 2). Rudolph (2003) employed network simulation to choose how to distribute the allowances optimally.

2.5.3 Deviations

Current railway system resources and railway system availability, as well as the timetable, decide which deviations will occur. Deviations might be delays, but also other differences to what is planned. Examples are that a foreign object is on the

track, that a track work takes longer time than was anticipated or derailments, for example, of hazardous loads. Other examples are that a train departs or arrives before timetable, or that the maximum allowed axle load of a railway line is reduced, forcing some trains to be rerouted. Such a deviation might, in turn, cause other trains to be more disposed to unpunctuality as their railway line gets more crowded.

How and when the customers should be informed about delays is discussed by McLay (2000). For example, by learning about an unpunctual train in advance, a freight customer does not need to have a truck standing waiting for the train.

2.5.4 Recovery

Recovery means to eliminate or reduce the effects of deviations. Train traffic controllers, faced with deviations, try countermeasures such as to get trains to travel over timetabled speed, to reshuffle trains and to reroute trains. They might even cancel some trains for a part of their journey or along their entire route. For example, a commuter train might not travel the entire distance to its terminus in a distant suburb. The consequences for passengers are not as great if there are few passengers on the last part of the train's journey.

Recovery might also be performed by the passengers, who choose what to do, when, for example, their train travels only a part of its timetabled distance. The stakeholders' valuation of unpunctuality includes how each individual passenger (or freight customer) values unpunctuality, but also how large risks an individual takes, for example, whether a longer timetabled route is preferred to a shorter, but (allegedly) more delay prone, route.

The deviations occurred underway, the timetable (and other expectations, for example, delays occurred earlier), as well individual traits, are needed to describe a passenger's valuation ('price tag') of unpunctuality.

2.5.5 Unpunctuality

One should make clear which (un)punctuality is described. It might be the punctuality of the passengers, the part of the journey travelled by train, or one of the trains in the journey. In the case of measuring the punctuality of passengers, the punctuality of the passenger's whole journey, for example bicycle plus train plus bus, should be considered. Even a small train-arrival unpunctuality (i.e., a late arrival) might result in the passenger missing the bus connection and hence the passenger will suffer from larger unpunctuality.

2.5.6 Unpunctuality Costs

The occurred unpunctuality, and the valuations of unpunctuality of the stakeholders suffering from it, are needed in order to calculate the unpunctuality cost, as discussed in Section 2.3.

The author proposes that unpunctuality costs might be divided into external and internal. *External unpunctuality cost* strikes stakeholders outside the railway sector, most visibly passengers and freight customers, who become late for work, whose perishable goods become unsaleable, who have to stop production due to lack of raw material or spares, and the like. It also strikes, for example, freight forwarders and motorists waiting at level crossings (the costs to the latter are studied by Ryan, 1990). *Internal unpunctuality cost* strikes inside the railway sector, for instance TOCs, CTTCs and MCs. An example is the secondary delays, which might also hit other trains than those run by the TOC that runs the primary delayed train. Another example is when unpunctual trains cut into the time allotted for maintenance of the railway infrastructure, which might lead to, example given, costs for personnel overtime.

This description of unpunctuality costs is similar to the one related to the concept of poor-quality costs, used in Quality Management. Bergman & Klefsjö (2003) mean that poor-quality costs consist of *Internal failure costs* and *External failure costs*. Here, Internal failure costs are costs due to bad quality, which is detected internally within the organisation before delivery of the product to the customer. External failure costs are costs due to defective products detected after delivery to the customer. A difference from manufacturing is that in transportation, it is not possible to rework or substitute the product for a fault-free one just before the delivery to the customer.

2.6 Explanations of Unpunctuality

To reduce unpunctuality requires consideration of how causes of unpunctuality might be reduced or eliminated by different possible remedying actions. The performance of the infrastructure, rolling stock and customers (passengers and freight customers), are all needed to produce output from the railway, and causes for unpunctuality might therefore be due to inferior performance related to any of these.

The difficulty to identify one single cause of unpunctuality is illustrated by the following example. The causes of a train wagon derailing are: trivially, the wagon is on the track, the construction of rolling stock and track permits derailments, and the track is badly aligned. As an explanatory cause, one would probably choose bad track alignment, because it is a deviation from the expected circumstances. The bad track alignment might, in turn, be explained by lack of track maintenance. This, in turn, might be caused by improper funding principles. It is realised that it is often not possible to describe all causes; one has to make a choice, based on the intended use of

the information. For a deeper discussion about different classifications of causes in general, see Hesslow (1983).

Unpunctuality and accidents are similar, in the respect that both are deviations from what is planned. There is a need to have ways of describing, modelling and investigating deviations. Therefore, different accident models are briefly described below.

2.6.1 Accident Models

Accident models describe how, and explain why, an accident²⁰ occurred. They give a frame of reference for the accident investigation (Hollnagel, 2004).

Reason's (1997) pathogen model, or Swiss cheese model, is illustrated in Figure 7. Accidents are prevented by safety barriers, which hinder substandard acts²¹ from causing accidents. In order to prevent an accident, it is enough that one barrier catches the substandard act. A *latent failure* is when one or several of the barriers do not function, that is, an accident waiting to happen.

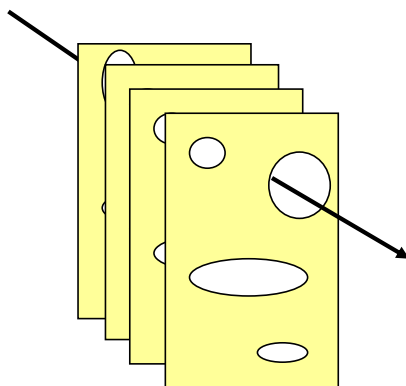


Figure 7. Reason's Swiss cheese model. If all barriers are penetrated, an accident occurs (Reason, 1997).

In early 20th century, personal traits were considered to make some workers prone to accidents (Kjellén, 2000). This theory was followed by Heinrich's (1959) domino theory, in which barriers hinder accidents, see Figure 8.

²⁰ An accident is an undesired event that results in physical harm to a person or damage to property (Bird & Loftus, 1976).

²¹ A substandard act is an act that deviates from the established standard, regulations or guidelines of the organisation (Groeneweg, 1998).

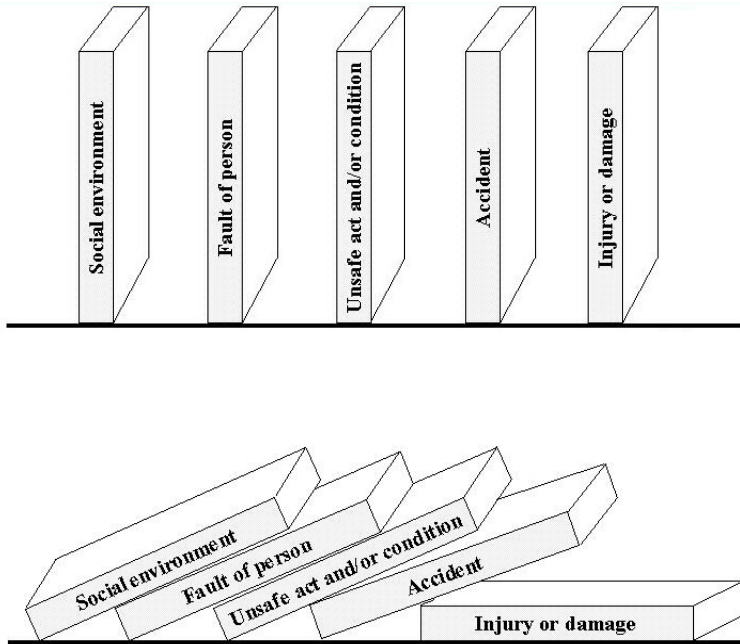


Figure 8. Domino theory. Successive barriers hinder accidents from happening. Injury or damage result only if all barriers fall. If one domino is removed, there will be no injury or damage (Heinrich, 1959).

The domino model has been the predecessor of many accident models, modified in order to account for loss in general (of people, property, environment, progress, among others). Bird & Loftus (1976) modified the left-most domino to *Lack of control*, defined as management not properly planning, organising, leading or controlling. One example of that is when management does not manage employee compliance to standards. Concerning the derailment example (page 24) the left-most domino might then be Improper funding principles, the next domino may be taken to be lack of track maintenance, the third as bad track alignment, the fourth is derailment and the last is the consequences (injury, damage, delay, bad-will). Holmgren (2006) employed the model of Bird & Loftus (1976) to classify maintenance-related incidents²² and accidents in the Swedish railway. Among his conclusions is that imperfect communication and information between maintenance personnel and train traffic controllers or TOC personnel is the most frequent cause of accidents manifested during maintenance execution.

Early domino theory has been criticised because it does not account for multiple causality. The so-called Finnish model for accident analysis is one model that accounts for multiple causes (Kjellén, 2000). It accounts for two chains of events, one in which a person comes in contact with the hazard, the other in which the hazard is

²² An incident is an undesired event that could (or does) result in loss (Bird & Loftus, 1976).

built up and released. The events in the chains may each have different contributing factors. The two chains of events join to cause the contact event, which, in turn, causes the injury event.

Accident models are classified by Hollnagel (2004) as being of one of the following types:

- *Sequential*: Events cause other events in a causal chain $A \rightarrow B \rightarrow C$, as in Heinrich (1959).
- *Epidemiological*: Agents and environmental factors, combined, create an unhappy setting to the barriers (defences) of the host. Latent (dormant) conditions, that is, conditions which are within the system well before the accident, but do not trigger accidents themselves, play an important role in models of this type.
- *Systemic*: Relations between parts in the production system cause variations and accidents must be expected. For this type of model, the distinction between sharp end and blunt end when describing the events leading to an accident is important. The sharp end refers to the people that are working at the time and place where the accident takes place, while the blunt end is the people who affect safety through their effect on the constraints and resources on the practices at the sharp end.

Train delays might emerge slowly, in which case there might be no easily identifiable cause. Another characteristic of delays is that events that do not occur might cause a train to be unpunctual. A simple example is lack of electric power. Unlike accidents, there is a tolerance, telling when a train is to be considered unpunctual.

The relations between the concepts deviation, unpunctuality, incident and accident are depicted in Figure 9. For example, Figure 9 illustrates that all accidents are incidents but not vice versa, and that unpunctuality is an example of deviation.

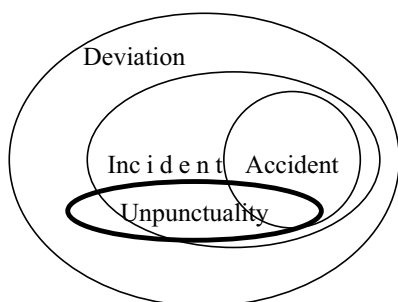


Figure 9. The relations between deviation, unpunctuality, incident and accident.

From the discussions, it is understood that a single cause is not always uniquely identifiable. The model guides the investigator in the enquiry and the underlying cause might simply be understood as where one chooses to halt. The deviations might be described regarding their role in the events associated with delays (such as

circumstance, action, symptom, consequence), where they occur (internal or external to an organisation, which regulation, which item) and which phenomenon (fatigue, lack of communication, flash-over). The three forthcoming subsections describe ways of looking at a deviation, which separately or in combination, might be used as explanations of unpunctuality and aid understanding of the situation.

2.6.2 Failures and Slips

Deviations from required function or work performance, for items²³ and humans, respectively, might cause unpunctuality. Anything from a tiny component in, for example, a turnout, to the entire railway system, might be struck by failure. Human deviations might range from individual reflexes to group decisions.

For humans, *slips* and *lapses*²⁴, that is, actions which fail to be carried out as planned, although the plan²⁵ is adequate, may occur (Reason, 1997). Examples of slips and lapses are that maintenance personnel, in a hurry, bring inaccurate spares to the place of a fault, or, while cleaning a signal box, accidentally crush components with the broom-stick, or that a locomotive driver oversleeps.

2.6.3 Unpunctuality Drivers

An *unpunctuality driver* is in this thesis defined as any factor that affects unpunctuality²⁶.

In a statistical investigation by Johansson & Carlsson (2003), adverse weather is found to partly explain train unpunctuality, that is, adverse weather is an unpunctuality driver. Having identified such an explanation does not necessarily mean knowing the underlying mechanism, although it might give clues. Heavy snowfall might impede train traffic, but might also slow maintenance work. Other examples of possible unpunctuality drivers are day of the week, type of train, time to repair, type of maintenance contract and utilisation of the railway line.

2.6.4 Human Issues

Human issues present explanations for unpunctuality different from, but complementary to, the ones presented in Sections 2.6.2 and 2.6.3. Examples include beliefs, attitudes, motivation, knowledge, cooperation, leadership and group behaviour. As an example, assume that a train is delayed due to the fact that the locomotive driver was given an inaccurate route order. Probably this was caused by a

²³ In the case of items, an item is said to have a fault, as defined in Section 2.4.1.

²⁴ Slips relate to observable actions and are commonly associated with attentional or perceptual failure. Lapses are more internal events and generally involve failures of memory, according to Reason (1997).

²⁵ A *plan* includes goal and means to achieve it.

²⁶ This definition is analogous to the term *cost driver*, defined in accounting as *any factor that affects cost* (Horngren et al., 1994).

*mistake*²⁷ by personnel at the TOC's office. If this happens more frequently for a certain TOC than for others, the TOC is an unpunctuality driver. The delay might be explained in terms of lack of knowledge, for instance, that TOC office personnel did not know that the driver in question did not have a driver's license for that type of locomotive. This, in turn, might have been caused by the fact that the personnel have not been adequately trained, or that it was assumed that someone else in the organisation was responsible for managing this issue. The author wishes to emphasise here that what is called human issues, and sometimes human failures, are not a result of malevolence. Human issues are due to deficiencies in the systems we work with, systems that rest on managers' shoulders. Juran and Deming both state that about 80-90% of the failures and mistakes that are made are due to managers (Deming, 1986; Juran, 1989).

An issue is the attitude towards improvements in general among stakeholders; for example, whether punctuality problems are seen as permanent evils or possibilities for improvement, whether problems are viewed as local or in a wider context, as exemplified by the empty-wagons problem described in the Section Introduction. The existence of a punctuality culture, that is, that people always think punctuality, is difficult to quantify. One can strive to foster it by, for instance, making certain that each driver sees an accurate clock from the driving cabin, where the time to departure is counted down second-wise, as is done by A-Train (Nyström & Karlsson, 2006).

Even passengers' behaviour affects punctuality. One example concerns the Arlanda Airport train, where the automatic doors are closed some seconds before timetabled departure time. Passengers, accustomed to other TOCs, have complained about this, as they see it as having caused them to miss the train (Nyström & Karlsson, 2006). However, by closing the doors in time, A-Train communicates to its passengers and personnel that timetabled departure time is to be respected, thereby instilling a certain attitude.

From Sections 2.6.2, 2.6.3 and 2.6.4, it is understood that the same failure might be viewed, and thus investigated, from several perspectives. For example, a certain kind of component failure that occurs frequently, although being of a well-known type, can be viewed in terms of the individual component being unreliable, from the perspective that having many of these components along a railway line is an unpunctuality driver, or that an organisation that introduces new technology without testing it appropriately causes unpunctuality, or even that the maintenance organisation does not learn from well-known failures. A description of the delays occurred over a period of time has to rely on the data that has been reported into databases. Therefore, data that presently is used to describe delays and their causes in the Swedish railway are now described.

²⁷ When humans fail to formulate an adequate plan, a *mistake* might occur (Wickens, 1992). Note the difference to slip, defined in Section 2.6.2. A mistake, for instance, intending to press the wrong button, might be cancelled out by a slip, for instance pressing another button than intended.

2.6.5 Delay Attribution Codes and Measurements

Banverket stores data on Swedish train delays, for all TOCs, in the TFÖR database. All extra delays, that is, the increase in time-lagging timetable between two places, usually stations, of five minutes or more, are flagged, indicating that a delay attribution has to be made. The delay attributors have about 100 codes to choose from when they report a delay, some of them requiring a free text report to be filled in. The structure of the codes, showing groups and sub-groups, is shown in Table 2 (Banverket, 1999). For an in-depth description of how infrastructure-related faults are attributed, see Paper I and Wahlström (2006).

Table 2. Classification of delays in groups and sub-groups according to TFÖR (Banverket, 1999).

<ul style="list-style-type: none"> • Planned track works • Train traffic control <ul style="list-style-type: none"> • Track works • Train traffic • Personnel • Incident/Accident • Train operating company <ul style="list-style-type: none"> • Departure/Dwell • Delivery • Shunting • Train composition • Personnel 	<ul style="list-style-type: none"> • Vehicle • Infrastructure <ul style="list-style-type: none"> • Track work • Signalling • Track • Power • Tele/radio • Other infrastructure • Other
--	--

Which TFÖR code to apply might be disputed in certain cases, for example, whether the pantograph (the code ‘Pantograph’, which belongs to the Vehicle group) or the contact wire (‘Catenary fault’ in the Infrastructure group) should be attributed. The expressed praxis for attributing a delay to pantograph or contact wire has changed over the years (Granström & Söderholm, 2005). However, the share of delays within each train-infrastructure interface should reflect the reality in a better way: pantograph-contact wire (6%), wheel-rail (3%), signalling (3%) and externalities (10%)²⁸. The remainder of the delays are “interior”, in essence, not at an interface. However, this does not necessarily mean that these are correctly coded from all perspectives. For example, a delay coded in TFÖR as ‘Turnout failure’, may have its origins in the interaction of rolling stock of different types and conditions, taken together with the characteristics of the turnout and the nearby environment (Nissen, 2005).

The TOC A-Train uses two databases on punctuality. Banverket’s TFÖR is employed, but as A-Train is not certain where the points of measurement are located

²⁸ The percentages are calculated from TFÖR data for 2007. The figures cover primary delays. Of the primary delay minutes in TFÖR, 7% were not attributed. Primary delay is explained in Section 2.2.

and the time precision might not be sufficient, A-Train also uses another, company-specific, measurement system. This system utilises reports that the drivers are obliged to fill in for each journey, where unpunctuality and other deviations are registered. The precision is 0.5 minutes, compared to TFÖR's 1 min²⁹ (Nyström & Karlsson, 2006).

For trains crossing national borders, for example the fish train from northern Norway to southern Norway via Sweden, train delay data might be difficult to find, as data from railway bodies in several countries might have to be used (Paper I) or dedicated measuring systems, for example, satellite navigation, might have to be installed on the trains (Rieckenberg, 2004).

Delay attribution codes for Banverket, A-train, the UK, the Japanese railway JR East and for air traffic are described in Appendix C.

2.7 Comparisons of Punctuality

In a benchmarking study, train punctuality in fourteen countries is compared by BEST (2002). It attributes its difficulties in doing so to different countries employing different tolerances³⁰, to whether departure or arrival time is considered and to whether all stops or only arrival at the end station are considered.

BOB (2003), a sister project to BEST (2002), observes that 5 minutes, or more, late to end station is a common definition of delayed train, and therefore converts other measures, such as 3 minutes late to end station and average delay, into this measure. This is done by assuming that the probability distribution of length of delays is the same as for NS, Nederlandsche Spoorwegen. Calculated in this way, the punctuality ranges from 84.5% to 98.3% for the studied railways (BOB, 2003). As a comparison, 73% of the flights departing or arriving at Arlanda Airport are punctual. Nevertheless, 48% of the passengers are delayed. An explanation for this is that the charter flights, which most often are full or nearly full, are almost never cancelled, regardless of long departure delays (Luftfartsverket, 2000:1).

Data on causes of delay differ much between railways in different countries, which makes it difficult to carry out fair comparisons. The share of delays caused by railway infrastructure ranges from 5% to 45% among the railways studied (BEST, 2002). The large difference indicates that the delay attribution systems, more than the factual circumstances, differ between the countries.

²⁹ Note that a precision of 1 minute for arrival time and departure time of a train renders the precision of length of dwell time to be 2 minutes.

³⁰ The term tolerance was introduced on page 11.

2.8 Unpunctuality Cost versus Maintenance Cost

A maintenance action should be motivated by its utility; for example, when the unpunctuality cost is sufficiently reduced as a consequence of the maintenance action. Figure 10 illustrates the total cost as a function of the maintenance volume. The different costs are calculated and added to obtain the total cost as a function of maintenance volume. Then, the maintenance volume that minimises the total cost is chosen. To improve punctuality, other means besides changing maintenance volume should also be considered, see Figure 5.

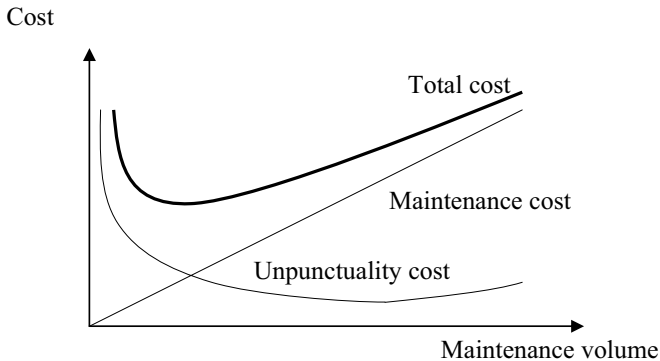


Figure 10. An on-the-spot account of the total cost as a function of the maintenance volume. The maintenance cost and the unpunctuality cost add up to total cost. Increasing the maintenance volume from zero, causes the unpunctuality cost to decrease until a certain volume of maintenance has been reached. Additional increase in maintenance volume causes the maintenance to interfere with traffic to a larger extent and the unpunctuality cost thus increases.

Maintenance optimisation requires calculations regarding when to carry out maintenance of railway infrastructure, both with respect to the life-length of the item (Jardine & Tsang, 2006) and with respect to scheduling of trains and infrastructure maintenance, studied in the operations research tradition (see Hessfeld et al., 1996; Higgins et al., 1995). Kufver (2002) gives an overview of the choice of maintenance methods with regard to how long they hinder train traffic. Choosing a more costly method to carry out a certain maintenance action might be motivated if it interferes less with traffic.

However, some objections could be raised to a maintenance optimisation scheme as in Figure 10. First, it is not certain that minimum total cost gives the highest profit, because demand for train traffic might be affected by the punctuality of the service. Second, maintenance optimisation might be contrasted to setting goals, that is to say, to create indicators and formulate goals using these. Goals might be of two different kinds; the *process goals* show the way to achieve the *result goals*. An example of result goal is that a specified punctuality is to be achieved, and an example of process goal is that the driver's cabin should be kept tidy so the driver easily finds the timetables, for example. Third, and perhaps most important, is to understand that the

figure is an “instant picture”. Over time, changes in, for example, technology, organisation and valuation of punctuality, and not least, added experiences among the stakeholders from the improvement work, mean that the cost picture changes and accordingly the curves. This is discussed by Bergman & Klefsjö (2003).

3 Research Design

In this chapter, some options for performing research are introduced. Also, the concepts of reliability and validity of a study are discussed. Thereafter, rationales for the performed methodological choices made in the research described in this thesis are related to the stated research questions.

3.1 Research Strategies

Yin (2003) describes five different research strategies to apply when collecting and analysing empirical evidence. Table 3 shows these research strategies and three criteria for determining the use of each of them.

Table 3. Criteria for selecting an appropriate research strategy (Yin, 2003).

Strategy	Form of research question	Requires control over behavioural events?	Focus on contemporary events?
Experiment	How, why	Yes	Yes
Survey	Who, what, where, how many, how much	No	Yes
Archival analysis	Who, what, where, how many, how much	No	Yes/No
History	How, why	No	No
Case study	How, why	No	Yes

The first criterion for selecting an appropriate research strategy is related to the way in which the research question is formulated (Yin, 2003). The main research question of the research presented in this thesis includes “how”, see Section 1.3. Following Table 3, this means that experiment, history and case study might be appropriate strategies. However, one strength related to the case study approach is that it can encompass the other research strategies (Yin, 2003). Hence, the case study strategy was selected as the overarching research strategy.

The organisation, which is in focus in this case study, is Banverket, due to its large role in the railway sector, as seen from Figure 1. The following organisations have also had important roles within the case study: A-Train, CargoNet, EuroMaint, Green Cargo, Jernbanelverket, MTAB, The Norwegian Post, Strukton Rail and SSAB.

Within the overarching case study, other research strategies, such as experiment, survey, and archival analysis are applied, based on the more detailed research questions. The rationales for the choices that were made in the research presented in this thesis are described in relation to each separate research question later in this chapter. In order to support these research strategies dealing with empirical evidence, a literature study was continuously applied in parallel. The literature study covered different aspects of the studied phenomenon (for example punctuality, availability, delay and maintenance). However, the literature study also covered methodological approaches for studying these phenomena, for example, methodologies and tools for data collection and analysis, such as the Seven Management Tools (Mizuno, 1988) and the Analytic Hierarchy Process (Saaty, 1980).

3.2 Data Collection

Yin (2003) gives six main sources of evidence to apply in a case study; archival records, direct observations, documentation, interviews, participant-observations and physical artefacts. The sources of evidence should be viewed as complementary and it is advisable to use multiple sources of evidence in order to achieve a triangulation, to strengthen the validity (Yin, 2003). In this research, interviews, observations, documents and archival records were used to extract empirical evidence. Some of these sources of evidence were also used to support the other applied research strategies; survey and experiment. Later in this chapter, this is described in further detail in relation to each stated research question.

The literature study was performed with a number of search engines, mainly Google Scholar, Scopus, Compendex, Emerald and Elsevier Science Direct. The searches were triggered by relevant keywords identified through the project description that was approved by the funding agency. In order to reduce the vast number of hits, a reduction process was performed. This reduction consisted of three main steps, where the first step was related to the reference titles, the second to their abstracts, and the third step to the whole or parts of the references. The result of the literature study can partly be found in the theoretical framework of this thesis and its reference list, but also in the reference lists of the appended papers and research reports & documentation filed at the university. The most relevant keywords used in the performed literature study can be found in the appended papers.

3.3 The Concepts of Reliability and Validity

Reliability denotes that a measurement is possible to repeat and *validity* denotes that a measurement actually measures what it is intended to measure. Reliability means that the operations of a study, for example, the data collection procedures, may be repeated by another researcher with the same results (Yin, 2003). Validity has several facets. *External validity* denotes that the findings of a study are possible to generalise (Yin, 2003). *Content validity* is defined as the extent to which an empirical measurement reflects a specific domain of interest (Carmines & Zeller, 1979).

In order to strengthen the validity of the performed research, a multitude of tactics was applied. One applied tactic was to establish a chain of evidence, by entering of documents in the university's records (such as bulky empirical material) and publishing at the university's library (analysed empirical and theoretical materials in research reports). This recording also strengthens the reliability. To further strengthen the validity, multiple sources of evidence were applied, for example, observations, interviews, archival records and document studies. Also, key informants have been allowed to study drafts of reports and papers.

3.4 Methodological Choices in this Thesis

Here, the performed methodological choices are discussed for each of the research questions stated in Section 1.3.

3.4.1 Methodological Choices for Research Question A

Research Question A is stated as: *How should (un)punctuality be described?* This research question asks how unpunctuality should be described, and emphasises measurement issues.

3.4.1.1 Research Approach

Literature study was the primary method used to answer Research Question A, complemented with empirical evidence, which was collected through interviews with stakeholders.

Another approach could be to answer the research question solely from the view of unpunctuality costs or, alternatively, from the view of psychological impact of unpunctuality. However, these approaches were only used to a limited extent.

3.4.1.2 The Study

The objects of study for Research Question A were descriptions of punctuality. Aspects of how punctuality is described for different railways, air traffic and the brewing industry were extracted by literature study and interviews. Descriptions of punctuality were exemplified and their pros and cons discussed.

3.4.2 Methodological Choices for Research Question B

Research Question B is stated as: *Which problems are encountered when trying to reduce unpunctuality?* This research question focuses on the problems, or obstacles, that are encountered when trying to improve punctuality by a reduction or elimination of unpunctuality causes.

3.4.2.1 Research Approach

The problems might be seen from different system perspectives. In order to give a wide view of the problems, different perspectives were used as vehicles to discover problems. The chosen perspectives from which to investigate unpunctuality problems were that of a certain train and that of a certain infrastructure asset. Since the two studies were of an explorative nature, any research strategy given in Table 3 could

have been applicable. However, due to accessibility, empirical evidence and archival data were collected through the Swedish railway.

Observations could have been made of, for example, punctuality improvement groups consisting of representatives from several stakeholders in order to find which problems they encounter. Also, document studies, for example, minutes from these groups, could have been used to identify issues. However, the author finds that the minutes are often too brief, as they do not thoroughly describe the problems and how they were dealt with. Direct observation and participant-observation might work well to identify practices, and, if used in several settings, to identify the more successful practices. However, this would have taken a long time, which was not available within the given time constraints. So, it was decided to investigate contemporary problems. The drawback with this was that it was not known beforehand whether any problem would be solved. On the other hand, the researcher retained some control and had the possibility to try different methods.

3.4.2.2 The Studies

Two main objects of study were related to Research Question B. In Paper III, a single-commodity train was studied and in Paper VII, one specific signal box was studied.

The single-commodity train was selected for its economic importance. By observations and interviews with managers, train drivers, factory personnel and yard workers, as well as by using databases on failures and delays, the causes for unpunctuality were explored and described. For further details, see Paper III.

The studied infrastructure asset, one specific signal box, was selected because of frequent failure recognition of one of its components (compared to similar signal boxes) without any corresponding cause identification. Differences between the signal box under scrutiny and other signal boxes played an important role in the study. By using partly redundant archival records of, for example, train traffic and infrastructure failures, as well as interviews, document studies, observations and measurements on site, empirical evidence was collected. For further details, see Paper VII.

3.4.3 Methodological Choices for Research Question C

Research Question C is stated: *How might availability of the railway infrastructure be described?* The term ‘availability’ is often used in the railway sector, without further specification. Availability might simply be viewed as whether a certain item, such as a specific metre of track, is ready to take a train forward, but because, for instance, the production of train traffic is not continuous, this does not give a useful measure of the availability of a larger part of the railway. Indicators related to availability were the objects of study for Research Question C.

3.4.3.1 Research Approach

In order to answer Research Question C, indicators of availability in the Swedish railway were collected through document studies. In addition, via a literature study, different indicators of availability from railway and a few other industries were identified. Thereafter, the availability indicators from the document and literature studies were compiled and the non-railway indicators were adapted for a railway context. Finally, the constructed list of availability indicators were assessed through a questionnaire aimed at practitioners within the Swedish railway.

Another possible strategy might have been to carry out a survey, in which the respondents are asked to write what they consider to be relevant indicators. This would have necessitated the researcher interpreting what was written, which would have lowered reliability and validity. Yet another possible strategy would have been to calculate currently defined indicators and look for relations between them, thereby identifying the most informative indicators. However, this would have had the drawback that a large amount of work would be needed to find the appropriate data and do the computations, as well as that only indicators currently defined would have been possible to test.

3.4.3.2 The Study

The study related to Research Question C is presented in Paper IV. It turned out to be rewarding to use the availability indicators of the electric power industry as analogies. This is due to the similarity in stakeholder structure between the electric power industry and the railway. The indicators were presented in a questionnaire, and the personnel at one CTTC³¹ were asked to judge the relevance of each indicator on a Likert scale. Thereby, the indicators that many of the personnel considered to be relevant, irrelevant and non-understandable were indicated.

3.4.4 Methodological Choices for Research Question D

Research Question D is stated as: *What is the suitability of the information, from systems measuring punctuality and maintenance, for improvement work?* Databases on infrastructure and punctuality are often used by individual managers, track workers and improvement groups. The users enter and/or extract data to be used by themselves or others. This places high demands on the suitability of the data; that is, data is appropriate for the intended use and it is correct. The research question was answered by performing two studies: one study was related to maintenance information and the other to punctuality information.

3.4.4.1 Research Approach

Applying Yin's (2003) criteria for selecting an appropriate research strategy, the word "what" included in Research Question D indicates that both a survey and an archival analysis were appropriate, see Table 3. At the same time, it would have been neither possible nor desirable to control behaviour events, and the focus was on

³¹ Personnel from CTTC Boden participated. Of 30 questionnaires, 27 were returned filled in.

contemporary events, which all supported both strategies (see Table 3). Hence, both an archival analysis and a survey were selected as appropriate research strategies to answer Research Question D. An archival analysis, combined with interviews, was selected to cover the maintenance part of the research question, while a survey was selected to cover the punctuality part.

Another way to gather information related to punctuality and maintenance, respectively, might have been to analyse the information upon which the improvement groups base their discussions, for example, by the use of a survey. Among the difficulties with this approach are that the topics vary from meeting to meeting and either are on a highly aggregated level or that only a specific item is focused upon. Another possible way would have been to look for anomalies in databases through an archival analysis, the drawback being that the accuracy of the possibly contradictory data is not known. Yet another way would have been to follow individual users of databases, the drawbacks being that they spend only a small share of their working hours in front of the computer and that interviewing might impact the behaviour of a user. A broad way might have been to distribute a questionnaire, asking which problems the users have experienced. A difficulty in this connection would have been that the users are often specialised, for example, repairers, data enterers and analysts, who may not see much of the discrepancies between reality and data interpretation. Also, a practical problem with the questionnaire approach would have been to identify suitable respondents, as many use databases only a little and the users are geographically scattered.

3.4.4.2 The Studies

By comparing the descriptions of recently occurred failures, found in the Ofelia database, to the descriptions obtained by interviews with maintenance personnel, the quality of the recorded data was assessed through triangulation. This was complemented by the use of Ofelia and by documentation. Hence, the research strategy related to the maintenance part of Research Question D was mainly an archival analysis covering the database on infrastructure failures, Ofelia. For further details, see Paper I.

Delay attributions were the objects of study for the research studying the punctuality part of Research Question D (for further details, see Papers V & VI). The purpose was to analyse to which extent delays are consistently attributed in the railway, in essence, to measure to which extent different persons report similar delays in similar ways. By archival studies it would have been possible to find data on how delays have been attributed, for example over time. However, to which extent such data reflects a changed reality or a changed way of perceiving and reporting this reality would have remained unknown. By carrying out experiments with delay attributors, their way of reporting different delays might have been investigated. But, as the delay attributors should be presented to realistic scenarios that develop on the computer screen, possibly with telephone calls with, for example, train drivers, as well as chats with colleagues about the appropriate delay attribution, such a study would have been

cumbersome and taken a long time. Therefore, it was decided that interviews would be used to develop a questionnaire with many cases (vignettes) and to perform a survey.

3.4.5 Methodological Choices for Research Question E

Research Question E is stated as: *How is the choice made between different maintenance actions?* In a document study of the maintenance process in Banverket's northern region, it was noticed that a large share of the indicators given for its sub-processes to a large extent measure the performance of the maintenance process as a whole, not the performance of the sub-process, which was stated in the documents (Nyström, 2005). Furthermore, in the sub-process which prioritises among prospective maintenance actions, only the expenditures of the actions themselves are quantified. Future costs and benefits (for instance punctuality) are not assigned figures. The prospective maintenance actions, and sometimes their consequences, are only briefly documented. These observations motivated a larger study, delving deeper into this decision-making.

3.4.5.1 Research Approach

Through an archival study one would have found the decisions made, but not their rationale, as these are usually not documented. Documents detail the performance criteria that Banverket's head office considers important. Alternatively, a case study on a specific maintenance action may have revealed the criteria advocating the action to be carried out, but not the ones opposing. To follow the decision-making process, by means of direct observation and interviews with the decision-makers over a long period of time (months), might have revealed what they consider important, as the decision-makers talk to colleagues and search for information. Still, in order to find out what is not important, the researcher would have had to pose questions to the decision-makers, thereby possibly affecting the decisions. An experiment would have allowed control of the alternatives and would have given the possibility of asking the same question in different ways, thus making it possible to estimate the consistency of decisions. However, an experiment in which the alternatives are controlled would not have given the possibility to study the behaviour of decision-makers faced with real problems. All in all, it was decided that criteria (effects of maintenance) be collected by means of a group interview and document studies, alternatives (maintenance actions) by document studies and data on decisions by experiments.

3.4.5.2 The Study

Maintenance decisions are the objects of study for Research Question E (for further details, see Paper VIII). Track managers choose between different actions when making maintenance decisions. These decision-makers have to subordinate to the allocated budget (made by decisions) and Banverket rules & regulations (also made by decisions). However, a track manager makes choices between distinct maintenance alternatives. This made these decisions easier to study than budgeting and regulation-writing, which have many more alternatives and are performed by many persons over a longer time.

The study included the analysis of empirical data from documents, interviews and experiments. Theories, relevant to the study, were related to decision-making and the Analytic Hierarchy Process, AHP (Saaty, 1980). The applied methodology is illustrated in Figure 11. In a group interview, decision-makers were asked which factors are needed to characterise any maintenance action. Then, they grouped the factors hierarchically using affinity diagrams (Mizuno, 1988), which were later supplemented with measures from Banverket’s strategic plan (Banverket, 2006). The factors on the highest level of aggregation are here called criteria. Using these criteria, experiments with individual track managers were carried out. In the experiments, each track manager prioritised maintenance actions, specific to the track manager, using two methodologies; ranking by criteria and ranking by alternatives. The ranking by criteria was calculated using the priorities of the criteria and the priorities of each of the alternatives with respect to each criterion. The ranking by alternatives was calculated using direct comparisons between the alternatives. The rankings and information from the interviews form the input to the analysis.

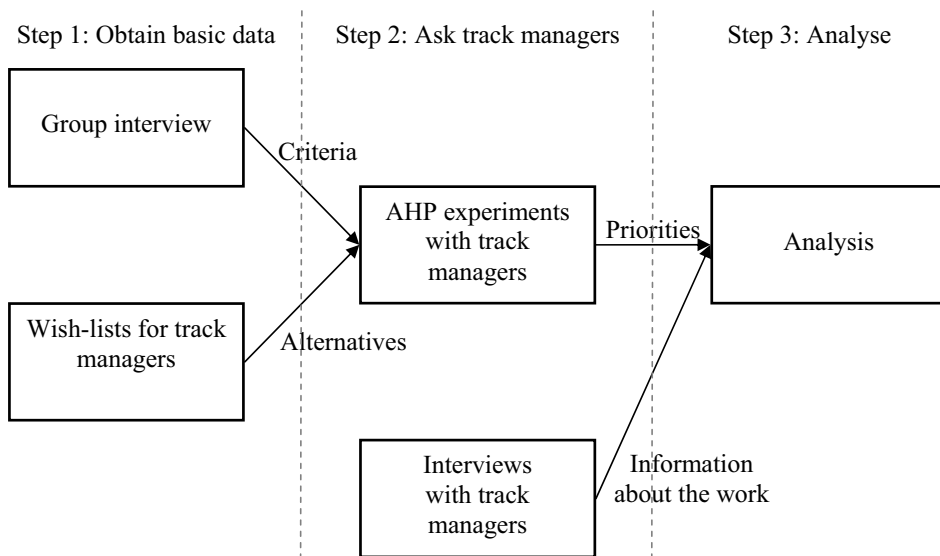


Figure 11. The applied research methodology began with a group interview that gave the basis for the criteria needed for the AHP experiments carried out with track managers. Subsequently, the resulting priorities were analysed, taking into consideration the information about their job, given by respective track manager in the individual interviews.

4 Description of Appended Papers

In this chapter, the appended papers are summarised. First, the papers are classified. Then, each paper is described regarding its purpose, study approach and findings.

The appended papers might be classified with respect to how practical or theoretical the investigated phenomenon is and with respect to how closely related it is to maintenance execution or unpunctuality cost. This is illustrated in Figure 12. One sees that the entities on the vertical axis might be put in different boxes in the earlier discussed Figure 5, which shows factors influencing unpunctuality cost. For example, in Figure 5, Unpunctuality and Unpunctuality cost are found in the two lower-most boxes.

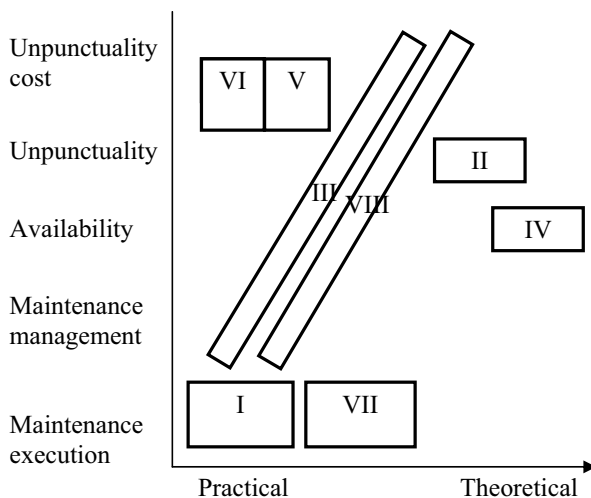


Figure 12. A classification of the appended papers.

The eight appended papers are summarised on the following pages, and are found in their entirety in Appendix E.

4.1 Description of Paper I

Nyström, B. & Kumar, U. (2003). Analysis of train delay information. In: *Proceedings of the World Congress on Railway Research (WCRR 2003)*, Edinburgh, Scotland, 28 September - 1 October, pp. 253-261.

4.1.1 Purpose of Paper

The purpose of the paper is to describe the judgment of accuracy and suitability of train delay and railway infrastructure failure data, for use primarily in maintenance, and to suggest improvements.

4.1.2 Study Approach

Databases and their interfaces, relevant for railway infrastructure failures, are studied. Personnel involved with corrective maintenance are interviewed regarding what happened in the case of specific records in the Ofelia database of infrastructure failures.

4.1.3 Findings

Data on train delays rests mainly on data collected by a nationally uniform automatic system, based on a division of railway lines in fixed blocks. The hereby generated train traffic data gets into the TFÖR database and hence is possible to investigate by use of the MAPS tool. However, the train traffic control system, used by train traffic controllers to manage the traffic, is unique to each CTTC in Sweden. Because the train positioning system, in some places, measures departure and arrival times at other locations than by the platform, the recorded data does not give an entirely accurate picture of the times of train arrival and departure, but the variation in travel time is reliably measured. The accuracy of data might be further improved, as one could adjust times based on a specific train's acceleration/deceleration data.

The entry of infrastructure failures into the Ofelia infrastructure failure database or the Bessy infrastructure inspection remarks database is not entirely consistent. Faults are not indicated to be train-stopping or not. The Ofelia database focuses on detailed item description, but the events leading to failure and delay are not well described in it. Repair difficulties affect remedial time, sometimes very much, which is seen from Ofelia data. However, the reasons for the difficulties are not described. Ofelia does not allow coding of multiple causes. Data fields to be inputted, by the use of scroll-down-menus, are 'Real fault', 'Cause' and 'Remedial action'.

A large proportion of the delays do not have a cause reported, which might overturn the ranking of causes of delay. Uncertain causal relationships are unavoidable; what can be done is to facilitate integration and ease of use of data sources.

Only delays of five minutes or longer get a cause attributed. This means that an investigation of the causes of slowly emerging delays is hampered. However, to

lower the threshold would increase the work load of the delay attribution personnel and thus lower the information quality.

The difference in train traffic over time and place, together with the time needed for repairs, might be an explanation for the differences in delay time between different causes. However, MTTR (mean time to repair) and average delay show low correlation among principal causes of infrastructure-related delay. This may possibly be attributed to the factor of time to repair distributions being highly skewed.

4.2 Description of Paper II

Nyström, B. & Söderholm, P. (2005). Improved railway punctuality by effective maintenance – a case study. In: Kołowrocki, K. (ed.), *Advances in Safety and Reliability*, pp. 1487-1491, Taylor & Francis, London. ISBN 0 415 38340 4. Presented at ESREL, Gdansk, June 2005.

4.2.1 Purpose of Paper

The purpose of the paper is to describe the analysis of requirements, related to punctuality, concerning how they are stated, measured and enforced by different stakeholders.

4.2.2 Study Approach

The study approach is to follow the requirement for punctuality starting from an end customer, who sends a letter by the Post. By means of interviews and document studies, data is collected in the transport chain.

4.2.3 Findings

As a basis for identification of requirements, the following flow of requirements was identified: The Post – TOC³² CargoNet – IM³³ JBV³⁴ traffic division – IM JBV infrastructure division. The requirements go in both directions of this hierarchy of the stakeholders³⁵. A chain of requirements is illustrated in Table 4 & Table 5. Inside the bold outlined boxes are requirements as measured by the stakeholders. The other, corresponding, requirement is formulated by the authors.

³² Train operating company.

³³ Infrastructure manager.

³⁴ JBV is short for Jernbanelverket.

³⁵ A stakeholder is “an interested party having a right, share or claim in the system or in its possession of characteristics that meets that party’s needs and/or expectations” (ISO/IEC 15288:2002).

Table 4. A chain of requirements. Inside the bold outlined boxes are requirements as found in the study. The other requirement is formulated by the authors.

Stakeholder	Requirement to stakeholder
Norwegian Post	Minimum 85% of Priority Class letters delivered over night (national average per calendar month).
CargoNet	90% of trains on time (< 1 min) to station (or customer's) and unloading has begun.
JBV – traffic division	90% of trains on time (< 5 min) to end station.
JBV – infrastructure division	99.999% of functioning trains pass 10 km of track without contact wire problems per year (approximate).

Table 5. A chain of requirements. Inside the bold outlined box is a requirement as found in the study. The other requirements are formulated by the authors.

Stakeholder	Equivalent requirement
Norwegian Post	Maximum 15% Priority Class letters handed in on time still belong to 'remaining post' in the morning after (national average per calendar month).
CargoNet	Train departure, underway transport and unloading delay totals ≥ 1 min for at most 10% of trains.
JBV – traffic division	10% of trains ≥ 5 min late to end station.
JBV – infrastructure division	Contact wire < 0.2 failures giving operations difficulties per year and 10 km of track.

The current performance measures and information used are not entirely suitable concerning feedback in general and for managing availability of railway track and thus its maintenance. Few requirements are tied to performance regimes; for example, unpunctuality caused by track works is penalised. However, unpunctuality due to track failures is not penalised. No stakeholder has a performance regime for 'track and trace' (that is, information on location, speed, and so on, of goods, carrier or vehicle).

All stakeholders except the Post consider an event to be "punctual or not", without considering the length of the delay. However, also considering the length of delay

would make it easier to account for the incurred costs, for example, in performance regimes. A more complete performance regime should also include that CargoNet and JBV traffic division clearly state in their agreements when force majeure applies. As seen from Table 4 & Table 5, JBV infrastructure division formulates its requirements in a ‘positive’ (that is, availability) way, the other stakeholders in a ‘negative’ (that is, faults) way.

4.3 Description of Paper III

Nyström, B. (2006). Delay analysis of a freight train – an improvement case study from a steel company. Submitted for publication.

4.3.1 Purpose of Paper

The purpose of the paper is to describe the identification of causes of unpunctuality of a single-commodity train, as well as suggestions of actions to give the greatest improvement in punctuality. It is also to discuss the feasibility of the procedure employed for investigation.

4.3.2 Study Approach

The transports of a company, producing steel slabs in one factory and rolling them in another, are investigated. The transport chain is described and the journey by train is analysed regarding punctuality and causes of unpunctuality. This is done by using interviews, databases and documents.

4.3.3 Findings

Large unpunctuality (in essence, long delays), might be caused by, for example, failures to infrastructure items such as turnouts, and/or by adverse weather. Small variations in train travel time, which in the short run might cause delays and in the long run cause timetabled time to be prolonged, are caused by, for example, minor locomotive problems or low contact wire voltage due to nearby trains.

It is not possible to calculate the unpunctuality cost to the factory receiving the slabs as a function of a train’s arrival unpunctuality. This means that it is hard to motivate punctuality-improving actions solely on the grounds of immediate benefit for the freight customer. However, in some cases, the punctuality might be improved by simple means. It is important to focus on departure punctuality, because it has a strong correlation to arrival punctuality. Instead of discovering faulty wheels with a certain wheel damage detector when the train is departing loaded, it would be better to adjust the detector to be able to do so also when the train is returning unloaded. The availability of infrastructure items differs between the route studied and the national average, for example, the turnout failure rate is about twice the national average. However, it remains to be investigated whether this reflects a hidden improvement potential along the route, or is due to, for instance, a higher traffic load.

The first step in the suggested procedure, to identify the train to investigate, is not facilitated by the current database layout. Therefore, the first step is not as simple as one may assume. The slack in the train's timetable is distributed unevenly along its route. Therefore, it is a good idea to study the variations in travel times. Delay attribution codes keyed in wrongly, and inconsistent coding practices, make the identification of causes of unpunctuality more difficult and unreliable.

4.4 Description of Paper IV

Nyström, B. (2008). The use of availability concepts in the railway system. Submitted for publication.

4.4.1 Purpose of Paper

The purpose of the paper is to describe indicators related to railway availability on a system level, from the perspectives of the railway stakeholders.

4.4.2 Study Approach

Indicators related to availability are collected within Banverket (the Swedish Rail Administration) and in railway literature. Also, in order to find fruitful analogies, literature concerning a few other branches of industry is surveyed for availability indicators. Personnel in the Swedish railway are asked to judge the relevance to their job of the indicators encountered or constructed during the study.

4.4.3 Findings

Some availability indicators are defined with respect to capacity and some with respect to punctuality. Indicators encountered in the railway literature, but not applied in Sweden, concern slack, travel time variation and wagons. Drawing analogies from the electric power industry turns out to be rewarding, since stakeholder structure and production characteristics are similar to those of the railway system. New railway indicators, found by analogy to the power industry, include indicators that concern passengers, traffic work not delivered and how to subtract the effect of adverse weather.

It is suggested that availability indicators be categorised with respect to

- stakeholder concerned
- whether external disturbances are included
- how they are calculated.

By the use of availability indicators, the performance of different train operating companies and maintenance contractors on different lines, under dissimilar conditions, might be compared. Prospective timetables and railway networks might also be compared. The stakeholders of the railway need to be aware that different availability indicators might give different results regarding, for example, the ranking of failure causes.

Train traffic controllers were asked to judge the relevance to their job of about 40 indicators. Among the indicators judged relevant were ‘Sum of arrival delay minutes’ and ‘Fraction of not accurately registered train data’, that is, that the TOC has not registered, for example, the correct train weight. Among the indicators judged irrelevant or non-understandable is ‘Fraction of passengers that are delayed more than a certain number of times per week or year’.

4.5 Description of Paper V

Nyström, B. (2008). A Methodology for Measuring the Quality of Deviation Reporting – Applied to Railway Delay Attribution. Submitted for publication.

4.5.1 Purpose of Paper

The purpose of the paper is to describe a methodology and its application to measure the consistency of railway delay attribution, that is, to which extent similar delays are similarly reported. This is done in order to inform on the experiences of using the methodology in this context.

4.5.2 Study Approach

A survey is developed to ask the personnel performing delay attribution how they would report the delays described in the vignettes of the survey. The survey development, application and analysis are thoroughly described in the paper. Paper V delivers the methodology which is applied in Paper VI.

4.5.3 Findings

In order to cover a large share of the many kinds of delays that exist, the survey consists of four, partly different, questionnaire versions, with 51 different vignettes altogether. The questionnaire versions were randomly assigned to respondents. The vignettes together span several dimensions (underlying variables), for example ‘Root cause vs symptom’, ‘Maintenance’, ‘Allocation on several causes’ and ‘Passengers’. Furthermore, to have different versions makes it possible to check for impact from other vignettes, as well as from slightly dissimilar wordings of a vignette. The methodology proved useful for measuring the consistency of delay attribution. However, to ask for the importance of each question, with the aim of finding the ones considered most important, did not work out well, as the questions were ticked to be about as important. A possible way forward is to instead, near the end of the questionnaire, ask the respondent to enumerate the most important vignettes, for example, to give a top-five. Furthermore, a procedure given in the paper, together with results of Paper VI, might be applied to data from the delay database in order to improve the accuracy of the data.

A survey similar to the one presented in this paper can be used by railways to estimate the accuracy of their delay attribution systems, as well as to continuously improve them. For instance, it can be used in the vocational training of delay attributors (for example train traffic controllers). By use of the survey, drawbacks of

the current delay attribution system can be identified. Continuous use of the questionnaire makes it possible to follow changes over time of the delay attribution and identify their possible causes, for example training. The methodology may be applied for measuring the consistency of deviation attribution in general.

4.6 Description of Paper VI

Nyström, B. (2007). Consistency of Railroad Delay Attribution in Sweden – Measurement and Analysis. Submitted for publication.

4.6.1 Purpose of Paper

The purpose of the paper is to describe the measurements and analyses of the consistency of the delay attribution, that is, to which extent similar delays are similarly reported. It also describes the attitudes towards delay attribution among delay attributors. In the case of low consistency, it describes possible improvements.

4.6.2 Study Approach

The methodology presented in Paper V is applied. Thus, a questionnaire is distributed to all delay attributors in the Swedish railway. In the questionnaire, vignettes tell stories about train delays. The respondents are asked to write how they would report the events, described in the vignettes, into the TFÖR delay database.

4.6.3 Findings

There are quite large differences in how similar delays are reported in different parts of the organisation; that is, the consistency of delay attribution is low. For most vignettes in the questionnaire, the respondents do not agree on the delay attribution. The highest number of different delay attribution codes, given as answers to one question, was 17. The shown inconsistencies might give consequences such as rolling stock being ascribed a lower share of the delays compared to infrastructure, information to identify causes being omitted, or they may have no consequences (as two different codes sometimes refer to the same thing).

An example of a vignette which shows high inconsistency describes a train that arrives late at a planned track work. Another vignette concerns replacement buses, where the delay attributors disagree whether to report the delayed train (56% of the answers), the delayed passengers³⁶ (16%) or to not follow any of these principles. Regarding the vignettes for which the respondents were asked to also state the length of the delay, 10% gave the absolute delay, not the extra delay, which would have been correct. One of the code pairs that most often were given as answers to the same questions is 'F14 Pantograph' & 'I40 Catenary Fault', which means that the shares of catenary wire teardowns that are attributed to pantograph and catenary fault, respectively, are not entirely trustworthy. In a vignette that describes that train 03's pantograph gets caught in the catenary wire, a lot of catenary wire is torn down and

³⁶ That is, the bus.

the track remains closed for eight hours. When the catenary has been repaired, the train is assigned a new train number, 13, and continues its journey. After this, the train suffers no more delays. The most popular answers were 'I40 Catenary fault' (42%) and 'F14 Pantograph' (20%). Remarkably, 2% of the respondents chose not to report. They justified it by the fact that the old train number (03) did not receive a delay, as the train was assigned a new train number (13). As the train got a new, designated, train path, the new train number might not be considered as delayed. The survey indicates that the delay attributors use heuristics (rules of thumb) that are specific to the particular kind of situation, rather than general rules.

Of the respondents, 56% judged delay attribution to be 'Important' or 'Very important'. Hence, most delay attributors are probably motivated to improve the quality of the delay attribution, although they complain about the software. Also, data from the delay database might be refined by applying a procedure, presented in Paper V, to results of Paper VI. Changes to computer software, instruction documents and training will improve the delay attribution system. Regular and individual feedback should be given to the delay attributors.

4.7 Description of Paper VII

Niska, S. & Nyström, B. (2008). Investigating Recurring Faults – a Case Study of Oxmyran Signal Box. Submitted for publication.

4.7.1 Purpose of Paper

The purpose of the paper is to describe ways to identify and remedy causes for frequent failures of a component in the Swedish railway infrastructure.

4.7.2 Study Approach

A specific signal box that has been problematic, without any clear cause identification, is selected. By using databases on maintenance and traffic, as well as interviews, observations and measurements on site, a multitude of possible failure mechanisms are investigated.

4.7.3 Findings

The failure reporting system is not so useful in identifying the most frequent failures in this case, as it, to some extent, conceals that the component recurrently fails in the same failure mode. This phenomenon makes it more difficult to learn how to improve design and maintenance. The failures are suspected to be caused by electromagnetic interference. The harsh environment close to the track and contact wire makes electromagnetic interference an issue. The investigated fault is hypothesised to be related to signal box construction, installation, lightning and/or the train traffic. In general, Banverket do not inspect components on delivery, which usually makes it harder to exclude non-compliant components as a cause of failure. The voltage over the component under scrutiny shows fluctuations considerably larger than at a nearby

signal box, despite similar specifications. An exchange of rectifiers is therefore recommended as the next step in identifying the causes of the failures.

Maintenance personnel were accustomed to the high frequency of failures. The study illustrates that even seemingly simple faults are complex to analyse. Other kinds of failures, which are especially frequent at specific places, might be investigated in a related manner. There is a need to develop ways to preserve and integrate data from different sources, as illustrated in the paper.

4.8 Description of Paper VIII

Nyström, B. & Söderholm, P. (2008). Selection of maintenance actions using the Analytic Hierarchy Process (AHP): decision-making in railway infrastructure. Accepted for publication in *Structure and Infrastructure Engineering*.

4.8.1 Purpose of Paper

The purpose of the paper is to describe the relative importance of different criteria, according to decision-makers. The studied criteria are differently influenced by different maintenance actions on the railway infrastructure. The second part of the purpose is to describe how consistent the selection of maintenance actions is.

4.8.2 Study Approach

Criteria that describe different effects of maintenance actions are developed by means of interview and document studies. The criteria are presented to track managers, together with a set of maintenance actions, specific for each of these decision-makers. In these experiments, the Analytic Hierarchy Process (AHP) is used to obtain the preferences for the criteria and for the different maintenance actions from these decision-makers.

4.8.3 Findings

A group interview showed that non-documented actions are a problem. The kinds of decisions to be made vary among track managers and over time, with respect to urgency, trade and effect. None of the informants knew about any evaluation of decisions, that is, planned effects are compared to attained effects. The decision-makers roughly agree on the ranking of the criteria, that is, the relative importance of different effects of maintenance. The criterion 'Safety' is ranked first (weight 0.41), then comes 'Punctuality and availability' (0.15), 'Track work time' (0.09), 'Cost' (0.08), 'Condition' (0.08), 'Own abilities and development' (0.06), 'Collaboration with stakeholders' (0.06) and 'Environmental impact' (0.06). However, the discrepancies between the results of the two ways employed to elicit the preferences for the maintenance actions are rather large. The decision-makers consider it easy to understand the rationale of AHP and to enter their preferences into the computer.

The criterion 'Cost', as ranked by the decision-makers, was partly compared to economical records. Similar comparisons might also be carried out for other criteria,

and this would make it possible to learn how big the inaccuracy is allowed to be, in order to not affect decisions. This would give indications of how much effort should be put into collecting data on each criterion. Different ways of presenting the decision problem can give different results. It is suggested that preferences of maintenance actions are recorded, as they are in this paper, in order to document the rationale of the decisions made and to facilitate learning.

5 Discussion and Conclusions

In this chapter, the analysis of the findings is presented. The description of the analysis is divided according to the research questions, and includes a discussion on the validity (reliability issues are mainly discussed in Section 3). The research questions all help in answering the main research question: How can punctuality of railways be improved?

5.1 Discussion and Conclusions Related to Research Question A

Research Question A. *How should (un)punctuality be described?*

This text draws on Section 2, as well as Nyström (2005), Paper I, Paper II and, to some extent, Paper VI. Here, the most important results and conclusions are summarised. The methodology applied is presented in Section 3.4.1.

5.1.1 Results in this Investigation

It is common for the studied railways of different countries to consider a train departing or arriving within 5 minutes from timetable as ‘punctual’. Also, for instance, 1, 3 and 15 minutes are used as tolerances. Railways often measure ‘punctuality’ as the percentage of punctual trains, the sum of delay minutes or the average delay. The fact that a multitude of partly synonymous terms related to punctuality are in use, adds to the heterogeneity.

If one not only considers each train as punctual or not, but also considers how large the unpunctuality is, it is easier to understand the consequences of unpunctuality. This means that it is easier to construct incentives that are better aligned with the cost of the occurred unpunctuality (see Paper II, which analyses a transport chain). Railway stakeholders emphasise punctuality as being relative to an agreement (explicit or implicit), not necessarily relative to a timetable. Therefore, clear definitions of terms should accompany punctuality statistics.

Railways usually do not include cancelled trains in the term punctuality. To do so in a single figure would value cancelled relative to delayed trains. Such a valuation is straightforward in some instances, for instance when commuters are assumed to catch the next train, one may consider them to be delayed the length of the interval between trains. In any case, cancellations should be included in some way when presenting punctuality figures, as both delays and cancellations reflect unreliability and result in consequences for passengers and freight customers.

The times of train departure and arrival should be determined when the train is by the platform. This is the preference of TOCs. However, the train positioning system in some cases measures departure and arrival at another position³⁷. Therefore, this data might be complemented with measurements from the train or driver's report, all of which may be stored in a delay attribution database.

5.1.2 Conclusions and Application

A large variation in train running time necessitates a TOC to have more rolling stock than otherwise needed. Therefore, to some TOCs, it is more important to uphold the train circulation, than to attain high punctuality (as illustrated in Paper I). 'Time lost', defined by Mattsson (2004) as the difference between 'Actual running time' and 'Basic running time', is a measure of the time lost for the customers, as well as of the extra time that rolling stock is needed. Hence, it is a valuable measure that should be used complementary to punctuality.

The choice of system borders is important when describing and measuring punctuality. This is because the passenger (goods), the part of the journey by train, or by one of the trains in the journey, might alternatively constitute borders. An example, showing the importance of clear specifications of which borders to use, is given by the case of replacement buses, where delay attributors disagree whether to report the delayed train, the delayed passengers, or not to follow any of these principles. Whether and how different kinds of passengers and goods, step effects³⁸ and passenger behaviour, for example, concerning change of trains, are modelled, should therefore be clearly stated when describing and measuring punctuality.

The perceived duration of a journey varies, for example with respect to the relative duration of the parts of the journey, time of day as well as from person to person. These observations advocate that punctuality should be described on the basis of measurements, not just customer polls. Because of different system borders, different stakeholders need different measures of punctuality.

The amount of unpunctuality should be put in relation to the traffic volume, in order to put the unpunctuality cost in relation to the benefit of the performed traffic. Valuation of social economy impact, in essence to calculate unpunctuality cost, is

³⁷ The train passage time is measured using the section blocks (explained in Appendix A: Basic Railway Technology and Paper I). In order to account for the train running distance between platform and section block, the recorded passage times are automatically offset before they are stored in the TFÖR database. The size of the offset is identical to each train. This means that differences in acceleration & deceleration properties between trains (and driving styles) cause the difference between TFÖR data and actual departure and arrival times, to vary.

³⁸ *Step effect* is the phenomenon that the valuation (that is, cost) of a delay is not a continuous function of delay length, but increases momentarily when reaching a certain length of delay. Examples include freight trains being delayed for ships with fixed departure time and passengers who miss their connections.

hard to perform due to the variety of passengers and goods (Ackermann, 1998; SIKA, 2002).

When constructing timetables, one needs to know cost of delay time relative to cost of timetabled running time in order to minimise the unpunctuality cost. However, unpunctuality has to be given financial figures in order to be comparable to the possible costs of improving punctuality by higher availability of railway items. Higher availability might be achieved by higher reliability and/or shorter repair time. In this connection, one should be aware that describing unpunctuality as a sum of delay minutes implies that the unpunctuality cost is assumed to be linear to delay time³⁹. One should also try to calculate the unpunctuality cost, that is, both external and internal unpunctuality cost, not just external unpunctuality cost.

5.1.3 Validity

A few larger railways have been studied. Some other railways might employ a different set of concepts. Interviews also admit the possibility that the interviewer “hears what he wants to hear”, and imposes his opinion on the interviewee. The probability of this has been reduced, and the validity hence increased, by, for example, recording interviews.

5.2 Discussion and Conclusions Related to Research Question B

Research Question B. *Which problems are encountered when trying to reduce unpunctuality?*

The studies that address this question are presented in more detail in Paper III and Paper VII. Here, the most important results and conclusions are summarised. The methodology applied is presented in Section 3.4.2.

5.2.1 Results in this Investigation

One problem is to identify the most important causes. Which causes of unpunctuality are the worst, and therefore the most important to eliminate or reduce, depend on how the valuation is made. As an example, among infrastructure-related failures, contact wire failures are relatively few in number, yet inflict a significant amount of delay-hours, as each failure results in long delays. The ranking of delay causes imposed by the number of delay-hours is quite different from the one based on the number of occurred failures, where turnout is the most frequent. A further complication is presented by secondary delays that have not been attributed to primary delays, as well as delays for which no cause has been reported. The relative magnitudes of the causes in the statistics on delays may be overturned by such delays.

³⁹ An economist, such as Bruzelius (1979), might have formulated this assumption as that the marginal value of time is constant.

The external unpunctuality cost, which in the study presented in Paper III is the cost incurred on the freight customer by the consequences of unpunctuality of a steel slab transport, is difficult to estimate with good accuracy. This is because the changes to the production programme in the factory that receives the slabs, due to unpunctual trains, have not been documented. It is also due to the fact that margins in the production programme and buffers of steel slabs to a certain extent delimit the effect of a single train running late. Therefore, it is not possible to calculate the cost caused by a certain train's arrival unpunctuality. The variation in transport time might be used to identify distances where the train is belated. Causes of this include low contact wire voltage, sloppy rails and train meetings. Steep slopes combined with narrow curves, on parts of the route, are circumstances obscuring the variation due to the causes mentioned. For the factory receiving the slabs, it is important to get information in advance about deviations, such as train delays and that certain slabs are missing. The current database might be slightly modified to make identification of groups of trains easier.

Unpunctuality costs, as well as costs for corrective maintenance, are caused by failures of the signal box at Oxmyran station (studied in Paper VII). There, a certain component fails, and therefore is replaced, considerably more often than identical components in other signal boxes. The voltage over the component under scrutiny shows fluctuations considerably larger than the corresponding component in a nearby signal box with similar specifications. However, positive fault localisation has not yet been achieved. Reasons for this include inaccurate information in the failure reporting system, that data from numerous sources have to be integrated and that the electromagnetic environment varies.

5.2.2 Conclusions and Application

What was found regarding unpunctuality of the steel slab transport is in agreement with the results by Östlund et al. (2001); it is difficult to estimate the cost of a delay of a certain length for a freight train, due to step effects. Also, Kreitz (2002) observed that it is hard to value freight train punctuality, as overall changes in the punctuality level might cause production systems to be altered, thus changing their susceptibility to train unpunctuality. In this investigation, the absence of detailed production data contributes to the difficulty. More detailed production data or a different approach with assessing overall effects of train unpunctuality might be two ways forward.

Even simple recurrent failures, such as that of the component in the Oxmyran signal box, are complex to analyse, due to the fact that information has to be integrated, and lack of suitable data and documentation. Studying the differences in the electromagnetic environments of similar components is a possible way forward.

Problems encountered when trying to reduce unpunctuality are also found in the Discussion and Conclusions sections regarding other research questions: see Sections

5.3, 5.4 & 5.5. For example, in Section 5.5.2, data on costs, relevant to research question E, is discussed.

5.2.3 Validity

The problem of reducing unpunctuality has been explored from the perspective of a train as well as the perspective of the infrastructure. However, it is not certain that the encountered problems are representative of unpunctuality-causing problems in general. However, by choosing these two perspectives, rolling stock and infrastructure, only one major perspective, that of train traffic control, is omitted. Nevertheless, this perspective is glimpsed in both studies.

5.3 Discussion and Conclusions Related to Research Question C

Research Question C. *How might availability of the railway infrastructure be described?*

The studies that address this question are presented in more detail in Paper IV (focused on potential availability indicators) and to some extent in Paper II (focused on the chain of requirements from the punctuality perceived by the customer to infrastructure availability in a real case). Here, the most important results and conclusions are summarised. The methodology applied is presented in Section 3.4.3.

5.3.1 Results in this Investigation

Availability indicators encountered in the railway literature, but not applied in Sweden, concern travel time variation, slack and wagons. New railway indicators found by analogy to the power industry include indicators that concern passengers (to a large extent), traffic work not delivered and how to subtract the effect of adverse weather. The power industry is rewarding to draw analogies from, since its stakeholder structure and production characteristics are similar to those of the railway system. The availability indicators found are most often calculated based on either punctuality or capacity.

Availability indicators might be categorised with respect to the type of concerned stakeholder (for instance, TOC), inclusion or not of external disturbances (for instance, effects of heavy snowfall) and how they are calculated. The fact that transports of unpunctual trains have a value, but not as high as punctual trains, advocates that availability should be calculated as a function of punctuality (not, for example, as a function of the planned capacity of the railway line).

5.3.2 Conclusions and Application

Availability indicators might be used to describe and compare different railways, different parts of one railway, or different ways of carrying out maintenance actions. This might facilitate benchmarking as well as business agreements, for instance,

between IM⁴⁰ and MC⁴¹. Potential users, for example, track managers, today often do not calculate indicators, because it is held to be too cumbersome or data is lacking. So, a choice of suitable indicators must be followed up to ensure that their use is feasible.

In Sweden, track works are planned according to an eight-week rolling plan⁴². Corrective maintenance actions also require time on track. These times are most often found in the respective databases⁴³. However, works that are carried out when there is no train traffic are not recorded in any database. Furthermore, inspection remarks designated as weekly are often not remedied within the specified time limit – it might well take a month. So, when calculating availability indicators, these complications need to be taken into account.

5.3.3 Validity

The literature study on availability indicators covered railway scientific literature and a few other branches of industry. Therefore, there might be other industries that are more suitable for purposes of comparison than the one identified; the electric power industry. The questionnaire used for asking personnel which indicators they judged most relevant briefly described the indicators. The answers might have been different if the indicators had been described differently, for example by presenting graphs of realistic data.

5.4 Discussion and Conclusions Related to Research Question D

Research Question D. *What is the suitability of the information, from systems measuring punctuality and maintenance, for improvement work?*

The studies that address this question are presented in more detail in Paper I and in Paper V & Paper VI. Here, the most important results and conclusions are summarised. The methodology applied is presented in Section 3.4.4.

5.4.1 Results in this Investigation

In general, data in the databases of Banverket, to a large extent lack descriptions of variables, their measurement and verification. This makes data aggregation and integration, which often are necessary in punctuality improvement work, more difficult. Reporting of railway infrastructure failures is managed using the Ofelia system, inspection remarks are in Bessy and no central database is used for condition-based maintenance measurements. This makes it more difficult to know of the condition of the railway items. Also, data on item replacements is hard or impossible to find. Ofelia to some extent interfaces with the TFÖR delay attribution system.

⁴⁰ Infrastructure manager.

⁴¹ Maintenance contractor.

⁴² This plan is called 'BUP'.

⁴³ Scheduled times of planned track works are found in a dedicated database, the times of the corrective maintenance actions in another (Ofelia).

Ofelia requires very detailed data regarding failing items, but not on the events leading to failure. By a broader description of the failures, including, for example, various symptoms and repair difficulties, fault localisation as well as learning could be facilitated, for instance when new technology is introduced in the infrastructure. To increase in-field learning, information on previous failures and actions taken should be readily available to personnel heading for maintenance work. Some failure records indicate causes that are likely not to be accurate. Causes given merely to fill in blank data fields may mislead, so the need to fill in all fields should not be over-emphasised. For further details, see Paper I.

Regarding the delay attribution, that is to report delays into the TFÖR database, there are quite large differences in how similar delays are attributed. This is true when comparing parts of the organisation as well as individuals at the same workplace. As an illustration, the highest number of different delay attribution codes given as answers to one question was 17. Furthermore, 10% of the delay attributors report the amount of absolute delay when they are asked to specify length of delays, despite that the amount of extra delay is to be reported. Training, regular feedback to the delay attributors and changes to computer programs will improve delay attribution consistency. In the future, a survey similar to the one used in the study might be used periodically to estimate the consistency and accuracy of the delay attribution. When considered relevant, such surveys might be carried out just for certain topical dimensions, such as ‘Root cause vs symptom’ or ‘Passengers’. Banverket’s goal that a certain percentage of the number of delays should have a cause attributed, rather than a percentage of the total length of the delays, might reduce focus on the attribution of long delays. The longer the delay, the more important it should be to find its cause/s. For further details, see Paper V & Paper VI.

5.4.2 Conclusions and Applications

As Banverket’s Ofelia, most CMMS (Computerised Maintenance Management Systems) on the Swedish market do not support condition-based maintenance (Kans, 2006). Also Granström (2008) points out inaccuracies and non-existence of follow-up of the maintenance of railway infrastructure items. He gives as an example the STRIX measurement wagon. The STRIX wagon is scheduled to be maintained and upgraded in midwinter, a period in which most problems in the interface contact wire – pantograph appear. Incorrect position of track and/or contact wire, too high or too low wire tension or too rapid change of contact wire height, are all failure modes, which might be detected using STRIX. So, measuring at the right time is important. Further, the unfeasibility of data for maintenance is exemplified by Gustavsson & Ådegren (2001), who report that a certain infrastructure work is not possible to associate with certainty with a cost in the Agresso economy database of Banverket, as the works are there specified according to the account date, not the date the work was carried out, and according to track section, without further detailing, for instance, between which stations along the track section.

A drawback in the Ofelia database is the lack of historical information on maintenance actions, as well as investments, carried out. The failure reporting is mostly designed for reporting faulty components. The events leading to failure are not as well described. Data integration with other systems would make it possible to detect input slips, as well as help in discovering causes of failure.

No studies on railway delay attribution comparing the answers from different delay attributors subjected to identical delay descriptions have been found. In the computer industry, Henningsson & Wohlin (2004) compare inter-rater agreement of software faults classification. They ask computer specialists to classify software faults, each described by a brief text. This was done using the Orthogonal Defect Classification (ODC) scheme, one of several existing classifications. They found that the inter-rater agreement was low. They consider the classification scheme not to be an important explanation for the low agreement, as other researchers have reported good agreements using this classification. Instead, they point out the description of faults to be an important factor. They conclude that in order to perform an accurate classification of faults, the classifier needs deep knowledge of the system, should have the source code, know what was done in order to remedy the fault and talk to the developer of the software.

The survey applied to investigate the consistency of the delay attribution, is not entirely comparable to the software faults study by Henningsson & Wohlin (2004). Firstly, the respondents of the survey have a good knowledge of their railway system, but the computer specialists in the software faults study did not have first-hand knowledge of the software in which the software faults occurred. Secondly, no studies on the inter-rater agreement of delay attribution have been found. Therefore, one should not exclude the TFÖR classification scheme in itself as an explanation of the differences between delay attributors. However, one similarity is that both the delay attribution survey and Henningsson & Wohlin (2004) suffer from not allowing subjects to ask for more information, for instance from personnel remedying the fault.

5.4.3 Validity

As Ofelia records were compared to interview data, obtained soon after a failure, the remedying personnel's memories should be reasonably accurate. However, as an interviewee might have performed work between the repair of the fault the interview is about and the interview, there is a possibility that the interviewee confused the maintenance works. That only one remedying person was interviewed, concerning each failure, lowers validity. Also, only failures reported during a brief period were investigated in this study, which means that they might not be representative of how failures are reported in general.

Regarding the questionnaire on delay attribution, the response rate (50%) makes validity an issue, but the small differences between the questionnaires received at different times indicate that validity is not affected negatively. However, at one CTTC, the response rate was so low that it is difficult to draw conclusions regarding

it. By using situations from reality, together with existing delay attribution codes, to create the vignettes, validity was increased. It was also increased by testing and modifying the vignettes in interviews with delay attributors. The external validity of the study was increased by presenting the delay attributor to realistic situations. The external validity was lowered by the fact that the events leading to delay were described in text and all facts (considered relevant by the researcher) were thus there, not taking into consideration that information (not) acquired by the delay attributor might affect the delay attribution. The validity was increased by asking the delay attributor to write down the answer, instead of giving alternatives to tick. This probably resulted in fewer casual answers and forced the delay attributor to think more as if it were a real situation. To have several versions of the questionnaire improved validity. By having an individual questionnaire, the validity was possibly slightly decreased. On the other hand, reliability was probably improved hereby. Content validity, in essence, that the questions are representative for actually occurred delays, is hard to estimate accurately, as it requires knowledge of the delays that actually occur, a quest in which this study seeks to contribute.

5.5 Discussion and Conclusions Related to Research Question E

Research Question E. *How is the choice made between different maintenance actions?*

The study that addresses this question is presented in more detail in Paper VIII. Here, the most important results and conclusions are summarised. The methodology applied is presented in Section 3.4.5.

5.5.1 Results in this Investigation

How the choices between different maintenance actions are made is due to the decision-making procedure used. When describing the decision-making procedure, Mora et al. (2006) identify five steps:

- Intelligence (Set the decisional agenda)
- Design (Represent the problem in a suitable model)
- Choice (Evaluate and select alternatives)
- Implementation (of the selected course of action)
- Learning (based on the decisions taken).

Mora et al. (2006) consider it to be a lack of research in the Design, Implementation and Learning steps. Although the study that answers Research Question E focused on the Choice step, it shows what the Design may look like, and sketches why this is beneficial for Learning in the long run.

The Design of the decision problem is in this study based on the assumption that decision-makers use several criteria to make choices between fixed alternatives. The

criteria constructed for the effects of maintenance are: 'Safety', 'Punctuality and availability', 'Track work time', 'Cost', 'Condition', 'Own abilities and development', 'Collaboration with stakeholders' and 'Environmental impact'. The most important criterion according to the track managers is 'Safety', which has 41% of the priority. The track managers who participated in the study showed differences, although not large, in their ranking of the criteria. The notable differences are that one track manager placed 'Safety' in second place, as he considered 'Safety' to be a hygiene criterion, and that another track manager ranked 'Punctuality and availability' fifth, while the others placed it in top three.

The correlation between two different rankings of the maintenance alternatives, obtained by the two ways of recording priorities, varies among the track managers. One of the track managers in the study shows a negative correlation, that is, definitely did not make decisions that account for all criteria in a consequent manner.

The work of the track manager includes more than purely making decisions. By discussions, both informal chats with colleagues and formal communication, for example, by proposals for tender, the involved personnel are prepared for the decisions, maintenance actions and their consequences.

5.5.2 Conclusions and Application

No studies comparing two ways of recording priorities; ranking by criteria and ranking by alternatives, have been found within maintenance, but comparisons between the criteria weights obtained by direct assignment and by AHP are prevalent and differing. For example, Hagquist (1994) reports large differences between the weights obtained thereby, while Ramadhan et al. (1999) report only minor differences.

The study, performed here, compares two methodologies of making decisions that should give the same ranking of the maintenance alternatives. The similarity between the two rankings indicates to which extent a track manager prioritises in a consistent way and therefore reveals something about his/her mental model. A low correlation between the rankings might be explained by the fact that the decision-maker may consider other aspects in the prioritisation than the presented criteria, or has a bad grasp of the pros and/or cons of some maintenance actions, or changes his/her mind during the experiment.

The mode used for making a choice between alternatives, whether by selection or bidding, might influence the choice, according to Tversky et al. (1990). As only one mode (selection) was applied in this study, it is not possible to say whether the choices would have been identical if decision-makers were to bid on the maintenance actions. Furthermore, bidding would not have been feasible in the experiment, as the maintenance actions do not have equal value. Also, in the experiments, the subjects did tell how much more important each alternative is to every other, a kind of bid.

To get approximate, but perhaps sufficiently accurate, data on the criterion 'Cost', that is, expenditures, there is a possibility to use cost data collected nation-wide. However, a first attempt to do so shows that the costs for specified maintenance actions, supplied by regional bodies in Banverket, differ up to a factor of eight (Banverket, 2003). This is probably due, not only to real differences in costs, but also to differences in what is included in a certain kind of maintenance action, for instance a turnout revision. Other criteria also suffer from difficulties, for example 'Environmental impact', as Banverket has no documented methodology for identifying impact on the environment (Lundberg, 2005).

Butler et al. (1993) ask managers to assess different investments, based on degree of corporate fit and several financial indicators such as length of payback period. The managers show reasonably good evaluation congruence within each company. Compared to the findings of Butler et al. (1993), the ranking of criteria are rather similar among the track managers, although maintenance actions are not so easily described, as are financial indicators.

Track managers lack a user-friendly overall picture of data, a difficulty observed by the author (in Section 3.4.5) as well as by Jovanovic (2004). The criteria presented in the study might be a step on the way to helping track managers organise a data framework and obtaining the data they need.

5.5.3 Validity

The external validity of the study is increased by presenting the decision-maker to real-life maintenance actions that are of contemporary interest to him/her. The alternative, to present identical maintenance actions to each decision-maker, would allow for comparisons between the decision-makers. However, this would require detailing a railway environment, which is cumbersome and takes a lot of time to make credible.

The AHP is a methodology with high reliability in itself, as many redundant questions are posed and the inconsistency measure tells to what degree the subject contradicts himself/herself (Saaty, 1980). This advocates that the results on an individual level are trustworthy. However, the low number of subjects in the study, six, makes it more difficult to judge whether or not the subject with a negative correlation coefficient is a rare exception.

5.6 Some Aspects of Improving Punctuality

This thesis focuses on punctuality and improvement of punctuality within the railway system. The emphasis is mainly on infrastructure maintenance activities. Different aspects of how punctuality might be improved have been described and analysed in Sections 5.1-5.5.

In this final section of Chapter 5, the main issues discussed in the thesis and the main results obtained are related to the Plan-Do-Study-Act-cycle (PDSA-cycle), which is an important mental model in quality management introduced in Japan by Deming in the 1950s; see Deming (1986) or Bergman & Klefsjö (2003).

The requirements set by stakeholders (Paper II) should give rise to steps taken to achieve the requirements. Also, methodologies that are to be used to measure their fulfilment should be developed and routines documented (Papers V & VIII). This includes equipment for measuring, databases to store data in and definitions of indicators to be used. Furthermore, strategies for preventive and corrective maintenance activities are decided upon (in the Plan-phase of the PDSA-cycle, see the circle in Figure 13).

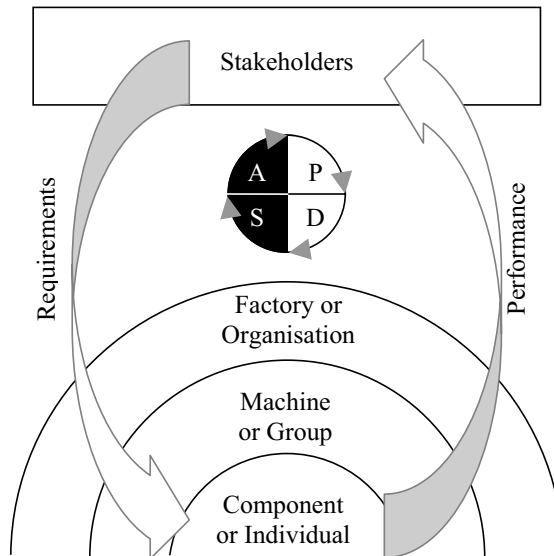


Figure 13. The thesis summarised. Requirements and performance measurements might be considered on different levels of the reality, both regarding items and humans. The requirements and performance considered in this thesis are related to punctuality and maintenance. Positively, knowledge about performance measurement also helps in stating requirements and facilitates decision-making by the stakeholders. This helps in the Study and Act phases of the PDSA-cycle.

Then, punctuality is measured and maintenance according to the decided strategies is performed. Furthermore, reasons for the deviations are investigated and relevant data stored in the databases (this is the Do-phase of the PDSA-cycle).

In the Study-phase the data is used for analysis. This means to select which of existing data to analyse and to carry out the analysis itself.

In the Act-phase lessons should be learnt based on the analysis in the Study-phase and steps should be taken for improvements. Here, the railway stakeholders should also learn from their own way of performing improvements in order to develop the way improvements are made in the next round of the PDSA-cycle.

This thesis has a focus on the Study-phase and the Act-phase. It has emphasised investigations and discussions related to a necessary conceptual framework, to the data available at present, and how the data is used for maintenance activities to improve punctuality (Papers I, III & VII). The steps of the maintenance process (Section 2.4.2) that mainly have been studied in this thesis are maintenance assessment and maintenance improvement, corresponding to the Study-phase and the Act-phase of the improvement cycle.

All this work starts with the very concepts of ‘punctual’ and ‘punctuality’. It has been seen that concepts related to punctuality are differently defined (Paper II) and interpreted (Paper VI). This makes it hard to check for the fulfilment of punctuality requirements and also makes comparisons between different railways more difficult. Furthermore, the improvement steps taken are based on maintenance activities, which in turn are based on decision criteria. The formulations of these criteria and the verification of their fulfilments are based on information from different databases (Appendix B). One problem is that the decision criteria used are not documented today, and thus probably not uniform (Paper VIII). It has also been seen that the information available in the present databases is not sufficiently well aligned to the need of the maintenance process.

Reality might be measured on different levels. Measurements might be performed at a much aggregated level down to a very detailed level, both regarding items and humans. Furthermore, punctuality might be measured as well as different availability concepts (Paper IV). Thus, Figure 13 illustrates that “faults are not found but sought after” (Hollnagel, 2004), that is, one has to consider where one chooses to stop the investigation, which will determine the assigned root cause. The cost of measuring and the benefit from measuring, should decide what is measured – but most of all the purpose. Regarding effects of maintenance actions, measuring is in some cases not even performed, for example, after tamping.

The development towards full competition in the Swedish railway sector is still ongoing. This includes finding forms for interaction between stakeholders, for example in improvement groups, as well as formulating contracts that give proper incitements for improvement, regarding infrastructure maintenance, the impact of trains on the railway infrastructure and vice versa, as well as publicly subsidised passenger traffic. However, compared to several other sectors of the industry, the culture still seems to be too weak for improvements. This might be due to the fact that the railway culture is strongly based on rules. However, the aviation sector is also strongly based on rules, but here we find a culture with more emphasis on

continuous improvement. Thus, there should be scope for benchmarking and improvements.

The thesis shows that there are deficiencies in performance measurement, data storage and use of data, and hence on which grounds decisions are made. It also presents some ways to improve punctuality, including performance measurement and decision-making. Although the empirical parts of this research mostly cover the Swedish railway and its context of deregulation, the literature study and some of the empirical work indicate that the situation is similar for other railways and that the results are applicable to them as well.

6 Suggestions for Improvements

In this chapter, some hands-on advice, based on the results and experience gained during this investigation, are given.

The suggestions for improvements are grouped under three headings, although this does not mean that the three groups are completely disjointed. A few examples are given under each heading. Some of them are easy to implement, others might be part of a long-term effort. The three headings are:

- Performance measurement, which helps to set goals and learn from better performers.
- Data management, which mainly concerns databases.
- Decisions, which is an area that borders with data management.

6.1 Performance Measurement

The world inside any organisation might seem static and improvements impossible, fostering a static attitude: “If people want to kill themselves by jumping in front of a train, they will do it”. But suicide rates, from jumping in front of trains or otherwise, differ considerably between countries and over time. Performance measurement thereby helps to identify possible areas of improvement.

Measure losses, for example slack. Slack is an example of a loss indicator, which, unlike delay minutes, cannot be easily administered away. This makes it a more appropriate goal.

6.2 Data Management

The number of databases is vast. Instead of having database managers, have information managers in charge of the data being collected and used in a good way, including database discontinuation and data migration to new databases. Involve users from all stakeholders from the beginning and test the user interfaces on novices at their workplace.

Continuously adapt software tools to the changing business environment, in order to facilitate, for example, follow-up of IM-MC contracts. Put more automation in the analysis of reported failures, for instance by the use of techniques that classify free text by content and search for patterns. This might be used for analysis as well as to facilitate input. Automatic alerts, like highlighting implausibly long repair times, might be easily implemented.

6.3 Decisions

Describe the pros and cons of proposed improvements in terms of the maintenance action criteria described in this thesis. Also, it should be the responsibility of a proponent of an action to define how the effects of the proposal are to be quantitatively measured in practice.

Informing others about what you are doing, is just as important as the activity itself. For example, make documents from local improvement groups easily retrievable. This includes data on performance of, for instance, TOCs.

Enforce made decisions. For example, do not allow trains that have not entered the appropriate data into the Opera train registering system to depart. Make certain that a MC has reported changed asset data into the BIS asset register, before paying for its work, whether maintenance or investment.

7 Suggestions for Further Research

In this chapter, some suggestions of possible further research are explored and discussed.

The suggestions for further research are grouped under three headings, although this does not mean that the three groups are completely disjointed. A few examples are given under each heading. The three headings are:

- Definitions that help communication in punctuality improvement work.
- Data collection, for punctuality and maintenance, involves large amounts of data which has to be carefully defined and collected to be useful.
- Culture and organisation offer many areas for improvement.

7.1 Definitions

The different descriptions of unpunctuality which are used today, which include length and frequency (that is, probability) of delay, are differently perceived. To facilitate a common understanding, descriptions that are perceived similarly by different persons, being passengers, freight customers or train traffic controllers, should be sought.

The tentative indicators of availability might be studied in an analogous manner. Also, the indicators might be computed for given, real railway lines and for different ways of carrying out maintenance actions, in order to see, for instance, the variation of different availability indicators. This might help to judge the suitability of the indicators, for example, in performance regimes.

Instead of focusing on describing one aspect of, for example, a track work, such as whether it takes longer than planned, one might define what constitutes a track work. Such a description might facilitate communication between databases and also between different stakeholders. This requires a change from the view that, for example, the Agresso system is an accounting system and nothing but an accounting system. Better data integration might be possible using an approach such as ontologies. This would require obtaining a thorough understanding of the stakeholders' perceptions of core concepts, such as track work.

7.2 Data Collection

Another important area of research is how the delay attribution should be integrated with future train traffic control centre technology in order to facilitate delay attribution. Another, minor, task might be to refine the current delay statistics, employing the procedure suggested in Paper V.

To learn more about the relation between how different criteria are judged for maintenance actions and the actual outcome of the maintenance actions requires a lot of data. In order to discover, for example, bias in judgments of decision-makers, data for many performed maintenance actions need to be collected. An important related issue is how accurate the figures of the criteria need to be, in order not to affect decisions. Yet another question is to which extent decision-makers improve from experience and how work and education should be designed in order to make them soon reach a high level of skill. To learn the effects of different (maintenance) actions, one needs to know where, when and which maintenance has been executed, data that is to some extent currently lacking. How this should be done incorporates considerations of the needs of how to describe and collect data not only on corrective and preventive maintenance, but also investments and reinvestments.

A practical issue, related to Section 7.1, concerns train traffic data. Maintenance work, both preventive and corrective, requires traffic data. Today, such data, when stored, is found in different formats at different CTTCs. An issue is how to make this data easily usable.

7.3 Culture and Organisation

Punctuality improvement groups and their ways of working might give rise to research regarding suitable ways of working. Questions include which members such groups should have, funding and also which kinds of problems are rewarding to scrutinise in this way.

A related area of research may be to find a best configuration of the decision capabilities, in essence, how decisions should be made, with respect to principles such as division according to area, type of technology or type of action. This includes how rules & regulations should be decided in a changing environment. In this connection, one might also investigate different models for how new technology may be introduced, for example, whether to introduce it where most needed, by area, by type of asset or by trade.

Another area of research is how IM-MC contracts should be designed, in order to improve punctuality. Difficulties include how to deal with randomness and deficiencies of the railway construction. A related question is how the condition of the items should be described.

Incentives for TOCs to avoid delays, thereby preventing secondary delays, even for others, have been tried in Sweden. A study of the effect of such programmes may be rewarding, including delay attribution and whether fewer persons have the attitude that delays are the sole business of the causing TOC.

References

Ackermann, T. (1998). *Die Bewertung der Pünktlichkeit als Qualitätsparameter im Schienenpersonenverkehr direkten Nutzenmessung (Valuation of Punctuality as a Quality Parameter in Passenger Rail Traffic through direct Measurement of Benefits)*. Dissertation, Forschungsarbeiten des Verkehrswissenschaftlichen Instituts an der Universität Stuttgart / Verkehrswissenschaftliches Institut, Stuttgart. ISBN 3-922203-21-3.

Al-Haimi, A.A. (1991). *Airline Schedule Punctuality Management*. Cranfield Institute of Technology.

Andersson, E. & Berg, M. (2001). *Järnvägssystem och spårfordon (Railway systems and railway vehicles)*. Järnvägsgruppen KTH, Stockholm. (In Swedish.)

Arias, E.F. (2005). The metrology of time. *Philosophical Transactions of the Royal Society A*, vol. 363, no. 1834, pp. 2289-2305.

ASQ (2005). *Quality Glossary (Alphabetical Listing)*. American Society for Quality. <http://www.asq.org/info/glossary/p.html> (07-11-03).

Bailey, C. (ed.) (1995). *European railway signalling*. Black, London. ISBN 0-7136-4167-3.

Banverket (1999). *Orsakskoder i TFÖR (Cause codes in TFÖR)*. Editors: T. Johansson/Å. Lundgren. Dated 1999-06-14. Banverket, Borlänge. (Pamphlet in Swedish.)

Banverket (2002). *Allmän järnvägsteknik (Basic railway technology)*. Banskolan, Ängelholm, Sweden. (In Swedish.)

Banverket (2003). *Slutrapport IDEAL-kalkyl (Final report IDEAL-costing)*. Banverket, Borlänge. (In Swedish.)

Banverket (2006). *Järnvägssektorns utveckling. Banverkets sektorsrapport 2005 (The development in the railway sector. Banverket's sector report 2005)*. Banverket, Borlänge. <http://banverket.se> (07-12-01).

Basu, K. & Weibull, J.W. (2002). *Punctuality – A Cultural Trait as Equilibrium*. Working paper No. 582, IUI, The Research Institute of Industrial Economics, Stockholm.

Bates, J., Polak, J., Jones, P. & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E*, vol. 37, issues 2-3, pp. 191-229.

Bente, H., Lüking, J. & Stalder, O. (2003). UIC study targets infrastructure costs. *Railway Gazette International*, vol. 159, no. 3, pp. 130-132.

Bergman, B. & Klefsjö, B. (2003). *Quality from customer needs to customer satisfaction*. Second edition. Studentlitteratur, Lund.

BEST (2002). *International Railway Benchmarking: Report of the Railway working group*. <http://www.besttransport.org/Conference05%5CBOB-rail-Hatch%20paper5b.PDF> (03-03-03)

Biederbick, C. & Suhl, L. (2007). Decision Support Tools for Customer-Oriented Dispatching. In: Geraets, F., Kroon, L., Schoebel, A., Wagner, D. & Zaroliagis, C. (eds.): *Algorithmic Methods for Railway Optimization*. Springer, Berlin/Heidelberg, pp. 171-183. ISBN 978-3-540-74245-6.

Bird, F.E. & Loftus, R.G. (1976). *Loss control management*. Institute Press, Loganville, GA.

BOB (2003). *BOB railway case - Benchmarking passenger transport in railways. Final report*. NEA Transport research and training, Rijswijk, The Netherlands.

Bruzelius, N. (1979). *The value of travel time: theory and measurement*. Croom Helm, London. ISBN 0-85664-821-3.

BS 3811:1993. *Glossary of terms used in terotechnology*. British Standards Institution, London, UK.

Butler, R., Davies, L., Pike, R. & Sharp, J. (1993). *Strategic investment decisions*. Routledge, London & New York. ISBN 0-415-07508-4.

BVH 706 (2005). *Beräkningshandledning - Hjälpmedel för samhällsekonomiska bedömningar inom järnvägssektorn (Calculation Instructions – Aids for Assessment of Social Economy in the Railway Sector)*. Banverket, Borlänge. (In Swedish.)

Carey, M. & Kwieciński, A. (1995). Properties of expected costs and performance measures in stochastic models of scheduled transport. *European Journal of Operational Research*, vol. 83, issue 1, pp. 182-199.

Carmines, E.G. & Zeller R.A (1979). *Reliability and validity assessment*. Sage, Beverly Hills. ISBN 0-8039-1371-0.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass. ISBN 0262530074.

Dahl, G. (1997). *Punktlighetsapplikationer i Essbase (Punctuality Applications in Essbase)*. SJ/Banverket/Unigrid.

Delay Attribution Board (2007). *Delay Attribution Guide*. Issue 16th September. Delay attribution board, London. http://www.rail-reg.gov.uk/upload/pdf/dag_0907.pdf (08-02-13)

Deming, W.E. (1986). *Out of the crisis: quality, productivity and competitive position*. Cambridge Univ. Press, Cambridge.

Di Marco, G., Malavasi, G. & Ricci, S. (2000). Affidabilità dei sistemi ferroviari. Analisi e valutazione mediante modelli di simulazione / Railway system reliability. Simulation models for analysis and evaluation. *Ingegneria ferroviaria*, no. 1-2, pp. 3-14. (In Italian.)

Durgin, F.H. & Sternberg, S. (2002). The Time of Consciousness and Vice Versa. *Consciousness and Cognition*, vol. 11, no. 2, pp. 284–290.

ERA (2000). *Glossary of Terms and Abbreviations, subset-023*. Issue 2.0.0. European Railway Agency. <http://www.era.europa.eu/public/Documents/ERTMS%20Documentation/Mandatory%20Specifications/SUBSET-023%20v200.pdf> (08-02-15)

Eriksen, T.B. (2000). *Tidens historia (The history of time)*. Atlantis, Stockholm. ISBN 91-7486-927-2. (In Swedish.)

Espling, U. (2007). *Maintenance strategy for a railway infrastructure in a regulated environment*. Doctoral thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology.

Esveld, C. (2001). *Modern railway track*. Second edition. MRT-Productions, Delft. ISBN 90-800324-3-3.

Genovesi, P. & Ronzino, C.D. (2003). Un nuovo modello di supporto alla normative di esercizio / A New Model to Support the Operating Regulations. *Ingegneria Ferroviaria*, no. 11, pp. 979-986. (In Italian.)

Glicksohn, J. (2001). Temporal Cognition and the Phenomenology of Time: A Multiplicative Function for Apparent Duration. *Consciousness and Cognition*, vol. 10, pp. 1-25.

Godstransportdelegationen (2001). *Godstransporter för tillväxt – en hållbar strategi (Freight transports for growth – a sustainable strategy)*. Godstransportdelegationen. <http://www.sou.gov.se> (06-09-30). (In Swedish.)

Granström, R. (2005). *Maintenance for improved punctuality: a study of condition monitoring technology for the Swedish railway sector*. Licentiate thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology.

Granström, R. (2008). *Management of condition information in punctuality perspectives*. Doctoral thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology. (Forthcoming.)

Granström, R. & Söderholm, P. (2005). Punctuality Measurements Effect on the Maintenance Process – A Study of Punctuality Statistics for the Swedish Railway. In: *Proceedings of the Eighth International Conference and Exhibit on Railway Engineering*, 29-30 June, London, UK, Engineering Technics Press, Edinburgh. (CD)

Groeneweg, J. (1998). *Controlling the controllable: the management of safety*. DSWO Press, Leiden, The Netherlands.

Gustavsson, C. & Ådegren, A.-B. (2001). *Sökvägar för identifiering av kostnadsdrivare: En jämförelse mellan två bandelar inom Luleå Banområde (Ways to identify cost drivers: a comparison of two track sections in Luleå track area)*. BSc Thesis, Luleå University of Technology. (In Swedish.)

Hagquist, R.F. (1994). High-precision prioritization using Analytic Hierarchy Process: determining state HPMS component weighting factors. *Transportation research record*, issue 1429, pp. 7-14.

Hansson, S. (1996). *Teknikhistoria: om tekniskt kunnande och dess betydelse för individ och samhälle (The history of technology: on technological knowledge and its importance for the individual and society)*. Studentlitteratur, Lund. (In Swedish.)

Harris, N.G. & Ramsey, J.B.H. (1994). Assessing the Effects of Railway Infrastructure Failure. *Journal of the Operational Research Society*, vol. 45, no. 6, pp. 635-640.

Heinrich, H.W. (1959). *Industrial accident prevention: a scientific approach*. Fourth edition. McGraw-Hill, New York.

Henningsson, K. & Wohlin, C. (2004). Assuring Fault Classification Agreement – An Empirical Evaluation. In: *Proceedings of the 2004 International Symposium on Empirical Software Engineering (ISESE'04)*.

Hesselfeld, M., Klahn, V. & Pulwer, H. (1996). Management of railway operations during maintenance by computer-aided construction of timetables. *Computers in railways V*, part 2, pp. 513-521.

Hesslow, G. (1983). *A Deterministic Approach to Causality*. Dissertation, University of Lund, Sweden.

Higgins, A., Kozan, E. & Ferreira, L. (1995). Modelling delay risks associated with train schedules. *Transportation Planning and Technology*, vol. 19, pp. 89-108.

Hollnagel, E. (2004). *Barriers and accident prevention*. Ashgate, Burlington, VT. ISBN 0-7546-4301-8.

Holmgren, M. (2003). *Maintenance related losses: a study at the Swedish National Rail Administration*. Licentiate thesis, Division of Quality and Environmental Management, Luleå University of Technology.

Holmgren, M. (2006). *Maintenance-related incidents and accidents. Aspects of hazard identification*. Doctoral thesis, Division of Operation and Maintenance Engineering & Division of Quality and Environmental Management, Luleå University of Technology.

Horngrén, C.T., Foster, G. & Datar, S.M. (1994). *Cost accounting: a managerial emphasis*. Eight edition. Prentice Hall, Englewood Cliffs, NJ.

Horowitz, A.J. (1978). The subjective value of the time spent in travel. *Transportation Research*, vol. 12, no. 6, pp. 385-393.

Höjer, M. & Mattsson, L-G (2000). Determinism and backcasting in future studies. *Futures*, vol. 32, no. 7, pp. 613-634.

IEC 60300-3-14:2004. *Dependability management – Application guide – Maintenance and maintenance support*, International Electrotechnical Commission.

Institut du Transport Aérien (2000). *Costs of air transport delay in Europe. Final report*. Institut du Transport Aérien.

<http://www.eurocontrol.int/prc/gallery/content/public/Docs/stu2.pdf> (07-10-07)

Inutsuka, F. (2006). Personal contact, Fumiaki Inutsuka, JR East, autumn 2006.

ISO 8601:2000(E). *Data elements and interchange formats – Information interchange – Representation of dates and times*. Second edition. ISO, International Organization for Standardization.

ISO 9000:2000. *Systemer for kvalitetsstyring: grunntrekk og terminologi - Quality management systems: fundamentals and vocabulary, Norsk/engelsk versjon*. Norges standardiseringsforbund, Oslo.

ISO/IEC 15288:2002. *Systems engineering – System life cycle processes*. ISO, International Organization for Standardization.

Jardine, A.K.S & Tsang, A.H.C (2006). *Maintenance, replacement, and reliability: theory and applications*. CRC, Boca Raton, FL. ISBN 9780849339660.

Joborn, M. (2001). *Optimization of Empty Freight Car Distribution in Scheduled Railways*. Dissertation, Division of Optimization, Linköping University, Linköping, Sweden.

Johansson, T. & Carlsson, O. (2003). *Tågtrafiken. Punktlighet och förseningar. Klimatets inflytande på tågtrafiken (Train traffic. Punctuality and delays. The impact of the climate on the train traffic)*. Banverket, Borlänge. (In Swedish.)

Jovanovic, S. (2004). Railway Track Quality Assessment and Related Decision Making. *IEEE International Conference on Systems, Man and Cybernetics*, pp. 5038-5043.

Juran, J.M. (1989). *Juran on leadership for quality: an executive handbook*. Free Press, New York.

Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decisions under risk. *Econometrica*, XLVII, pp. 263-291.

Kans, M. (2006). A holistic perspective of computerised maintenance management support. In: Kumar, U., Parida, A. & Rao, R. B. K. N. (eds.): *Proceedings of the 19th International Congress: Condition Monitoring and Diagnostic Engineering Management (COMADEM 2006)*, Luleå, June 12-15. Luleå University of Technology, Luleå, pp. 231-241. ISBN 978-91-631-8806-0.

Kjellén, U. (2000). *Prevention of accidents through experience feedback*. Taylor & Francis, London. ISBN 0-7484-0925-4.

Konieczny, H. & Weiner, A. (1995). *Järnvägen och dess underhåll / The railway and its maintenance*. Ingenjörsoch sjöbefälsskolan vid Chalmers tekniska högskola, Göteborg. (In Swedish.)

Kreitz, P.-A. (2002). *Förbättringspotential för godskalkyler (Improvement potential for freight costing)*. Rapport 2002:13, TFK – Institutet för transportforskning, Stockholm. ISBN 9188752453. (In Swedish.)

Kufver, B. (2002). *Effektiva produktionsmetoder med hänsyn till tågtrafiken (Efficient production methods considering the train traffic)*. VTI notat 17-2002. VTI, Sweden. (In Swedish.)

Kuriyama, K., Uchiyama, M., Suzuki, H., Tagaya, H., Ozaki, A., Aritake, S., Kamei, Y., Nishikawa, T. & Takahashi, K. (2003). Circadian fluctuation of time perception in healthy human subjects. *Neuroscience Research*, vol. 46, issue 1, pp. 23-31.

König, A. (2004). *Messung und Modellierung der Verlässlichkeit des Verkehrsangebots – Experimente mit Schweizer Befragten (Measurement and modelling of dependability – experiments with Swiss interviewees)*. Dissertation, IVT Institut für Verkehrsplanung und Transportsysteme, ETHZ, Zürich. (In German.)

König, A. & Axhausen, K.W. (2002). *Verlässlichkeit als Entscheidungsvariable (Dependability as decision parameter)*. Arbeitsbericht 110, Verkehrs- und Raumplanung, ETH Eidgenössische Technische Hochschule, Zürich. (In German.)

Lammari, N. & Métais, E. (2004). Building and maintaining ontologies: a set of algorithms. *Data & Knowledge Engineering*, vol. 48, issue 2, pp. 155-176.

Landex, A. & Nielsen, O.A. (2006). Simulering af passagerforsinkelser på jernbaner (Simulation of passenger delays in railways). *Trafikdage på Aalborg Universitet*. (In Danish.) <http://www.trafikdage.dk/td/papers/papers06/Trafikdage-2006-575.pdf> (07-10-30).

Levine, R.V. & Norenzayan, A. (1999). The pace of life in 31 countries. *Journal of Cross-Cultural Psychology*, vol. 30, no. 2, pp. 178-205.

Levine, R.V., West, L.J. & Reis, H.T. (1980). Perceptions of Time and Punctuality in the United States and Brazil. *Journal of personality and social psychology*, vol. 38, no. 4, pp. 541-550.

Luftfartsverket (2000:1). *Delay analysis report at Swedish airports DARSA Arlanda år 2000*. Luftfartsverket, Norrköping, Sweden. <http://www.lfv.se> (04-01-28).

Luftfartsverket (2000:2). *LURT Slutrapport med DARSA årsrapport 2000 (LURT Final report with DARSA yearly report 2000)*. Luftfartsverket, Norrköping, Sweden. <http://www.lfv.se> (04-01-28). (In Swedish.)

Lundberg, K. (2005). *Rail operation and maintenance in an Environmental Management System context – current practice and potential improvement*. Licentiate thesis, KTH, Stockholm ISBN 91-7178-072-6.

Mattsson, L.-G. (2004). *Train service reliability. A survey of methods for deriving relationships for train delays*. Unit for Transport and Location Analysis, Division of Systems Analysis, Department of Infrastructure, KTH, Stockholm.

McLay, G.D. (2000). *Understanding the role of information in the passenger rail industry*. Dissertation, University of Southampton.

Metz, D. (2004). Travel Time – Variable or Constant? *Journal of Transport Economics and Policy*, vol. 38, part 3, pp. 333-344.

Mizuno, S. (1988). *Management for Quality Improvement: The Seven New QC Tools*. Productivity Press, Cambridge, MA.

Mohns, P. (2007). Personal contact, Peter Mohns, SL Stockholm Local Traffic, autumn 2007.

Mora, M., Forgionne, G., Gupta, J.N.D., Garrido, L., Cervantes, F. & Gelman, O. (2006). A Strategic Descriptive Review of the Intelligent Decision-making Support Systems Research: the 1980-2004 Period. In: Gupta, J.N.D, Forgionne, G.A. & Mora, M. (eds.): *Intelligent decision-making support systems: foundations, applications and challenges*, Springer, London, pp. 441-462. ISBN 1-84628-228-4.

Mücke, W. (2002). *Operations Control Systems in Public Transport*. Tetzlaff, Hamburg, 2002. ISBN 3-87814-801-1.

Nelldal, B.-L., Fröidh, O., Kottenhoff, K., Lindahl, A. & Troche, G. (2000). *Tågtrafikplanering (Train traffic planning)*. Andra upplagan. Avdelningen för Trafik- och transportplanering, KTH, Stockholm. (In Swedish.)

Nelson, D. & O'Neil, K. (2000). Commuter Rail Service Reliability: On-Time Performance and Causes for Delays. *Transportation research record*, vol. 1704, pp. 42-50.

von Neumann, J. & Morgenstern, O. (1953). *Theory of games and economic behavior*. Third edition. Oxford U.P., New York, Princeton U.P., Princeton. ISBN 0-691-00362-9.

Nissen, A. (2005). *Analys av statistik om spårväxlars underhållsbehov (Analysis of the need for maintenance of railway turnouts)*. Licentiate thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology. (In Swedish.)

Nyman, D. & Levitt, J. (2001). *Maintenance Planning, Scheduling, & Coordination*. Industrial Press, New York.

Nyström, B. (2005). *Punktlighet (Punctuality)*. Forskningsrapport, Järnvägstekniskt centrum, Luleå University of Technology, Luleå. (In Swedish.)

Nyström, B. & Karlsson, S. (2006). Description of railway delays – a study from the views of the stakeholders. In: Kumar, U., Parida, A., Rao, R. B. K. N. (eds.): *Proceedings of the 19th International Congress, Condition Monitoring and Diagnostic Engineering Management (COMADEM 2006)*, Luleå, June 12-15. Luleå University of Technology, Luleå, pp. 701-710. ISBN 978-91-631-8806-0.

Parsons, T. (1990). *Events in the semantics of English: a study in subatomic semantics*. MIT Press, Cambridge, Mass. ISBN 0-262-16120-6.

Profillidis, V.A. (2000). *Railway Engineering*. Second edition. Aldershot, Ashgate.

Ramadhan, R.H., Wahhab, H.I.A. & Duffuaa, S.O. (1999). The use of an analytical hierarchy process in pavement maintenance priority ranking. *Journal of Quality in Maintenance Engineering*, vol. 5, no. 1, pp. 25–39.

Reason, J.T. (1997). *Managing the Risks of Organizational Accidents*. Ashgate, Hants, England & Burlington, VT.

Rieckenberg, T. (2004). *Telematik im Schienengüterverkehr – ein konzeptionell-technischer Beitrag zur Steigerung der Sicherheit und Effektivität (Telematics in railway freight transport – a conceptual-technical contribution to improved safety and efficiency)*. Dissertation, Technische Universität Berlin. (In German.)

Rietveld, P., Bruinsma, F.R. & van Vuuren, D.J. (2001). Coping with unreliability in public transport chains: A case study for Netherlands. *Transportation Research Part A*, vol. 35, no. 6, pp. 539-559.

Riksdagens revisorer (2002). *Tåg i otid – tågplanens betydelse för punktligheten (Trains on time – the impact of the train plan on punctuality)*. Rapport 2002/03:18, Riksdagens revisorer, Stockholm. <http://www2.riksdagen.se> (08-02-11) (In Swedish).

Roeckelein, J.E. (2000). *The Concept of Time in Psychology: a Resource Book and Annotated Bibliography*. Greenwood Press, Westport, Conn.

Rosenlind, S., Lind, G. & Troche, G. (2001). *LIME: model for capacity utilisation and profitability of a railway line*. Division of Traffic and Transport Planning, Department of Infrastructure and Planning, KTH, Stockholm.

Rudnicki, A. (1997). Measures of regularity and punctuality in public transport operation, In: Papageorgiou, M. & Pouliezios, A. (eds.): *Transportation systems 1997: a proceedings volume from the 8th IFAC/IFIP/IFORS Symposium*, Chania, Greece, June 16-18. Oxford, pp. 661-666.

Rudolph, R. (2003). Allowances and margins in railway scheduling. In: *Proceedings of the World Congress on Railway Research (WCRR 2003)*, Edinburgh, Scotland, 28 September - 1 October, pp. 230-238.

Ryan, T.A. (1990). Roadway Vehicle Delay Costs at Rail-Highway Grade Crossings. *Transportation research record*, issue 1262, pp. 31-38.

Saaty, T.L. (1980). *The analytic hierarchy process: planning, priority setting, resource allocation*. McGraw-Hill, New York. ISBN 0-07-054371-2.

SFS 2007:1027. *Förordning med instruktion för Banverket (Ordinance with instruction for Swedish Rail Administration)*.

<http://www.samhallsguiden.riksdagen.se> (08-01-29). (In Swedish).

Shaw, J. (2001). Winning Territory. In: May, J. & Thrift, N. (eds.): *TimeSpace: Geographies of Temporality*. Routledge, London, pp. 120-132.

SIKA (2002). *Översyn av samhällsekonomiska metoder och kalkylvärden på transportområdet (Review of methods and parameters of social economy within transportation)*. Rapport 2002:4, SIKA (Statens institut för kommunikationsanalys – Swedish Institute for Transport and Communications Analysis), Stockholm. <http://www.sika-institute.se> (08-02-11) (In Swedish).

SJF 615.2 (1990). *TFÖR instruktioner (TFÖR instructions)*. SJ.

SJK (1999). *Järnvägsdata 1999 (Railway data 1999)*. Tredje upplagan. Svenska järnvägsklubben, Stockholm. ISBN 91-85098-85-X.

Simson, S.A., Ferreira L. & Murray, M.H. (2000). Railway Track Maintenance Planning. In: *Railroad track engineering and maintenance: Passenger rail planning and operations*, pp. 29-33. Transportation Research Board, National Research Council National Academy Press, Washington, D.C. ISBN 0-309-06692-1.

Slagmolen, M. H. (1980). *Train choice: measurement of the time-table-quality of rail services based on an analysis of train choice behaviour*. Dissertation, Rotterdam.

SS 4410505. *Tillförlitlighet – Ordlista. / Dependability – Terminology. Utgåva 3*. SIS (Swedish Standards Institute), Stockholm, 2000.

SS-EN 13306:2001. *Underhåll – Terminologi. Maintenance terminology*. SIS (Swedish Standards Institute), Stockholm.

SS-EN 50126:1999. *Järnvägsanläggningar – Specifikation av tillförlitlighet, funktionssannolikhet, driftsäkerhet, tillgänglighet, underhållsmässighet och säkerhet (RAMS) / Railway specifications – The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)*. SIS (Swedish Standards Institute), Stockholm.

SS-EN ISO 9001:2000. *Ledningssystem för kvalitet: krav – Quality management systems: Requirements*. SIS (Standardiseringen i Sverige / Swedish Standards Institute), Stockholm. (In Swedish.)

The Swedish Ministry of Enterprise, Energy and Communications (2005). *Regleringsbrev för budgetåret 2006 avseende Banverket m.m. inom utgiftsområde 22 Kommunikationer (Appropriation directions for the fiscal year 2006 to the Swedish Rail Administration, etc., within expenditure area 22 Communication)*. <http://webapp.esv.se/statsliggaren> (07-05-28). (In Swedish.)

Söderholm, P., Holmgren, M. & Klefsjö, B. (2007). A process view of maintenance and its stakeholders. *Journal of Quality in Maintenance Engineering*, vol. 13, no. 1, pp. 19-32.

TF601 (2000). *Riktlinjer för tidtabellskonstruktion för tåg på statens spåranläggningar (Guidelines for timetable construction on the Swedish state railway lines)*. Banverket, Borlänge. (In Swedish.)

Transek (2006). *Resenärernas upplevelser av inställda turer och förseningar. SL som ett exempel – kartläggning och värdering av konsekvenserna (Passengers affected by cancellations and delays. SL as an example – Mapping and valuation of the consequences)*. Transek Consultancy. (In Swedish.)

Tversky A., Slovic, P. & Kahneman, D. (1990). The Causes of Preference Reversal. *The American Economic Review*, vol. 80, no. 1, pp. 204-217.

UIC (1996). *Code 405: Links between railway infrastructure capacity and the quality of operations*. First edition. UIC (Union Internationale des Chemins de fer – International Union of Railways).

UIC (2000). *Code 451-1: Timetable recovery margins to guarantee timekeeping – Recovery margins*. Fourth edition. UIC (Union Internationale des Chemins de fer – International Union of Railways).

UIC (2001). *RailLexic Version 3.0*. UIC (Union Internationale des Chemins de fer – International Union of Railways). (Database.)

Vägverket (2002). *Vägverket Region Skånes dialogunderlag till Nationell Plan för Vägtransportsystemet perioden 2004-2015 (Swedish Road Administration Skåne county's basis for dialogue on the National Plan for the Road Transport System the period 2004-2015)*. Vägverket, Region Skåne. http://www.vv.se/filer/2747/Underlagsmaterial_VSK.pdf (08-02-11). (In Swedish).

Wahlström, A. (2006). *Kommunikationsflöden vid akut felavhjälpning inom Banverket (Flows of communication during acute fault correction within Banverket)*. MSc thesis, Luleå University of Technology. <http://epubl.ltu.se/1402-1617/2006/146/LTU-EX-06146-SE.pdf> (In Swedish.)

Whitrow, G.J. (1988). *Time in history: the evolution of our general awareness of time and temporal perspective*. Oxford Univ. Press, Oxford. ISBN 0-19-215361-7.

Wickens, C.D. (1992). *Engineering psychology and human performance*. Second edition. HarperCollins, New York.

Widlert, S. (1998). Stated Preference Studies: The Design Affects the Results. In: de Dios Ortúzar, J., Hensher, D. & Jara-Díaz, S. (eds.): *Travel behaviour research: updating the state of play*. Elsevier, New York, pp. 105-122.

Witlox, F. & Vandaele, E. (2005). Determining the Monetary Value of Quality Attributes in Freight Transportation Using a Stated Preference Approach. *Transportation Planning and Technology*, vol. 28, no. 2, pp. 77-92.

Yin, R.K. (2003). *Case study research: design and methods*. Third edition. Sage, Thousand Oaks, Calif. ISBN 0-7619-2553-8.

Zoeteman, A. (2004). *Railway Design and Maintenance from a Life-Cycle Cost perspective: A Decision-Support Approach*. Dissertation, TRAIL Thesis Series, Delft, The Netherlands. ISBN 90-5584-058-0.

Östlund, B., Bolinder, C., Janson, G. & Andersson, M. (2001). *Samhällsekonomisk värdering av järnvägsinvesteringar för godstrafik (Social economy valuation of railway investments for freight traffic)*. TFK - Institutet för transportforskning, Borlänge. (In Swedish.)

Appendix A: Basic Railway Technology

In this appendix, the basics of railway technology are described.

The fundamentals of railways are the same all over the world, but the specifications differ. The information in this appendix applies to the Swedish railway and comments are given on a few major international differences. A *railway system* is defined as consisting of vehicles, steered by a track on a dedicated area, which are governed by a signalling system. Internationally, the degree to which railways are distinguished from underground and tramway systems varies (Andersson & Berg, 2001). In the form of one-dimensional guided movement of metal surfaced wheels on surfaced metal track, railways appeared in the beginning of the 19th century in British mines (Profillidis, 2000).

This appendix describes the track, different modes of propulsion, mainly electricity, as well as signalling, used for collecting information and giving orders in the railway system. Finally, telecommunications are briefly explained. The partition of the railway into these four disciplines reflects the traditional division of personnel at Banverket.

A.1 Track

Most railway tracks are of ballasted type, although several non-ballasted types are in operation throughout the world (Esveld, 2001). In Figure A1, a cross section of a ballasted track is illustrated. At the top is the rail, connected to sleepers, made of concrete or wood. The rails rest on the sleepers via rail pads⁴⁴ in order to attenuate vibrations. The sleepers rest in ballast, which rests on sub-ballast. The ballast consists most often of macadam (crushed stone), which allows water to drain, damps vibrations and distributes load to levels beneath it. In some cases a geotextile fibre sheet is placed below the ballast to hinder the finer material of the sub-ballast from penetrating it. In this way, the capillary transport of water, which might cause freezing or lower the limit shear stress of the ballast, is prohibited (Konieczny & Weiner, 1995). The sub-ballast, consisting of gravel-sand, ensures quick drainage, protects the subgrade from being penetrated by ballast stones and further distributes load. Together, rails, sleepers, ballast and sub-ballast make up the track.

The track rests on the subgrade of soil, which flattens uneven terrain. The subgrade accounts for a big share of the total investment cost of a new railway; recently 30-45% in continental Europe (Profillidis, 2000).

⁴⁴ Rail pads are not used in the case of wooden sleepers.

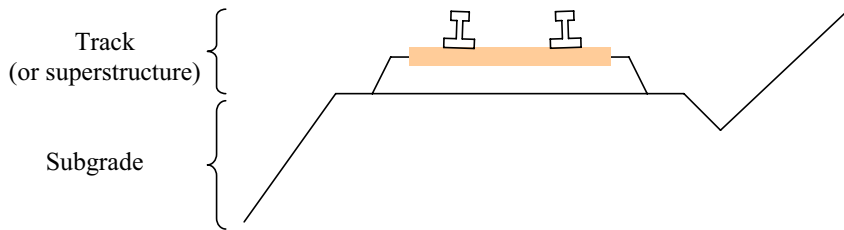


Figure A1. A ballasted track system schematically. Note the inward tilting of the rails. To the right, a trench is seen.

Rail fastenings may be of different types, the simplest consisting of two spikes for wooden sleepers. The rail fastenings are to fix the rail along the three axes and provide torsion resistance. In this way the track gauge is kept constant and slipping of the rails is hindered (Profillidis, 2000). The type seen in Figure A2 is in wide use in Sweden today.

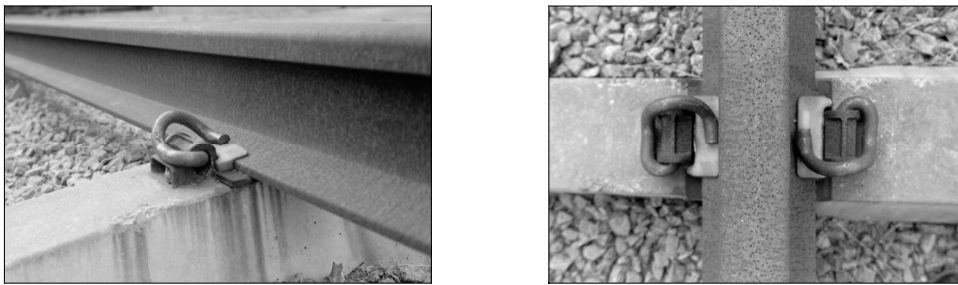


Figure A2. Pandrol rail fastenings on a concrete sleeper. Left: the fastening on the gauge side of rail. Right: Gauge side and field side fastenings seen from above. Hydraulic equipment is used to obtain the force needed to assemble the springs of the fastenings. As concrete conducts electricity, the plastic rail pads, located between the fastenings and the rail, hinder electric connection between the two rails of the track, thereby ensuring track circuit integrity (explained in the Section A.3.1 Train Traffic Control and Section Block, p. 89).

A.1.1 Track Geometry

The track gauge is defined as the distance between the rails, measured 14 mm below top of rail between opposing rail surfaces. Track gauge differs between countries and even within countries. In Sweden so called standard gauge, 1435 mm, is used. In Finland and Russia, for example, the track gauge is 1524 mm.

The rail is continuous⁴⁵ – the rail segments are welded together. The forces in the rail caused by temperature variations are to be withstood by fastenings, sleepers and ballast; otherwise, the track will be deformed, even causing derailment. Great axial

⁴⁵ This is entirely true for only one of the two rails of a railway track, as the signalling system requires gaps on rail, at intervals of usually about one kilometre, shorter at stations. This is explained in Section A.3 Signalling (page 89).

forces in the rail are also caused by braking trains, which may cause slipping of the rails (Profillidis, 2000; Banverket, 2002; Nelldal et al., 2000; <http://www.rhk.fi> (08-01-29)).

Culverts are placed perpendicular to the track to allow drainage. In the past they were made of stone; nowadays, they are made of corrugated sheet or concrete (Konieczny & Weiner, 1995).

A.1.2 Possible Damages

Each item in the track system described above is intended to keep the track in place. To inspect the track bed, measurement vehicles⁴⁶ are used.

If a wheel is forced to slide instead of roll, wheel flats may result. These might cause damage to both rolling stock and rail, and may even break the rail (Banverket, 2002).

Lateral buckling of the rails may be caused by elongation of the rail due to temperature differences. This phenomenon may cause derailment. Nowadays, this happens rarely. See Figure A3 and Figure A4.



Figure A3. Old track laid with gaps in between rails, which are intended allow for temperature differences. If a gap is not wide enough, lateral buckling may occur. The gaps (exaggerated in the figure) give a characteristic sound when travelling.



Figure A4. Modern long welded rails have tough fastenings and sleepers to withstand the expansion forces in the rails due to the temperature differences.

A.1.3 Turnouts and Crossings

Turnouts allow a track to be split in two (sometimes three) tracks and the train to change course. At crossings, two tracks meet at a grade with no change of course.

The demands on a turnout (and crossing) are manifold; the vehicles should be able to pass it at high speed, it should be built in the right place to support operational demands and be compatible with signalling equipment. The basic parts of a turnout are given in Figure A5. The position of the switch (or tongue) rails decides the course of the next train.

⁴⁶ Measurement car STRIX (up to 200 km/h), Plasser rail trolley (two models; 50 and 70 km/h, respectively).

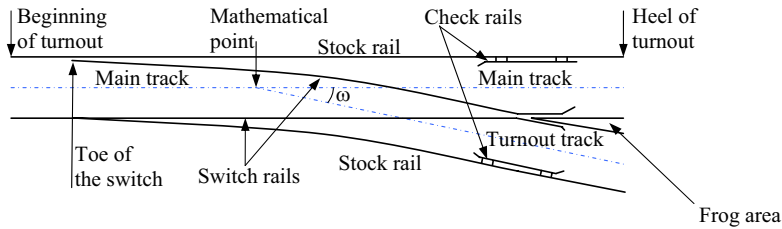


Figure A5. Components of a turnout. The turnout equals the switch plus the check (or guard) rails and the frog. The toe of the switch is operated by the switch drive (or manually). The frog angle ω is often given as the tangent of the angle, for example 1:9. A smaller angle, such as 1:27, allows higher train speeds. (After Profillidis, 2000.)

A.2 Power Supply

The first railway powered by steam was opened in 1825 between Stockton and Darlington in England (Hansson, 1996) and trains pulled by steam locomotives are still in use in Africa and Asia. Among their disadvantages are low fuel efficiency, limited power and expensive maintenance. During the 1930s, diesel locomotives appeared, becoming more popular after the second world war. The first electric train dates back to 1879, and after 1920 electric traction was extensively used. The operating cost of electric traction is lower than for diesel traction, but the need for power delivery subsystems makes the initial investment greater (Profillidis, 2000).

A.2.1 Electric Traction

The electricity may be transferred to the locomotive by a conductor rail, and back via one or both of the ordinary rails or a dedicated ground rail. An advantage of conductor rail compared to overhead transmission line is that tunnels do not have to be built so high, but a disadvantage is their sensitivity to snow. As with track gauges, countries have chosen different designs (Profillidis, 2000). In Sweden, 15 kV 16 2/3 Hz, delivered via contact wire, is used for traction. The electricity is supplied by from the three-phase 130 kV 50 Hz public grid (Banverket, 2002).

A.2.2 Contact Wire and Pantograph

The contact wire is zigzagged relative to the centre line of the track, in order to avoid wearing the same spots on the slippers of the pantograph on the locomotive, see Figure A6. The pantograph's slippers are made of carbon.



Figure A6. A pantograph of a Green Cargo locomotive receiving power from the contact wire at Älvsbyn station. The pantograph's two carbon slippers simultaneously touch the contact wire. Photo: Rikard Granström.

A.2.3 Auxiliary Power

Auxiliary power, distributed using 10 kV or 20 kV, 50 Hz, is used for non-traction infrastructure, for example, switch-heaters that keep switches from freezing (Banverket, 2002).

A.3 Signalling

The signalling system ensures safe operation and guides the trains in the railway network. Railway signalling is generally fail-safe, meaning that failures always occur “towards a safe state”. For example, a faulty light bulb causes the signal to show stop, and if all bulbs are dark, it is to be interpreted as stop (Andersson & Berg, 2001).

A.3.1 Train Traffic Control and Section Block

To prevent train collisions, the train traffic controller at station A calls station B's controller when the train passes A towards B. Then, they block the section in both directions by turning the stop signs on. When the train passes B, the controller at station B calls station A to report that the section between the stations is free. This manual system⁴⁷ may not be safe enough. Furthermore, if the stations are not very close to each other, train traffic will have to be sparse as only one train at a time may travel between the stations (Nelldal et al., 2000).

On most of the Swedish railway network, the manual system has been exchanged for an automatic system. The principle of section block using track circuiting is illustrated in Figure A7. The rails in a section have a voltage over them. The rails are connected to a relay, which functions due to the potential over it. When a wheel set connects the rails, the voltage over the relay will drop, and the relay will release. Then, the signs leading to that section are automatically turned to stop, so no other train may enter it. One rail is continuous, the other has insulators between rail

⁴⁷ In Sweden, this is called ‘TAM’.

sections (Banverket, 2002). Section length varies up to about 2500 m (Nelldal et al., 2000).

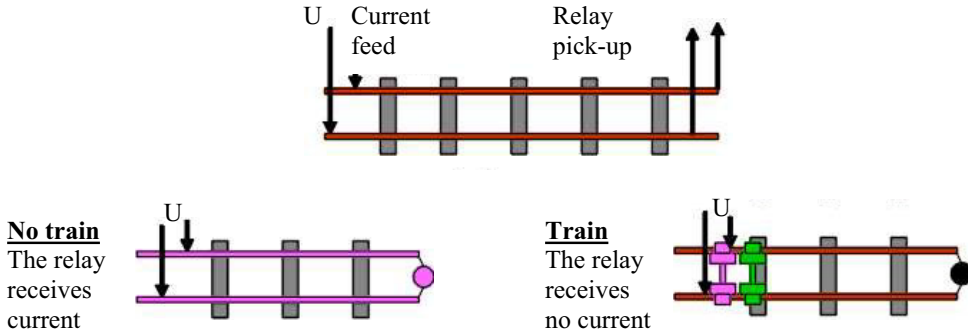


Figure A7. The principle of section block. The voltage differential between rails is approximately 7 volts. When a wheel axle of a train short-circuits the two rails, the corresponding relay gets no current. A wheel axle is allowed to have a resistance of maximum 0.1Ω (Banverket, 2002).

A.3.2 Light Signals

In the early days of railways, semaphores were used for control. With the advent of highly luminous light sources, possible to see even during daytime, semaphores were outdated. A commencement signal tells the driver the maximum allowed speed or to stop before the signal. See Figure A8.

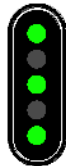


Figure A8. A commencement signal. Three green lights mean “Drive 40, short route”. Therefore, the train is allowed to drive maximum 40 km/h and the distance to a signal showing stop might be as short as 250 metres. Next to stop (one red light), this is the most restrictive signal (Banverket, 2002).

Trains often move too fast to be able to stop before a red signal if the brakes are not applied before the driver can see the commencement signal. Hence, the driver needs to know in advance what the next commencement signal will show. Therefore, an announcement signal, see Figure A9, is located before the commencement signal in order to make it possible for the driver to reach a full stop before the commencement signal. Signals differ a lot between countries; see Bailey (1995) for an overview.



Figure A9. An announcement signal. Two flashing green lights mean “Expect 40” (km/h) (Banverket, 2002).

To ensure that the trains do not exceed the speed limits, locomotives are equipped with the ATC (Automatic Train Control) system, specified in the 1970s (Andersson & Berg, 2001). The ATC transmitter of the locomotive emits radio waves to the track. When it encounters a balise in the track the balise responds with its information. The balise may be of the type that transmits fixed data (for example, maximum allowed speed, if it is the same for all vehicles) or dynamic data (for example that a signal is currently set to stop). If the speed of the train is too high, the ATC system beeps the driver to slow down. If the train does not slow down, ATC brakes the train.

The train traffic controllers direct the movements of trains by deciding which train will take which route. This is performed by appropriate shunting and setting of signals. This is done from a computer at a train traffic control office or directly by the turnout (rare nowadays). The controller has to deal with deviations, for example when there is an obstacle such as a road vehicle on a level crossing. This might be indicated by a vehicle sensor installed in the level crossing, which sends the information to the controller's screen. At the same time, signals are automatically set to red. There are also wires in the gates to detect if any gate is broken.

A.4 Telecommunications

Telecommunications in Banverket includes radio communication and detectors.

A.4.1 Detectors

There are different detectors in the track. Here, the most common ones are mentioned.

Hot wheel/blocked braking detectors bolometrically measure temperature in the wheel-rail contact. High temperature indicates hindered wheel movement.

Wheel flat detectors detect wheel flats. Overheating detectors are used to measure temperature in wheel bearings. Profile supervision detectors use light beams in order to check train width (Banverket, 2002).

Appendix B: Databases

In this appendix, some of the databases most important to maintenance and punctuality in the Swedish railway are described.

In the Swedish railway sector, there are approximately 200 databases and data collection systems. However, some outdated and highly specialised programs are included in that figure. A few systems most related to maintenance and punctuality are described below (<http://www.banverket.se> (08-02-15)).

The BIS track information system is a database that describes the railway infrastructure. Several other systems use BIS as a reference system. BIS gives an “instant picture” of the infrastructure, as it only stores the present state of affairs. Ofelia is a database on occurred railway infrastructure failures. All items are described in a predefined structure, in which failures and repairs are entered. Bessy is a system for inspection of railway infrastructure. The inspector in the field uses a handheld computer that is a part of the system to enter data and transfers the results to the database. The Bessy database thus contains railway infrastructure inspection plans and remarks from all inspectors.

TrainPlan is a graphical system for constructing timetables. Argus is the train traffic control (dispatching) system used at CTTC Boden. It is used by the train traffic controllers and traffic informers. For a shorter period of time, it stores the routes of the trains, a function that might be used, for example, in accident investigations. TFÖR is a database on train delays. MAPS is an Excel-based tool for investigating TFÖR data (SJF 615.2, 1990; Dahl, 1997). A newer tool, Duvan, combines data from a number of systems, including TFÖR. Opera is a system where the TOCs enter the data for their trains and one might monitor the train traffic.

The economy system Agresso is used for accounting and billing.

Appendix C: Delay Attribution Codes

In this appendix, the delay attribution codes of some railways and air traffic are described, in order to give the reader a sense for different delay attribution systems.

C.1 Banverket

About one hundred delay codes are used by Banverket, of which about 20 concern infrastructure (Banverket, 1999). TFÖR codes often tell which phenomenon occurred (in the code's associated data), but not which role the described event played. However, codes such as 'F24 Detector alarm wheel damage' tell both which phenomenon (wheel damage) and role (symptom). TFÖR codes most often describe symptoms (and consequences, as delay minutes are most often given), making it difficult to find causes using any of the (accident) models. However, the secondary TFÖR delay codes, per definition, reflect a sequential model. Some, however, also facilitate a systemic view, for example 'L26 Scarcity of track' tells that the volume of train traffic and the volume of track do not have a good relation.

C.2 A-Train

The ten delay codes that are used by A-Train, which operates the airport train between Stockholm Central Railway Station and Arlanda Airport, are listed in Table C1, and A-Train says the codes 1-9 cover 85-90% of total delays. It is also possible for the drivers to complement the selected code with free text. The reports are entered into the Diver database (Nyström & Karlsson, 2006).

- | |
|--|
| <ol style="list-style-type: none">0. Unknown/Other1. Stop signal2. Train ahead3. Crossing train path4. Train traffic control5. Infrastructure of Banverket6. Infrastructure of A-Train7. Person in track environs8. Passenger9. Own vehicle |
|--|

Table C1. Delay codes on the quality report filled in by A-Train's drivers.

C.3 Delay Attribution Board

The Delay Attribution Board is a London-based joint railway industry body (Delay Attribution Board, 2007). The entries in the TRUST code set number more than 200. The codes are grouped according to responsibility, for example, 'Non Passenger's Charter Excludable infrastructure problems'.

Some of the codes are very specific, for example, ‘Delays to other trains caused by a Railhead Conditioning train taking an unusually long time in section or at a location’. The author remarks that such a delay may, in general, alternatively be coded as a train that is delayed and gives rise to subsequent secondary delays. The infrastructure code ‘Possession cancellation’, that is, the planned track work has been cancelled, indicates a systemic view on the problem.

C.4 JR East

The Japanese railway JR East uses 42 codes, of which about 10 concern railway infrastructure failures (Inutsuka, 2006). About as many concern accidents. Specific codes include ‘Train fire’, ‘Wagon failure’, ‘Late driver’, ‘Signal passed at danger’ and ‘Speed limit exceeded’. Another code is ‘Wrong stop position’, that is, the train did not halt at the designated spot by the platform.

C.5 Air Traffic

The air traffic controllers do not perform delay attribution. It is the ground personnel that report delays⁴⁸ and air traffic controllers are not involved (Nyström & Karlsson, 2006). This is because it is the events that take place before the aeroplane leaves the gate that airlines might influence and the airlines have designed the delay codes. Scandinavian uses many codes in addition to the codes specified by IATA (International Air Transport Association). IATA’s delay attribution codes⁴⁹ number about 100. Compared to TFÖR, a larger share of IATA codes concerns passengers and events before departure and to a greater extent gives the cause, not the symptom. Different from TFÖR, the IATA delay codes specifically describe that aeroplane maintenance has run late, that spare parts or maintenance equipment have been lacking, that personnel have been delayed from other flights (‘Crew rotation’), that work on the ground has been delayed by bad weather, that the ground time allotted has been shorter than needed for the aeroplane, as well as strikes.

⁴⁸ At least at Scandinavian airlines.

⁴⁹ http://www.b737mrg.net/downloads/b737mrg_delaycodes.pdf (08-02-18).

Appendix D: Glossary of Terms and Abbreviations

In this appendix, standards and jargon are explained. The terms given without references are either defined by the author or are common usage. In cases of varying terminology, synonyms are given. (Swedish terms are given in brackets.)

ATC, Automatic Train Control. A safety system that hinders the train from running over the speed set and passing signals at danger. It consists of balises in the track and emitters attached to the train. Another term is Automatic Train Protection (Banverket, 2002).

Availability / Performance (Driftsäkerhet). The ability of an item to be in a state to perform a required function under given conditions at a given instant of time or during a given time interval, assuming that the required external resources are provided (SS-EN 13306). From this definition, it is understood that high reliability and swift corrective maintenance without disturbing operations give high availability.

Backlog. Plannable but not yet complete maintenance work (Nyman & Levitt, 2001).

Balise (Balis). A passive transponder mounted on the track which can communicate with a train passing over it (ERA, 2000)

Banverket, The Swedish Rail Administration. The Swedish Railway Infrastructure Manager. Banverket is an authority that has an overall responsibility for the railway transport system. The responsibility also includes underground and tramway systems (SFS 2007:1027).

Bessy. Database containing railway infrastructure inspection remarks. Managed by Banverket.

BIS. Database on railway infrastructure assets. Managed by Banverket.

Condition-based maintenance (Tillståndsbaserat underhåll). Preventive maintenance based on performance and/or parameter monitoring and the subsequent actions. Performance and parameter monitoring may be scheduled, on request or continuous (SS-EN 13306).

Corrective maintenance (Avhjälpande underhåll). Maintenance carried out after fault recognition and intended to put an item into a state in which it can perform a required function (SS-EN 13306).

Delay (Försening). That the train lags compared to the timetable. It is usually measured in minutes.

Extra delay / Cumulative delay (Merförsening). The increase in time-lagging timetable between two places, usually stations.

Failure (Felhändelse). The termination of the ability of an item to perform a required function.

Note 1: After failure the item has a fault, which may be complete or partial.

Note 2: ‘Failure’ is an event, as distinguished from ‘fault’ which is a state (SS-EN 13306).

Failure code. Tells why the item failed.

Inspection (Inspektion / Besiktning / Kontroll). To check for conformity by measuring, observing, testing or gauging the relevant characteristics of an item.

Note: Generally inspection can be carried out on, before, during or after other maintenance activity (SS-EN 13306).

Item (Enhet). Any part, component, device, subsystem, functional unit, equipment or system that can be individually considered (SS-EN 13306). So, item is a very general concept, as an item might consist of items.

JBV, Jernbaneverket. The Norwegian Railway Infrastructure Manager.

LCC, Life Cycle Cost (Livscykelkostnad). The cost incurred by an item from cradle to grave.

Maintainability (Underhållsmässighet). The ability of an item under given conditions of use, to be retained in, or restored to, a state in which it can perform a required function, when maintenance is performed under given conditions and using stated procedures and resources (SS-EN 13306).

Maintenance supportability (Underhållssäkerhet). The ability of a maintenance organization of having the right maintenance support at the necessary place to perform the required maintenance activity at a given instant of time or during a given time interval (SS-EN 13306).

Ofelia. Database on occurred railway infrastructure failures. Managed by Banverket.

Ontology (Ontologi). Formal conceptualisation of a real world, sharing a common understanding of this real world (Lammari & Métais, 2004).

Pantograph (Strömavtagare). Device for transmitting power from the contact wire to the train.

Preventive maintenance (Förebyggande underhåll). Maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item (SS-EN 13306). Preventive maintenance is predetermined (occurring at intervals in time or, example given, production volume) or condition based, in essence based on results from inspection or monitoring.

Primary delay / Exogenous delay (Primärförsening). A delay that directly strikes that particular train (Dahl, 1997). Examples include the train being delayed by turnout failure, engine malfunction or driver's mistake.

Process (Process). Transforms input to output through applying one or more activities (SS-EN ISO 9001:2000).

Punctual (Punktlig). That one or several events occur when agreed between involved stakeholders. Punctual is an attribute variable in the sense that an event is either punctual or not. Note that whether an event is punctual or not, is undefined when an agreement does not exist or is unclear.

Punctuality (Punktlighet). The extent to which one or several events occur when agreed. It might be operationally defined and documented as of below:

Punctuality of an event is one or several instances of a tuple on the form
<Stakeholder, Place, Event, TimeAgreedForTheEvent, TimeOccurredForTheEvent >, where

Stakeholder → Passenger | Freight customer | TOC | Maintainer of infrastructure |
Maintainer of rolling stock | IM

Place → AnyPlace

Event → Departure | Arrival | Passage of Passenger | Freight | Train | Vehicle

TimeAgreedForTheEvent → DateAndTime

TimeOccurredForTheEvent → DateAndTime | Cancelled

Reliability (Funktionssäkerhet). The ability of an item to perform a required function under given conditions for a given time interval.

Note: The term 'reliability' is also used as a measure of reliability performance and may also be defined as a probability (SS-EN 13306).

Root cause (Grundorsak). The underlying cause of a problem.

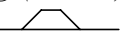
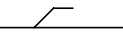
Secondary delay / Knock-on delay / Reactionary delay / Cascading delay (Sekundär försening / Följdförsening). A delay caused by another train's deviation from the timetable, as opposed to primary delay (Dahl, 1997).

Signalling system (Signalsystem). Among the tasks of a railway signalling system is to give information on where the trains are and assure safety.

Slack. The difference between the running time according to the train plan (timetable) and basic running time. It might alternatively be expressed as a quotient, such as 1.07.

SL. Stockholm Local Traffic (Stockholms lokaltrafik). Manages public transport in the Greater Stockholm area (<http://www.sl.se> (04-04-01)).

Stakeholder (Intressent). An interested part having a right, share or claim in the system or in its possession of characteristics that meets that party's needs and/or expectations (ISO/IEC 15288:2002).

Station / Am eng: Siding (Station). A station permits trains to meet or overtake on single-track lines:  This is called siding in American English, which should not to be confused with a siding in British English, which usually is a dead end:  In this thesis, the term *railway station* is used for places where passenger exchange may take place.

Step effect (Språngeffekt). The phenomenon that the valuation (that is, cost) of a delay is not a continuous function of delay length, but increases momentarily when reaching a certain length of delay. Examples include freight trains being delayed for ships with fixed departure time and passengers who miss their connections.

TFÖR (TågFÖRingsregistret). Database on train delays. Managed by Banverket. MAPS is a tool for investigating TFÖR data (Dahl, 1997; SJF 615.2, 1990).

Timeliness (Läglighet). The extent to which a journey or transport occurs when desired. Note that timeliness is highly subjective, and varies between stakeholders. Timeliness includes the facet of length of journey time.

TOC, Train Operating Company (Operatör / Trafikutövare / Järnvägsföretag). An organisation that runs trains, for example Green Cargo, SJ and Veolia.

Track work (Banarbete). Examples of track works are to install a new turnout, inspect the sleepers and change a light bulb.

Train number (Tågnummer). A reserved train path in the timetable. A train might be assigned different train numbers on different parts of its route.

Train plan / Timetable / Schedule (Tågplan / Tidtabell). A train plan tells where each train is supposed to be at a given instant.

Train traffic controller, Train dispatcher, Dispatcher (Tågledare / Tågklarare / Fjärrtågklarare). Person controlling the trains.

Turnout (Spårväxel). An item that allows the railway track to be split into two (sometimes three) and the moving vehicle to change course (Profillidis, 2000).

Single-commodity train (Systemtåg). Freight train for specific transport task, most often for a single customer and constituting a link in a production chain (SJK, 1999).

Valuation / Assessment (Värdering). In this thesis, to calculate the cost of unpunctuality, given unpunctuality data.

Vignette (Scenario). A story that describes a hypothetical situation.

Wheel flat (Hjulplatta). If a wheel is forced to slide instead of roll, it may result in out-of-round wheels, that is, wheel flats (Banverket, 2002; UIC, 2001).

Appendix E: Appended Papers

Paper I

Analysis of Train Delay Information

B. Nyström & U. Kumar,
Luleå Railway Research Centre,
Luleå University of Technology,
SE-97187 Luleå, Sweden;
e-mails: Birre.Nystrom@mt.luth.se, Uday.Kumar@ce.luth.se.

Abstract

The iron ore railway in northern Scandinavia is a single track line that stretches from the Gulf of Bothnia to the Norwegian Sea. Mixed traffic on a single-track railway makes the traffic sensitive and prone to disturbances. To improve punctuality and efficiency in the traffic flow, the link between disturbances and their impacts on train delays needs to be analysed and understood more precisely than today.

This paper investigates the underlying causes of train delays for the iron ore transport section of the Swedish railway. The analysis presented in the paper is based on data pertinent to train delays, which are retrieved, classified and analysed in close cooperation with operation and maintenance personnel from both infrastructure and train operating companies. The focal point of our analysis is on discovering the sequence of events and the linkages therein, if any, leading to train delays. These event descriptions are subsequently compared to the records of concerned reporting systems of the Swedish railway administration. As an end result of the analysis and comparison with kept data, we are able to perceive a wider view of the causes for delays.

This is expected to provide us with a more relevant and reliable feedback, which should prove useful for prioritization of maintenance work.

Keywords: railway, delay, root cause, failure, database

Introduction and Background

During 1988, the Swedish State Railway (SJ) authority was restructured to enhance its competitiveness and make travel and transport by railway economically viable. The restructuring program divided the Swedish State Railway system into two major groups, namely organisations with responsibility for infrastructure management and train operators, i.e. organisations or companies with responsibility for trains operation. The train operators were expected to take the responsibility for transportation of goods and passengers in close cooperation with infrastructure managers. Today, there are about twenty train operating companies (TOCs) functioning in Sweden. The railway infrastructure is managed by 'Banverket', which is a public entity. In 1998, Banverket was restructured into two distinct divisions, purchaser or 'Service buyers' and contractors, or Outsourced service providers. For administrative purposes, Banverket is divided into five regions, each of which handles its own purchasing. In recent years, maintenance contracts have increasingly been awarded through open tender, and thus being subjected to market competition. Full competition is planned to be achieved by 2006.

In all, the Railway Inspectorate oversees the railway safety issues. The governmental body Rikstrafiken and the diverse county authorities provide funds to the train operating companies to run passenger services which would otherwise be unviable due to their non-profitability. There are several rolling stock maintainers, e.g. EuroMaint. The company Samtrafiken manages cooperation regarding tickets and timetables, making it possible to complete a journey on a single ticket.¹ The multitude of entities involved in the railways after its restructuring has made it considerably difficult to locate the responsible arm for problems encountered and to ascertain the course of action to be taken to rectify it. Such issues as how to classify the anomalies and how to achieve cooperation among the varying bodies in order to solve them are of utmost importance.

Along with the organisational complexities, there is also the element of 'Mixed Traffic'. The iron ore-mines in Gällivare and Kiruna deliver their goods to the harbours of Luleå (by the Baltic) and Narvik (by the Norwegian Sea), but the route is also used by trains carrying passengers or other goods e.g. timber, fresh fish. This mixed traffic on a single-track railway, with varying train speeds, makes the traffic sensitive to disturbances.

This paper is based on the project "Improved punctuality by effective maintenance" sponsored by Banverket and EU regional fund. In this project we are investigating ways of reducing delays through identification and elimination of critical factors behind the train delays. Both technical and organisational measures are to be proposed in this regard. Initially, failure events that have occurred and their recounting and classification in the databases are to be investigated.

In this paper, we investigate and analyze the failures and delays on the iron ore transport section of the Swedish railway. The focus of our analysis is on finding the sequence of events leading to train delays. Subsequently, we compare our findings related to the event descriptions to records in the relevant reporting systems of the Swedish railway administration for correctness, completeness and clarity. This contributes to provision of a broader understanding of the causes for delays.

The following section briefly explains some of the relevant keywords used in this paper, so there may be a common context of understanding between the readers and the authors.

Definitions and Concepts

Meaning of Delay and Punctuality

A delay in railway traffic is defined as a train lagging behind schedule. A train in Sweden is considered punctual if it is less than five minutes off schedule at a station (Banverket's definition). This definition does not apply to commuter trains. It should be noted that a train might suffer a number of delays along its route, yet arrive punctually to the final destination, due to a higher speed attained than planned for on certain parts of the journey.

The delays may be classified into two major categories: primary and secondary delays.

Primary and Secondary Delay Definitions

A primary delay is a direct outcome of the cause of delay, e.g. a turnout failure. If another train, in turn, is delayed as a result of meeting the primary delayed train, it suffers a secondary, or knock-on delay. Thus the resultant delay chain might proceed in several more steps.

The statistics that emerge from our analysed databases contain several types of delays. Train traffic control centre personnel establish the linkages between the primary delay and knock-on delays either automatically or manually during their operations. Delays are classified as follows:

- Primary reported extra delay, i.e. an extra delay to a train initiated directly by the cause, reported on a primary cause code, e.g. Track work.
- Secondary derived extra delay, i.e. an extra delay on one's own or another train, caused by a primary reported extra delay. This figure is calculated from the data obtained on primary reported extra delay.
- Secondary reported extra delay, i.e. an extra delay on own train, reported using a secondary delay code, e.g. Trains meet.²

Track availability may be defined as the time the track is de facto available divided by the planned available time. Thus, planned maintenance does not lower the availability, but unplanned maintenance work lowers the level of availability of the tracks.

The Five-minutes Limit and Slowly Emerging Delays

Detailed graphs of train traffic running on the iron ore lines were generated and inspected. A cursory examination of each train graph makes apparent some slowly emerging delays over long distances. These are the accumulation of the extra delays between any two stations that are below the five-minutes limit, yet which eventually sum up to a delay of well over five minutes.

Some of the data in the data base was found to be anomalous, or in an illegal format, at a significant proportion of departure/arrival events, and as such was deemed useless in computing slowly emerging delays, also rendering the aforementioned visual inspection more difficult.

Delay Assessment

Calculating delay severity is a significantly more complex process than a mere consideration of the number of trains involved and the aggregation of the time delayed. For example, a small delay might be acceptable if one is subject to a wait regardless of the delay for any subsequent connection, or goods need to be stored before latter shipment. Delay assessment is a complex issue that merits further independent and protracted discussion, which is beyond the scope of this paper. It should be pointed out that a punctuality measure for passenger trains should relate to passengers, not trains.³

In the next section we are providing a discussion about the components involved in the overall process of data collection, classification, and analysis.

Data Collection and Data Classification

This section is segmented into three components that constitute a base for further analysis of data:

- A brief description of the different groups interested in information and data concerning infrastructure, trains and traffic etc.

- The actual data collection mechanism from train traffic event generation to logging of data in the master computers at the traffic control centres.
- Finally, data bases where information concerning train delays and punctuality data are manually or automatically recorded based on the criteria of whether the data type is an inspection report or failure report.

Users of Information and Railway Databases

Collection of data is an expensive activity and as such needs to be economically justified. Studies⁴ show that, in general, 20-30% of total companies' expenses are used to re-discover what is already known. There is a need to measure asset performance in order to gain experiences that may be used for assessing current performance, future trends and benchmarking. We need usable data on expenses, weather, track characteristics (length, profile, radii, age, no of switches etc), track standard, failures, traffic load, accidents, delays and organizational characteristics.^{5,6} Field data is required for these objectives, along with a system for data collection and storage.

As data forms the basis for discussion and decisions regarding maintenance and investments in cross-functional improvement groups, it is important that the collected data is correct, complete and meaningful. The quality of data also needs to be made known to the users, so the users/participants may agree on which data to trust in order to reach a common understanding.

Some of the users are:

- Infrastructure managers
- Train Operating Companies
- Transport authority and Government bodies
- Customers organisations
- Research Organisations

There are approximately 200 data bases or data collection systems directly or indirectly related to railway systems in Sweden. Out of these, about 40 are directly related to infrastructure performance and train operations etc. According to personnel involved in preparing reports used for decision-making, data retrieval out of the myriad of databases is a cumbersome process, leaving little time for analysis.

Collection of Data: The Event Flow

The data path consists of train traffic event data generated directly from block sections of the tracks, being subsequently integrated into the TFÖR and MAPS systems at the Train Traffic Control Centre. The specific event chain is discussed below.

Train Traffic Event Data to the Train Traffic Control Centre

Train traffic event data is generated when a train entering a certain block section of the track electrically connects the two rails, and the relay associated with that block section is closed. In Figure 1, a part of single track line between two stations is shown. The three block sections between the stations indicate where the trains are, as well as the location of the main tracks and sidings of the stations.

It may take a few moments before messages on departure and arrival events reach the computer system of the train traffic control centre, resulting in a maximum of a few seconds deviation from true time. This might pose a problem when dealing with the dense traffic at heavily congested railway stations, but not in the scenario that is the scope of this investigation.

In data from the Train Traffic Control Centre dispatching system for northern Sweden (“Argus”), only arrival and departure times at stations were possible to obtain, not times at individual block sections.

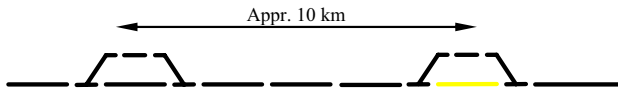


Figure 1. In this schematic figure of a single track railway, three block sections make up the distance between two stations.

The arrival and departure times here refer to track circuit closing – not the time of the train in front of the platform, as the track circuit may start somewhere else. To adjust for measurements performed a distance away from the platform; a certain number of seconds are added or subtracted. However, this figure is the same for all train types regardless of slowly or rapidly accelerating, so the approach sometimes leads to noting on-time trains as departing too early.⁷

Data from Train Traffic Control Centre: TFÖR and MAPS systems

A train entering a station’s block section triggers an automatically generated message giving indication of its position to the main master computer at the train traffic control centre as shown in Figure 2.

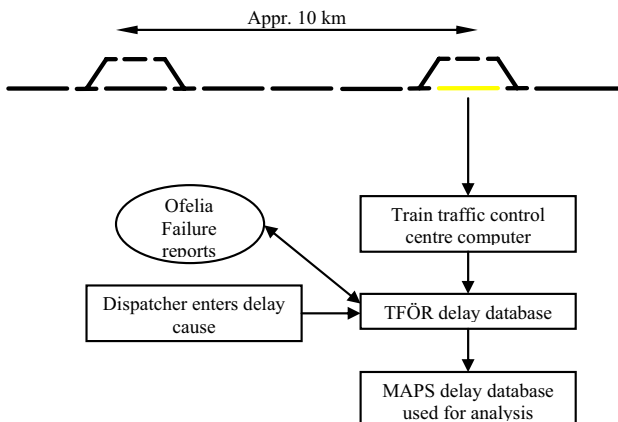


Figure 2. A train passing a station (coloured) triggers a message to be sent to the train traffic control centre computer system. This data is forwarded to delay databases.

The TFÖR (TågFÖRseningar, Train Delays) database obtains train operations reports from stations and stores delay information in the form of predetermined codes assigning probable causes for train delay. The train dispatcher must enter a "cause code" for any extra delay over five minutes. Extra delay means that the train under observation lagged at least five minutes more at a station than it did at the earlier station⁸. The MAPS (Multidimensional Analysis of Punctuality and Disturbances) database obtains data from TFÖR for analysis.

In the TFÖR (and Argus system), reasonably complete data was found only for Swedish stations, so data on train operations in Narvik was gathered manually by the iron ore transportation company MTAB for this study.

Train delay data may be slightly unreliable in some cases; however, positioning measurements are performed at the same locations every time. This makes the variation in transport time consistent and reliable. In order to get more correct arrival and departure times, track circuits could be moved, which is expensive. Alternatively, one could adjust times, based on a specific train's acceleration/deceleration data.

Having a five minute limit on recording delays means that slowly emerging delays must be treated in a different way. Lowering the limit would increase the load of encoding delay causes on train dispatchers and thus lower the information quality.

Classification of Data

Thus, the condition of the railway infrastructure is described in several systems e.g. Bessy containing inspection reports, Ofelia containing failure reports and Strix containing track regularity data. A brief overview is given below of the two systems:

Inspection Data: The Bessy System

The inspector carries out safety and maintenance inspections, who in turn gives each notification a priority, regarding how soon it must be remedied: Acute, Week, Month, Year, Others. When the first remedial action is taken, its date is registered in Bessy.⁹

Failure Data: The Ofelia System

When a failure is reported to the train traffic control centre the track operation leader calls the contractor's contact person, who in turn calls out his personnel. Even when immediately remedying a failure observed by them, maintenance workers have to report to the train traffic control centre.

Ofelia contains two computer sheets, one containing e.g. symptom, asset identification and failure-remedying personnel. Figure 3 shows the Ofelia fault time-line. The track operation leader decides on when to report further (i.e. call the contractor) and when to defer remedies.

oReported	oFurther reported	oDeferred	•Repair begun	oRemedied	oFinished
-----------	-------------------	-----------	---------------	-----------	-----------

Figure 3. An Ofelia report shows status of fault. Here, repair has started.

The other sheet contains detailed asset structure. Scrolling lists are used to fill in Asset type, Asset part, Component, Unit, as well as respective models. Here, also real fault, cause and

remedy are to be filled in and free text (or additional comments) may complement, see Figure 4.




Real fault	Cause	Remedial action
Short circuit 	Brake flake 	Cleaning 
<input type="text"/>	<input type="text"/>	<input type="text"/>

Figure 4. An Ofelia report allows real fault, cause and remedial action to be chosen from scroll-down-menus. Also free text may be entered. Here, the track circuit is corrupted by a conductive brake flake (that is, two rails become electrically connected) and the action taken is cleaning (i.e. removal of the flake).

By real fault Ofelia means what is remedied, e.g. short circuit (of track circuit), the cause being brake flake and the remedy cleaning.¹⁰

For example, if in the course of inspection an inspector comes across a loose bolt, the notification is entered into Bessy. If, instead, the incident is discovered by another railway employee, the fault is entered into Ofelia. The consequence may be that places with high density of employees may encounter a lot of (Ofelia) failures, whilst other places would receive more (Bessy) inspection notifications instead.

Not all Ofelia records contain entries that need to be regarded as faults. An item is considered faulty if it has lost the ability to perform a required function according to standard.¹¹ So the bolt in the above-mentioned example is questionable, in such that if one considers the bolt from the maintainer's viewpoint, it has lost its function, however, considering the turnout (TOC's viewpoint) of which it is a part of, the turnout has not lost its function and is thus not faulty. An example of an Ofelia record that is not a fault is a worn crossing. It is thus advisable to look for both Ofelia faults and Bessy Acute notifications.

TFÖR–Ofelia Linkage

A TFÖR report with delay caused by infrastructure failure is linked to the Ofelia report describing the reason for failure.¹² There is no existing linkage from TFÖR to the Bessy system.

From here on, we proceed on to an analysis of the possible causes of delay based on compiled data.

Analysis of Causes of Delay in the Swedish Railway System

In 2002, 77814 hours of train delays were reported¹³. In Figure 5, the delay time is distributed by causes. This picture aggregates delays for all trains on all tracks – however, it does not divulge any information about individual customers or individual parts of the network.

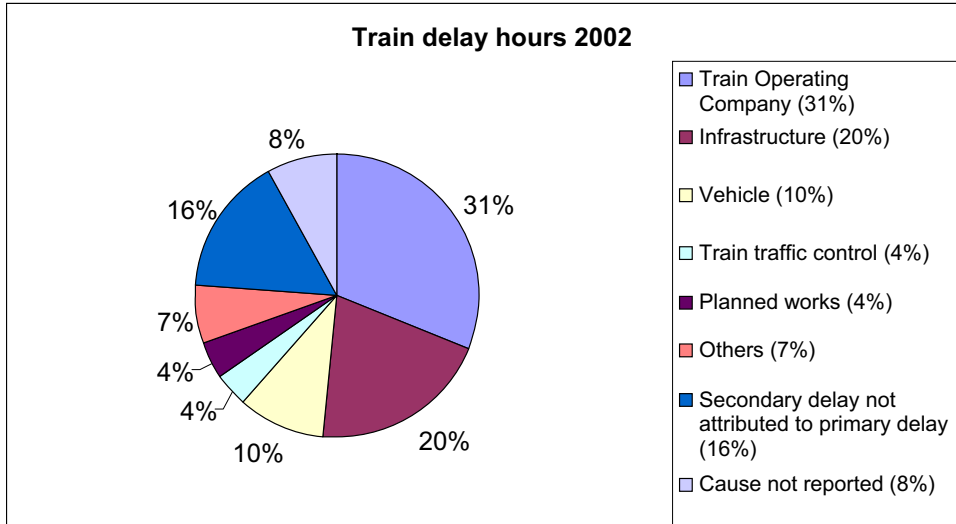


Figure 5. Train delay hours 2002 in Sweden, assigned to causes.

When we investigated infrastructure-related failures, we identified the top ten-failures regarding caused train-delay hours, see Table 1. It should be noted that the sum of all delays caused by elements such as overhead wires is added up in the second column of Table 1. The overhead wire-failures are relatively few in numbers (see column 1), yet cause a significant amount of delay-hours, as each failure results in a long individual delay, 2.0 h on average, as seen in column 5. The ordering imposed by the number of delay-hours attributed to causes are quite different from the one based on the number of occurred failures, where turnout is the most frequent. The leading cause for average delay hours is attributed to failures to feeder lines (on average 77.8 h of delays each time it happens) and overhead wires (2.0 h). Turnout, though most frequent in number of occurrences, does not feature among the top-ten reasons for delays with longest average delay time, since each failure on average results in a minuscule and non-critical delay.

Table 1. Top-ten number of total delay-hours on infrastructure (Ofelia, year 2002). The six most important causes are responsible for 50% of the infrastructure-related delays.

1. Asset	2. Number of failures	3. Delay attributed to asset failure (h)	4. Percentage of delay attributed to infrastructure	5. Average delay attributed to each asset failure (h)
Overhead wire	1121	2208	14%	2.0
Track	4746	1706	11%	0.4
Turnout	9866	1495	10%	0.2
Signal box and section block	5291	1458	9%	0.3
Positioning system	3669	640	4%	0.2
Feeder line	7	544	3%	77.8
Shunting equipment	204	245	2%	1.2
Centralised traffic control	909	221	1%	0.2
Level crossing	4666	190	1%	0.04
Embankment	327	156	1%	0.5

In Figure 5, secondary delays are included in the percentages, but it may be seen that secondary delays have not been attributed to primary delays, and hence make up 16% of the total causes. Depending on actual causes, the relative magnitudes of the causes presented may be overturned by these delays. So, including secondary delays in our statistics may give a truer picture of the delays, but we need to be aware of the share that has not been attributed to primary delay.

In the next section, we take a brief look at the aspects of delay time distribution.

Delay Time Distribution

We investigate the time distributions on the stretch of track between Kiruna and Narvik. The iron ore trains from Kiruna to Narvik are scheduled to cover the distance in about four hours varying by up to about 20 minutes in both directions.

A cumulative plot shows the probability of reaching the destination under a certain time. In Figure 6 we see that the 90%-percentile of transportation time from Kiruna to Narvik is about 6.5 h. To the iron ore logistics company MTAB, being able to maintain the circulation is more critical than being a few minutes late. Knowledge of transportation time distribution might facilitate train scheduling.

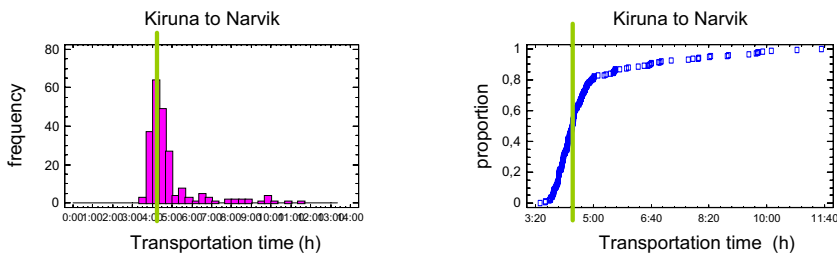


Figure 6. Distributive respectively cumulative plot of Iron ore train journey time from Kiruna to Narvik (03-03-01 – 23). The vertical line shows approximate scheduled time.

Delay Time Explanation

Apart from different traffic in different places and times, time needed for maintenance and repairs might be an explanation for the differences in delay time between different causes. However, MTTR (mean time to repair) and average delay show low correlation among the principal causes listed earlier. This may possibly be attributed to the factor of time to repair distributions being highly skewed. Figure 7 shows the distribution of time to repair for turnout failures. The graph shows the timeframe from maintenance personnel reaching the fault until it is remedied (Ofelia, year 2002, longer times than 10 h are not shown).

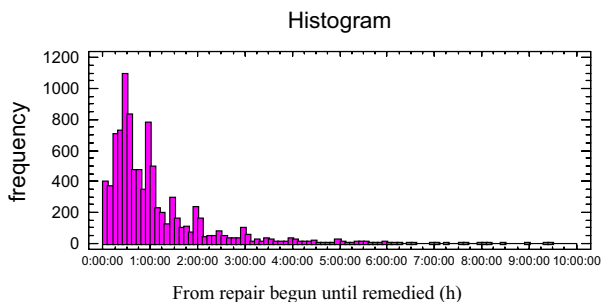


Figure 7. Distribution of turnout time to repair.

Figure 7 shows the time to repair, that is, from when maintenance workers start remedying the failure until its completion. But as the Ofelia and Bessy databases do not detail certain faults being train-stopping or not, we can not conclude that the track is non-available all the time. Nor is it possible to tell whether the tracks were functional despite the fault all the time.

To be able to model the impact of failure time on delay, the distribution of remediation times should be taken into account. The non-availability of track may lie at a point of time somewhere between the time to repair (Figure 7) and failure time, but it may also be overestimated by both and hence the availability of the track underestimated.

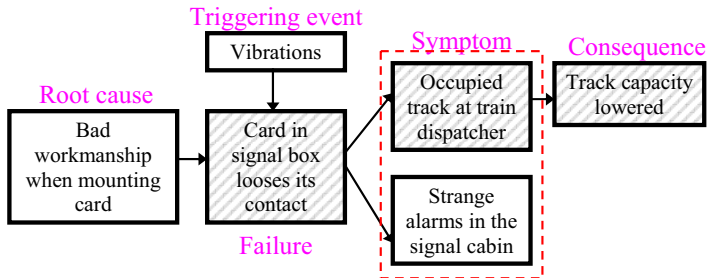
Ofelia focuses on component failures, but the system has its limitations. For example, mistakes in personal communication are not visible in it. In Ofelia, the real fault is not equated with the reported root cause. To illustrate, if a wagon catches fire on a marshalling yard, and the fire brigade turns off the electricity in order to extinguish the fire, hereby delayed trains are considered delayed due to powerless wires. As such, by “fault” the Ofelia system perceives what is remedied; in this case the remedy is to switch the electricity on again. One needs to be aware of this, when analyzing statistics related to the cycle of events.

Analysis of Train Delay Data and Information: Examples from the Field

As part of the investigative process, Failure analyses of infrastructure failures were performed. Personnel were interviewed the day after. The Ofelia system served as a guideline for selecting personnel to interview. The next two cases work as functional examples form a basis to discuss to what extent we are able to obtain a true picture of the failures.

Case I

Figure 8 illustrates a failure analysis of an occupied track event. The symptom in Ofelia was given as “Occupied track” when no trains were supposed to be there. As the reporter category was given as train dispatcher, it was understood that the symptom actually was that the track appeared to be occupied on the dispatcher’s screen. When personnel called in on the incident investigated, strange alarms in the signal cabin in a pattern familiar to the repair personnel were seen. Matched with earlier experiences, this indicated a lost contact. After adjustments, the alarms disappeared and the failure was remedied. The lost contact was probably due to bad workmanship when the card was mounted (this was done recently) and triggered by vibrations.



Ofelia FR00190842 03-03-02

Figure 8. Failure analysis of "Occupied track" events. Hatched boxes' information was found in Ofelia.

The hatched boxes of Figure 8 show what was possible to learn from Ofelia (in this case there was no traffic, so the lowered track capacity did not matter, which was seen in the Ofelia-TFÖR link showing no delays). We see that neither the root cause, nor the triggering event, was indicated in Ofelia. An inclusion of these might help when finding out possible weak points in the system, deserving further investigation. Ofelia data is very detailed regarding components, including asset, part of asset, component and unit (or, more correctly, demanding detailed fill-in), but not on describing the events leading to failure. Only one of the symptoms was recorded in the system. By describing the strange alarms symptom, fault-finding could be facilitated. This way, the new-comers to certain equipment may learn from other people's experiences. For example, usage of this data could form training modules that focus on important failure modes.

Case II

Figure 9 shows a failure analysis of an overhead wire tear-down. It began with the breaking of a carrier cable, which could have been due to carrier cable fatigue. As in the former case, we have two symptoms on display. But in this case, one symptom – the blown fuse – became known before the rapid part of the event chain began (starting with the triggering event). As we can see from the figure, contributing factors to the severity of the consequence (traffic was stopped for about sixteen hours) were the ability of the pantograph to grab broken overhead wire, overhead wire not stocked nearby and the storm. Possibly, the whole event could have been avoided if operations had made another decision, or if investments in carrier cable had been made.

In the Ofelia report, contact wire break is given as the real fault and aging as the cause. Also, the severity of the failure was indicated by writing that about 300 m were torn down. (This was the only record in this study where times were erroneously given (as three days after the event).)

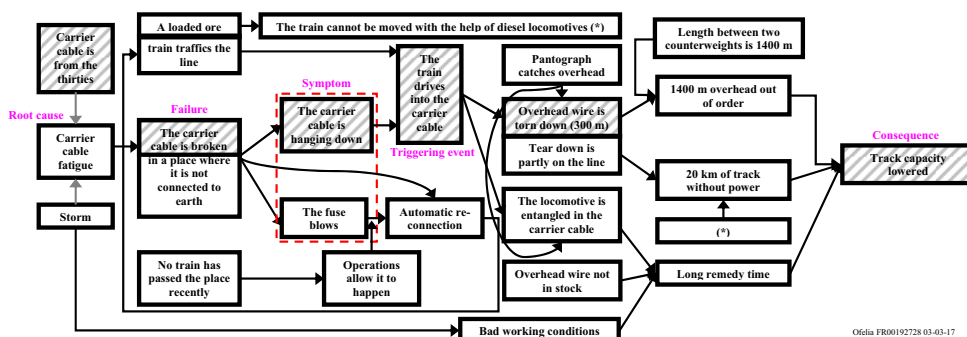


Figure 9. Overhead wire failure analysis showing failure chain, from root cause to remedial actions. Grey arrows represent uncertain casual relations.

Discussion

Ofelia focuses on detailed asset description, but the failure-event chain is not well described in it. Root causes, triggering events and symptoms are only sometimes recorded in the Ofelia database, as it does not facilitate recording of multiple causes.

Some of the Ofelia records indicate causes that are likely not to be accurate. Causes given merely to fill in blank data fields may be misleading, so the necessity of filling in all fields should not be over-emphasised. Hard facts, like item, time and participating personnel (although not everyone is given in the lists) are likely to be true. Repair difficulties in some cases have a large impact on remedial time, which is seen in the database. The reasons for difficulties are, however, not allowed to be described in Ofelia.

It is important to consider each customer's punctuality for the whole route, not delays along the route. The delay information are of great use, e.g. when computing knock-on relations.

Secondary derived delays give a measure of how much delay a certain area's trains cause themselves and trains in other areas, thus telling how well these trains are run.

A failure reporting system should facilitate optimal decision making by maintenance personnel. In the contact wire example of Figure 9 the wrong dates may be discovered by also seeing that train traffic is running.

Uncertain causal relationships will always be present in the scenarios. In order to get better understanding we need to ease access to data for hypothesis testing: for example, are the old contact wires more prone to failure? (Failures might be found in Ofelia, age in another system). Follow-up questions may be posed in context with the previous question: "or is it the amount of traffic that is decisive; how many trains have actually passed? (In TFÖR). What were the pantograph types on these trains?" The data-mining process may continue further along this line of thinking.

To increase in-field learning, all failure reports and actions previously taken on the object should be easily accessible when heading for a maintenance work. This information may help to avoid lap and patch as (non-working) actions taken before are seen, as well as facilitate the

mastery of the remediation personnel. The quality of reports is likely to improve, as the maintainers themselves will be using this data later on. Allowing maintainers to write the report themselves might improve quality further; however this imposes additional measures of educating maintainers and increasing interface user-friendliness.

We have ascertained that failure remediation is dependent on equipment design, information from train traffic control, maintenance personnel competence, stock policy, weather and other things. Each of these factors needs to be considered in order to bring about improvement.

Concluding Remarks

Absolute timing of train punctuality data is somewhat unreliable, but as data is gathered in a similar manner over time, it does not make a notable difference. Software adjustments could result in improved data quality.

Uncertain causal relationships are unavoidable; and ultimate root causes are not always to be found. What can be done is to facilitate integration and ease of use of data sources. This will also help in everyday maintenance work.

It is important that the data retrieval process is kept easy and that data from infrastructure, maintenance, traffic, and accounting are compatible. This can help involved parties to reach a common understanding, and also facilitate discussions on non-technical ways of improving punctuality. In addition, Ofelia and Bessy lacks a check box to be filled in to show whether a fault is train-stopping or not.

We conclude that making data retrieval easier should be a priority. In addition, data on asset, maintenance measures taken and their costs as well as production (i.e. traffic load) etc should be compatible.

Acknowledgements

The authors would like to thank Banverket Traffic and EU Regional Fund for sponsoring this study. We would also like to thank MTAB for providing data and other support during the course of the study.

References

-
- ¹ <http://www.tagoperatorerna.se>, <http://www.banverket.se>, <http://www.rikstrafiken.se>, <http://www.jarnvagsinspektionen.se>, <http://www.euromaint.se>, <http://www.samtrafiken.se> (03-06-22)).
- ² Dahl, G: Punktlighetsapplikationer i Essbase, 1997-08-21. (In Swedish.)
- ³ Ackermann, T: Die Bewertung der Pünktlichkeit als Qualitätsparameter im Schienenpersonenverkehr direkten Nutzenmessung. Series: Forschungsarbeiten des Verkehrswissenschaftlichen Instituts an der Universität Stuttgart / Verkehrswissenschaftliches Institut; Bericht 21. Dissertation. Stuttgart 1998. ISBN 3-922203-21-3. (In German.)
- ⁴ Boshyk, Y: Beyond knowledge management: how companies mobilize experience. From: Mastering information management: [complete MBA companion in information management]. Academic editors: D. A. Marchand, T. H. Davenport. Financial Times/Prentice Hall, London 2000. ISBN 0-273-64352-5.
- ⁵ Gustavsson, C. & Ådegren, A: Sökvägar för identifiering av kostnadsdrivare: En jämförelse mellan två bandelar inom Luleå Banområde. Examensarbete Högskoleingenjörsprogrammet 2001:71 HIP. Luleå University of Technology, Luleå 2001. (In Swedish.)
- ⁶ Banverket employees, Personal communication winter/spring 2002/2003.
- ⁷ Eriksson, Thomas, Quality – Punctuality, SJ Göteborg. Personal communication April 2003.
- ⁸ SJF 615.2, 1990-04-12. TFÖR instruktioner. A SJ (State Railway) Regulation. (In Swedish.)
- ⁹ BVH 807. A Banverket handbook. (In Swedish.)
- ¹⁰ BVH 808.11. A Banverket handbook. (In Swedish.)
- ¹¹ Svensk standard SS-EN 13306. Underhåll – Terminologi. Maintenance terminology. SIS, Swedish Standards Institute, Stockholm 2001.
- ¹² Banverket's intranet (03-02-05). (In Swedish.)
- ¹³ Nilsson, M et al: Incitament till ökad punktlighet på järnväg – förstudie. TFK – Institutet för transportforskning. Stockholm 2003. ISBN 91 88752 51 8. (In Swedish.)

Paper II

Improved railway punctuality by effective maintenance – a case study

B. Nyström & P. Söderholm

Luleå University of Technology, Sweden

ABSTRACT: A railway has many different requirements, such as safety and punctuality. The purpose of this study was to explore and describe how maintenance can contribute to improved punctuality. In an industry branch that is becoming more open, as is the case in Norway where this study was undertaken, it is important to have clear requirements and performance measures.

This study has focused on railway stakeholders and their requirements regarding punctuality, the measures employed, and the traceability between requirements, goals, and measurements. Linking the end customer's punctuality requirements to reliability of the track is illustrated.

The requirements flow is presented as well as a discussion about its pros and cons. The results of the study can be applied in order to compare different designs of maintenance to fulfil punctuality requirements. The information flow and the measures currently used are not entirely suitable concerning feedback and usefulness for managing availability and thus maintenance.

1 INTRODUCTION

The railway industry is in a state of transition, with new stakeholders emerging and old ones finding themselves in a new environment. This puts a heavier emphasis on the need for systematic ways to manage the requirements of stakeholders. Clearly specified performance measures might facilitate changes in division of labour and new actors entering the market.

The restructuring of the Norwegian railway began in 1989 with the partition of Norwegian State Railway (NSB) into divisions. Since 1996, Jernbaneverket (JBV) has been a separate body responsible for infrastructure management with NSB as a passenger train operating company (TOC). (www.jbv.no 04-04-01)

One important requirement of the railway is punctuality, which is influenced by several stakeholders and factors. The passenger or goods should be there when the train is bound for departure. The TOC's rolling stock should function and the driver should be on time. Infrastructure equipment such as turnouts should have high availability. A delayed train meeting another train at a single track might force the meeting train to wait at a siding, leading to even more delayed trains. The prioritizing between

different trains in such conflicts is managed by train traffic controllers who try to minimize the spread of disruptions. Hence, customer behaviour, TOC equipment and personnel, planning and scheduling of train traffic, and infrastructure maintenance are factors that influence punctuality. Factors outside stakeholder control, such as weather, also influence punctuality.

The purpose of this paper is to describe and analyse requirements related to punctuality concerning how they are stated, measured, and enforced by different stakeholders.

2 CONCEPTS AND DEFINITIONS

In this section, some important concepts and definitions of this paper are briefly presented.

A stakeholder is defined as: "an interested part having a right, share or claim in the system or in its possession of characteristics that meets that party's needs and/or expectations" (ISO/IEC STD 15288 2002). In this definition stakeholders include, but are not limited to, users, supporters, developers, producers, trainers, maintainers, disposers, acquirer and supplier organisations, regulatory bodies, and members of society. "Stakeholder requirements are expressed in terms of the needs, wants, desires, expect-

tations, and perceived constraints of identified stakeholders” (ISO/IEC STD 15288 2002).

In this paper punctuality is seen as a railway requirement. Punctuality is a characteristic of arrival and departure times at places important to a stakeholder. For further discussion about punctuality, see Nyström (2005).

Requirements Management is a systematic approach to elicit, organise, document, and manage both the initial and changing requirements of a system. A principal result of this work is the development of one or more requirement specifications that define and document the complete external behaviour of the system to be constructed. (Davis & Lefingwell 1996)

“Maintenance is the combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function” (EN 13306 2001). From the definition, it is clear that maintenance aims to upkeep the item (i.e. equipment, machine or system) in a functioning state, and to not make it perform over the original requirements.

3 METHOD

The organisations were investigated to facilitate the description and analysis of a chain of requirements. Hence, the main unit of investigation was the value chain with the sub units being requirements and measures related to punctuality. This is illustrated by following the requirements of the end customer for punctuality of a post package down to the availability requirements of a railway infrastructure, which is produced by maintenance. To allow for comparisons between countries, Norwegian Post was chosen as the freight customer.

Empirical data was collected by personal interviews, a focus group interview, and documentation. As a support to the case study, a literature study was also performed (Nyström 2005), focusing on investigating different perspectives and dimensions of punctuality. A focus group of personnel from the railway manager’s organisation, train traffic control, and TOCs, was used for filling in missing classifications as a preparation for the individual interviews.

The stakeholders are organisations, parts of organisations, or individuals, but the interviewees are persons. The investigators view the roles of people as stakeholders, e.g. the role as quality manager.

Figure 1 shows the main stakeholders in the value chain for delivery of a letter. The scope of the study is within the dashed line. The study is limited to availability measured as punctuality and maintenance, while operation and modification issues are not considered.

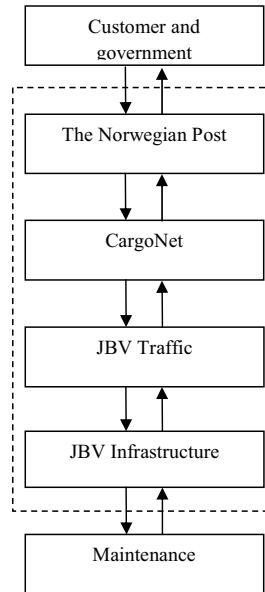


Figure 1. The value chain in transporting a letter by rail. In this model, requirements are symbolised by downward arrows and feedback, while measuring the requirement by upward arrows. As mentioned later, requirements also go upwards.

4 FINDINGS

In this section, findings from the investigated value chain in Figure 1 are presented. It begins with the Norwegian Post and continues down the supply chain, discussing requirements to and from the stakeholders.

4.1 *The Norwegian Post*

About half of the requirements on the Norwegian Post lack any method of measurement. Of these, several concern timeliness (how much a planned journey/transport is convenient to a stakeholder), not punctuality. Whether or not transports are delayed is followed up. Although the number of minutes delayed is measured, it is not possible to follow up, according to one informant. The Post uses the term ‘remaining post’ to denote post remaining, exceeding plan, at post sorting facilities or in transport.

Of the requirements lacking any method of measurement, several concern timeliness or information. The Post requires its business customers to allow post pick-up in a certain time span (or several time spans, if there is much to collect) for the Post’s cars to travel the shortest route in the shortest time with as much as possible. The performance regime of the Post has only one step for train punctuality, 30 min-

utes, compared to three for road transport (Transport manager & Senior consultant in Nyström 2004). Hence, the Post's marginal cost might not be covered for considerably longer delays, and the regime does not offer incentives to reduce the causes of long delays.

4.2 CargoNet

Some requirements placed on CargoNet are quantified, especially concerning punctuality where CargoNet is typically allowed about half an hour to unload after train arrival to the train station. Another requirement placed on CargoNet lacking any measurement is timeliness of train routes and informing the Post of which wagons contain Priority Class letters, so that these might be unloaded first (Quality manager in Nyström 2004).

CargoNet's requirement on track time for its trains also lacks any method of measurement. Furthermore, a clear definition of force majeure, i.e. when requirements are dropped, was not found. It is seen from a JBV data base that such causes amount to about 5-10% of total train delay.

4.3 JBV Traffic

Many requirements are aggregated to measure punctual trains. The word delay was used as being the antonym of punctual/punctuality by most informants, but this is not necessarily so from JBV's definitions, see Figure 2 (Jernbaneverket 2003, Traffic control analyst in Nyström 2004).

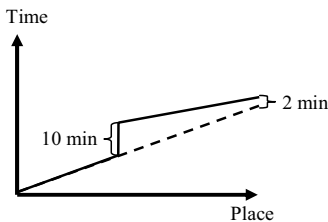


Figure 2. A train suffering a 10 min delay underway, but ending up punctual at end station.

Figure 2 shows a train path (dashed line is the scheduled path). Underway the train receives a 10 min delay (due to e.g. turnout failure), but arrives at end station only 2 min after its scheduled time, and is therefore considered punctual thanks to timetable slack. This phenomenon as well as when a train departs before the scheduled time counts as punctual if it arrives punctually to the end station, though it might conceal underway problems. By also measuring punctuality as length of delay at the end station,

one might indicate, e.g., to what extent train circulations are possible to withhold.

Today, the length of delay to the end station is somewhat indicated by measuring average end station delays, requiring it to be at most 2.5 min/train, and delays to 50 h/million train-km. These goals consider only delays due to infrastructure (Jernbaneverket 2004b). Trains might cause delays to other trains; therefore, having requirements on TOC-caused delays (with incentive scheme) might be advantageous.

4.4 JBV Infrastructure

For track maintenance, about ten goals (Jernbaneverket 2004a) are formulated as track quality (affects travel comfort and risk of derailment), numbers of failures per km track, delays, and expenses. Infrastructure equipment-caused delay depends on downtime due to each failure, traffic volume, and mix, not only the number of failures. A low number of failures might indicate low traffic volume, but not necessarily high track availability. It does not reveal much when, e.g., investigating prospective train plans. However, comparing one's number of failures with similar lines provides a motivating as well as realistic target to strive towards for rail infrastructure companies.

Infrastructure failures are classified as either track, contact wire, or signalling. The length of delay caused by failures of either track or signalling has roughly the same mean and standard deviation, while contact wire has many more rather longer delay times. (For freight in Norway 2003, the count, mean, and standard deviation of infrastructure-caused train delay minutes were extracted from JBV database: track (86, 36.4, 60.0), signalling (751, 26.2, 36.4), contact wire (223, 118.0, 153.6).) The number of contact wire failures probably indicates the delay quite inaccurately, and might be misleading when comparing, for example track sections. Failure causes in the JBV database were not classified entirely correctly, e.g. as maintenance-induced (e.g. due to track machine breakdown). Correct classification is important for action to be required from the appropriate stakeholder.

4.5 Summary of the value chain

The chain of requirement concerning punctuality of delivery is seen in Table 1 and Table 2. The bold outlined boxes contain requirements as found in the case study. The regular boxes contain the same requirements, but formulated as "the other way around". Table 1 shows the requirements formulated in an "availability" way, while Table 2 shows the corresponding requirements using a "failure-analogue" performance measure. As seen from Table 1 and Table 2, only one of the four stakeholders

(JBV infrastructure) formulates its requirement in the latter way. To give a complete view of requirements a network would be needed, as some requirements also go in other directions, e.g. from TOC to JBV infrastructure.

Table 1. A chain of requirements. Inside the bold outlined boxes are requirements as found in the case study. The other requirement is formulated by the authors.

Stakeholder	Requirement to stakeholder
Norwegian Post	Minimum 85% of Priority Class letters delivered over night (national average per calendar month).
CargoNet	90% of trains on time (< 1 min) to station (or customer's) and unloading has begun.
JBV – traffic division	90% of trains on time (< 5 min) to end station.
JBV – infrastructure division	99.999% of functioning trains pass 10 km of track without contact wire problems per year (approximate).

Table 2. A chain of requirements. Inside the bold outlined box is a requirement as found in the case study. The other requirements are formulated by the authors.

Stakeholder	Equivalent requirement
Norwegian Post	Maximum 15% Priority Class letters handed in on time still belong to 'remaining post' in the morning after (national average per calendar month).
CargoNet	Train departure, underway transport and unloading delay totals ≥ 1 min for at most 10% of trains.
JBV – traffic division	10% of trains ≥ 5 min late to end station.
JBV – infrastructure division	Contact wire < 0.2 failures giving operations difficulties per year and 10 km of track.

5 DISCUSSION AND CONCLUSIONS

From the interview with the focus group, no additional dimensions of punctuality were found other than those already known from Nyström (2005). However, the group studied did emphasise punctuality as being relative to an agreement (or expectation), not necessarily relative to a timetable. A multi-

tude of punctuality related expressions were used by the participants (no specific order): punctuality, in time agreed, according to decided timetable, right time, expected time, precision of delivery, satisfied-customer-index (broader term?), punctual, delay minutes (opposite?), right time-minutes, delay, quality of traffic, arriving in time, and in time. Punctuality as perceived by TOCs was more oriented towards customers' punctuality, while other organisations related to asset reliability and causes of unpunctuality. A positive measure, thus avoiding delay minutes or the like, was sought by several participants. Furthermore, the focus group interview revealed the themes, viz. information in advance concerning unpunctuality and formal agreements versus expectations, were important.

All stakeholders except the Post generally subscribe to the view "punctual or not" without explicitly considering the length of delay to the end station, though the costs incurred might vary with delay length. The meaning of delay was interpreted as not being punctual by most interviewees, even if its exact interpretation was somewhat unclear. The stakeholders' definitions of punctuality differ slightly. CargoNet has a smaller tolerance than JBV Traffic, but its time includes the time to shunt the wagons to unloading. CargoNet considers unloading time for trains in their punctuality statistics, while JBV considers trains to the end station. JBV infrastructure primarily measures number of failures.

The word regularity was interpreted differently by the informants, meaning the share of punctual post, punctual trains, trains de facto run, or trains ready for departure at scheduled time. The conclusion is that the ambiguous use of the word might be confusing in customer-supplier relations. Therefore, it is important to be clear when using the term regularity, by either specifying it further or using other terms: e.g. "arrival punctuality of trains" or "fraction not cancelled trains".

The important term 'cancelled train' means the same, a train not travelling its planned route, for TOCs as well as JBV, though some informants were uncertain of the exact meaning of the phrase. Cancelling a train and substituting it with (slow-going) buses, does not record as a delay. This might lead to not enough prioritising preventive maintenance of equipment that causes cancellations (i.e. with long time to repair). That there is only one step in the incentive scheme, might further contribute to this. Keeping records of, and setting goals for, time to repair might facilitate benchmarking and thus improvement of track availability. Declaring the share of trains cancelled when reporting punctuality should be encouraged. It is important to infrastructure maintainers to know from TOCs when they cancel trains (or will run early), as this leaves maintainers with the advantage of having longer, continuous track access.

The measure punctuality to the end station aggregates the effect of infrastructure failure, TOC failure, and the amount of slack time in the time table, making the individual effect difficult to distinguish. The underway delay measure is useful when investigating what happens due to infrastructure failures on a certain track section. It facilitates learning where to improve track availability, e.g. by shortening preventive maintenance intervals or striving for faster repair. By also investigating variation in transport time on underway distances, a discussion on slack issues important to punctuality might be facilitated. Investigation is today hampered by the limited access to reliable underway train delay data.

Few requirements were tied to penalties. To the Post: 85% of Priority Class Letters in time (and some other products always in time). To CargoNet: trains being not more than 30 min late. JBV infrastructure has to pay TOC's alternative transport if it carries out train hindering track works not agreed upon. Penalties rank the associated requirement higher and give it priority. Only unpunctuality caused by track works, not for example failures, are penalized. An incentive scheme for failure-induced delays would make the economic incitement chain more complete.

No stakeholder has a performance regime for "track and trace" (information on location, speed, etc., of goods, carrier, or vehicle), possibly due to difficulties in implementing it or satisfaction with current information. Prognoses of JBV infrastructure on failure remediation times are not entirely satisfying. When choosing maintenance method, the average time on track as well as its reliability should be considered. There is probably an improvement potential for maintainers to give prognoses of repair time, thereby allowing JBV Traffic, CargoNet, and the Post to give good estimations of arrival times.

Few requirements were quantified, even where they lend themselves to quantification. Quantifying requirements are especially important when dealing with requirements that are contradictory. An example is the wish to have a robust timetable as well as small allowances and margins. Another example is planned delays due to planned maintenance activities (has quantified goal), possibly opposing the wish to have high track utilisation (lacks quantified goal).

A requirement might be *proactive* or *reactive*. A proactive requirement indicates that measures should be taken to attain a certain desirable consequence (or avoid an undesirable). To have proactive requirements was considered important by the focus group, e.g., when not entirely so, was safety inspections made by the railway inspectorate. Who should act and how they should act to improve was also considered important.

There is no quantification of track uptime, needed to provide the link between track failure and train

delay and hence, estimate the consequences of prospective traffic changes.

The temporal dimension of requirements shows that most requirements are only measured using lag indicators (outcome measures), i.e. they tell the past (see Kumar & Ellingsen 2000), making analysis of prospective actions more difficult. A requirement for the size of the maintenance backlog would be proactive, as it indicates to keep the track in good condition (i.e. high probability of functioning) and how much track time maintenance will need to occupy the track, thus facilitating the planning of traffic and maintenance, and higher punctuality.

ACKNOWLEDGEMENTS

The authors would like to thank the involved organisations for making this study possible. We would also like to thank Professor Jørn Vatn, NTNU Trondheim, Norway, for valuable comments on drafts of this paper. The authors are also most grateful for the financial support from Banverket, the Swedish railway infrastructure manager.

REFERENCES

- Davis, A.M. & Leffingwell, D. 1996. Requirements Management in Medical Device Development. *Medical Device & Diagnostic Industry Magazine*, 18, 100-116.
- EN 13306. 2001. *Maintenance terminology*. Stockholm: Swedish standards institute.
- ISO/IEC STD 15288. 2002. *Systems Engineering – System Life Cycle Processes*. ISO/IEC.
- Jernbaneverket. 2003. *Punktlighetsrapport*. (In Norwegian.)
- Jernbaneverket. 2004a. *Vedlikeholdshåndbok Rev 0.1 1B-Ve 29.01.04* (In Norwegian.)
- Jernbaneverket. 2004b. *Virksomhetsplan 2004-2007. Infrastruktur RØ. "Hoved-, Gjøvik- og Gardemobanen"* 08.03.2004. (In Norwegian.)
- Kumar, U. & Ellingsen, H. P. 2000. Design and development of maintenance performance indicators for the Norwegian oil and gas industry. In *The 15th European Maintenance Congress: Euromaintenance 2000; Proc., Gothenburg, Sweden, March 2000*.
- Nyström, B. 2004. *Improved railway punctuality by maintenance – a case study*. Project report. Luleå University of Technology. (Unpubl.)
- Nyström, B. 2005. *Punctuality and railway maintenance*. Licentiate thesis. Luleå University of Technology. (In press.)

Paper III

Delay analysis of a freight train – an improvement case study from a steel company

BIRRE NYSTRÖM

Division of Operation and Maintenance Engineering

Luleå University of Technology

SE-971 87 Luleå

Sweden

birre.nystrom@ltu.se

Abstract

This paper proposes a procedure for the systematic definition of actions to improve the punctuality of trains, giving the greatest possible improvement. The transports of a company producing steel slabs in one factory and rolling them in another are investigated. Reliable transports between the two plants are crucial to market competitiveness. The transport chain for the steel slabs is described and the journey by train is investigated regarding the punctuality, transportation times and causes for any lack of punctuality. Ways to improve the punctuality, e.g. the more proactive use of track-side detectors, and the pros and cons of the procedure employed are discussed.

Keywords: punctuality; delays; failures; railway infrastructure; maintenance; freight trains

1. Introduction

The railway industry is in a state of transition, with new stakeholders emerging and the old ones having to adjust to the new environment. The restructuring of the Swedish railway system began in 1989 with the partition of the Swedish State Railways (SJ®) into the infrastructure manager Banverket® and separate train operating companies (TOCs) for passengers and freight.

Transport customers are requiring ever-shorter transport times and higher punctuality. Therefore, the railway industry is striving to improve performance by increasing train lengths and axle loads and introducing measures to ensure train punctuality. Punctuality is dependent on and influenced by several stakeholders and factors. The goods have to be in place when the train is bound for departure, the engine and wagons must function and the driver is to be available on time. If a delayed train meets another train on a single track, one of the trains is forced to wait at a siding, possibly leading to even more delays. Train traffic controllers try to minimize the spread of delays by prioritizing between different trains. The railway infrastructure manager has to provide a highly reliable infrastructure. Therefore, freight customer behaviour, TOC equipment and personnel, the planning and scheduling of train traffic, and infrastructure maintenance all influence punctuality. Another significant factor is the weather, e.g. when it causes slippery rails due to moisture and leaves.

The purpose of this paper is twofold. Firstly, it describes and analyses circumstances affecting the punctuality of steel freight shuttles and gives proposals for improvement. Secondly, it discusses the feasibility of the procedure employed. This paper is a revised version of Nyström (2005:1).

2. Concepts and definitions

In this section, some important concepts and definitions of this paper are briefly presented; see Nyström (2005:2) for a more extensive discussion.

Punctuality is defined as the extent to which the arrival and departure correspond to the times agreed upon by the stakeholders, here the steel company SSAB®, the TOC Green Cargo® and the infrastructure manager, Banverket.

Delay is understood as a negative deviation from the timetable, i.e. the train is lagging behind schedule. The term *delay minutes* denotes the length of the delay at a specific place. A train traversing a specific distance, e.g. between two stations, can suffer an *extra delay*, e.g. a train lagging behind schedule by 7 minutes when entering the distance and 17 when leaving it has suffered an extra delay of 10 minutes.

A *train number* is a reserved train path in the timetable.

3. Method

The study was performed as a single-case study (Yin 2003). Empirical data was collected using interviews, documentation and archival records. Statistics related to punctuality played an important role. The procedure proposed here has the steps M1–M7, and goes from symptoms to causes.

- M1. Choose a train to investigate.
- M2. Establish the relationship between the unpunctuality cost and the unpunctuality.
- M3. Establish the relationship between the punctuality and the deviations.
- M4. Establish the relationship between the deviations & the timetable and the deviations & the availability.
- M5. Establish the relationship between low availability and its causes.
- M6. Establish the improvement potential.
- M7. Seek remedies with large improvement potentials.

This paper primarily discusses steps M1–M4, and touches on M5–M7.

3.1. A few aspects of the problem

Rieckenberg (2004) noticed the low average speed, 22-23 km/h, of a system train from the Ruhr area to Finland, pointing to, e.g., long stops at a Swedish shunting yard and borders as contributing factors. Granström and Söderholm (2005) illustrated and

discussed the necessity of suspicion when searching for the root cause of railway failures. Harris and Ramsey (1994) employed network simulation to assess the effect of different infrastructure failures on the London underground, though they did not investigate the probability of different types of failure. Higgins *et al.* (1995) carried out simulations by altering the probabilities of different delay causes which were related to the train, track or terminal, including scheduled stops. The probability of delay may be specified as, e.g., the probability of delay per train-km. Moreover, Kaas (2000) calibrated a simulation model using the Danish database on train delays between stations. On the basis of these works, it can be concluded that to improve punctuality, an overall understanding of the transport, the probabilities of failure, the effects and causes of failure, and computations is needed.

3.2. Databases used

In this study, three of Banverket's databases were used. The first database (TFÖR) registers the train traffic performed, including the causes of delay, though mostly the symptoms are registered. The second database (Ofelia) contains records of infrastructure failure. TFÖR and Ofelia recordings differ slightly in some instances, e.g. the place of delay may be given as one or two stations apart. TFÖR data is arranged according to timetable adjustments; e.g. timetable period T04.1 covers the half-year period from June 13, 2004 – Jan 8, 2005. MAPS is a tool for retrieving TFÖR data. The third database was the railway infrastructure asset register (BIS).

4. The case-study

Findings from the steel freight shuttle delay analysis are presented here. Firstly, a presentation of the transport is given, followed by a description of the transport preparations in Luleå (including loading the wagons), the transport en route, and the delivery in Borlänge, before the empty wagons are returned to Luleå, see Figure 1.

SSAB produces steel slabs at its Luleå steel factory and performs rolling at its Borlänge steel mill. In Luleå, the carbon fraction of the crude steel is reduced and alloying elements are added to charges of 100 tonnes. The liquid steel is then cast cold in a continuous process, with slabs of approximately 20 tonnes being cut. In Borlänge, some 1000 km away, the slabs are rolled according to the customer's specifications. Slab wagons travel loaded southbound to Borlänge and unloaded northbound, whereas dedicated coil wagons travel empty to Borlänge and then loaded with coils northbound for coating at the Luleå factory.

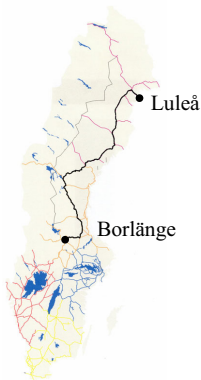


Figure 1. The steel shuttle traverses a route of approximately 1000 km between the northern coastal city of Luleå and the town of Borlänge in central Sweden. Its route may be temporarily changed due to track work.

4.1. In Luleå

The wagons are loaded with the ordered slabs, from indoor and outdoor storage. One tries to come as close as possible to the maximum allowed train weight.

When the train is registered in the computer as loaded, SSAB employees drive it to Luleå Railway Station. Here, Green Cargo visually inspects the wagons, which includes ensuring that the slabs are properly positioned, checking for damage, e.g. to flaps, listening for wheel damage and conducting a brake test. However, some types of wheel damage are not discovered. After the inspection, the train is reported as ready for departure. The trains are scheduled to leave from Luleå Station at the times 4.25, 8.15, 11.57 and 21.00 (for train numbers 9100, 9102, 9104, and 9106, respectively). The scheduled departures from SSAB to the railway station in Luleå vary between 2 and 3.5 hours before the departure for Borlänge. The journey between Luleå and Borlänge requires about 16 hours each way.

4.2. In Borlänge

As rolling at the Borlänge mill predominantly uses steel slabs from its stock and some 20% goes directly from the train to production, a buffer for variations of the train arrival time is available. A small delay (roughly 30 minutes) is considered normal, though when a delay approaches 5-8 hours, the current production run may be shortened or production orders cancelled.

4.3. M1: Choose a train to investigate

Choosing a train to investigate is the first step of the proposed procedure. The steel freight shuttle, which travels back and forth between two fixed places, is heavy

(approximately 100 tonnes/wagon, equating with a 25-tonne axle load and totalling about 2,400 tonnes) and short (500-600 metres), and should represent a rather common type of transport. It should be possible to generalize the results achieved to similar kinds of transport.

The steel shuttle has been running since the early 1980s. Nowadays, three or four trains run daily in each direction. The train is turned in Boden, a major railway hub some 30 km from Luleå. Here, the train number changes between an even and an odd number. Train rerouting due to planned track work causes a change of the train number in a non-systematic fashion historically. Consequently, identifying the trains of interest is not as simple as might be assumed. However, the MAPS database lists the steel shuttle numbers for most of the recent timetables.

4.4. M2: Relation between the unpunctuality cost and the unpunctuality

Calculating the cost of unpunctuality (step M2) means giving the loss as a function of the delay. From the point of view of SSAB production, long delays (more than some 5 hours) are costly, as production runs have to be shortened or annulled. However, information on this is missing. One should also be aware of the fact that drastically improved punctuality may enable SSAB to transport a greater share of slabs directly from the train to production, without passing through the stock, thus minimising the handling as well as the capital tied up.

4.5. M3: Punctuality – delay relationship

Departing ahead of or on schedule correlates with arriving on time, although trains departing on time may also be subject to en-route delays. The data shows that departing 5 minutes or more behind schedule lowers the probability of arriving on schedule from approximately 40% to 30%.

4.6. M3: Localisation of delay causes

The sum of extra delays for a given journey for a steel shuttle should equal the delay at the end station. However, as extra delays are registered only if they reach at least five minutes between consecutive stations and as the train might go faster than scheduled, exact correspondence is not to be anticipated. The correlation between the sum of extra delays (from Boden Central to Borlänge) and the arrival delay at Borlänge is about 0.94 for all the steel trains combined (T02.2). The sum of extra delays mainly explains the arrival delay, but this sum explains the arrival delay to a lesser extent when studying train numbers separately.

The top five places where delays occur most often are quite stable over the years, though the values vary.

4.7. M3: Arrival punctuality

The punctuality of arrival at Borlänge of all the steel shuttles over 2.5 years is given in Figure 2, divided into five consecutive timetable periods. Negative deviation means that the train arrives after the time scheduled in the timetable, positive deviation means arrival ahead of schedule. (The boxes in the figures contain 50% of the number of values, the median is denoted by the vertical line, and the left whisker is drawn from the lower quartile to the smallest point within 1.5 interquartile ranges from the lower quartile.) The arrival in Luleå may be considered slightly better than the arrival in Borlänge, in view of the smaller average of deviation from the scheduled arrival time.

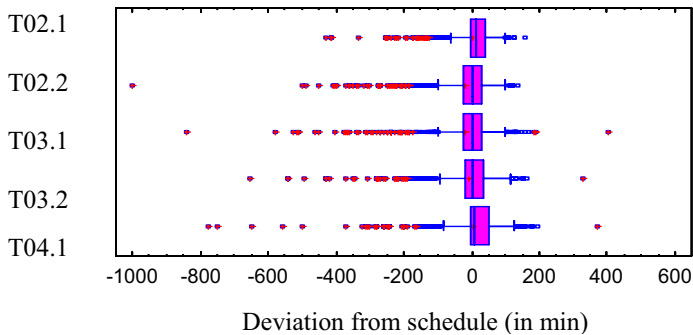


Figure 2. Deviation (in minutes) from the scheduled arrival time at Borlänge for steel shuttles during five consecutive timetable periods, denoted T02.1-T04.1 (corresponding to 2.5 years ending on January 8, 2005). A negative deviation means a train lagging behind schedule.

4.8. M4: Deviation from the timetable

In Figure 3, the relative deviation (in min/km) is shown for one train (9107, with the other trains being comparable). Notable here is the large variation shown at Tväråbäck (i.e. the time from the departure from Hällnäs to the departure from Tväråbäck) and Vallsta-Bolnäs. From Karsjö, trains often depart some 20-30 minutes ahead of schedule, due to planned but non-appearing oncoming trains. However, the non-appearance of oncoming trains is not coded in the TFÖR database, although it represents a deviation from the timetable.

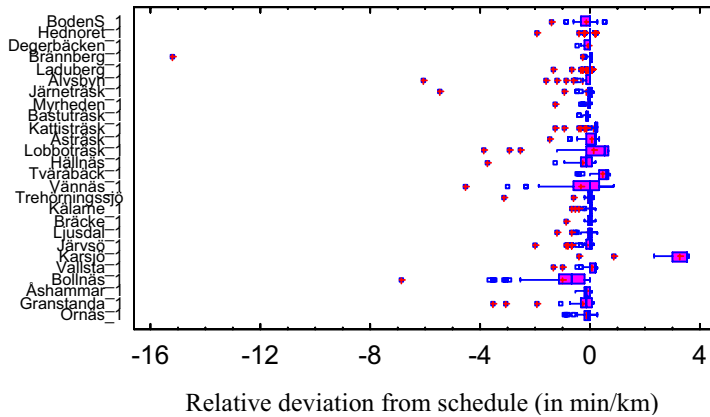


Figure 3. Deviations from the scheduled transport time (in min/km) as departure from one station to departure from the next. Here, a southbound train is shown (9107, T04.1).

It should be noted that the times in Figure 3 include large positive deviations from the scheduled times, large negative deviations (delays caused by, e.g., infrastructure failure where only an extra delay of 5 minutes or more is assigned a cause) and small deviations, positive or negative. These deviations may be caused by, e.g., minor rolling stock problems, varying climate conditions, low voltage and bad track. Furthermore, driving styles vary between steel shuttle drivers, e.g. concerning settings and decisions as to whether or not to run always at speeds close to ATC (Automatic Train Control) intervention. Furthermore, the variation may also depend on whether or not the driver looks out for animals or pedestrians. The latter problem may be diminished by, e.g., removing pedestrian crossings close to road crossings. How much variation is due to such factors remains to be established, as filtering out trains and stations where no cause of delay has been recorded is not always straightforward. However, excluding extra delays of 5 minutes or more leaves the ranking between stations with respect to variation practically unaltered. Such filtering should show how hard the distance is to drive, rather than the effect of failures. The distribution of the transport time, as in Figure 3, may be used when timetabling.

The shortest possible transport time could be calculated by using the shortest time for each distance between stations. When doing so, an approximately 200-minute shorter transport time at least is possible (9107, T04.1). However, one should be aware of the fact that about 40 minutes of this is due to certain drivers not using their right to a meal break at Ljusdal. The stretches Hällnäs-Tväråbäck and Tväråbäck-Vännäs have among the largest deviations per km. The Hällnäs-Tväråbäck deviations are explained by non-appearing oncoming trains. However, the data for Vännäs is manually reported and a few minutes incorrectly reported on this short distance may lower the calculated deviation per km considerably. However, the absence of unbelievable values indicates that the train-registering dispatchers are careful.

In most cases, the variation between train numbers is larger than that between timetable periods, compare Figure 2 and Figure 4.

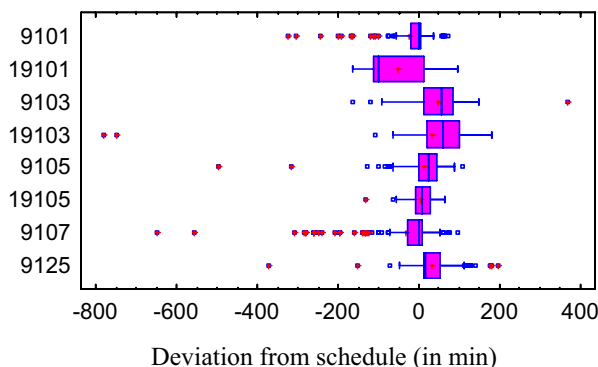


Figure 4. Deviation (in minutes) from the scheduled arrival time at Borlänge for different train numbers (T04.1).

The first few kilometres, from Luleå to Gammelstad, but not the return stretch, have a large variation in the travelling time, possibly due to the heavy trains, with slightly different weights and wagon mixes (new and old), as well as failures. Damaged wheels may be discovered by a track-side detector and force trains to stop at the Gammelstad siding. Informed by the train traffic control as to which wheel axle(s) have caused the alarm, the driver then walks along the train to inspect. The driver then informs the train traffic control if the wheel is considered faulty and, if so, he must call for workshop personnel for shunting assistance.

4.9. M5: Secondary delays

Figure 5 shows the delay causes 'oncoming train', 'overtaking train' and 'train ahead', i.e. delays caused by other trains lowering the availability of the track to northbound steel trains. Such delays occur predominantly at Borlänge, Långsele and Tväråbäck. In general, delays occur there and in Boden. Train 9116 suffers numerous delays in Bastuträsk, mostly due to oncoming trains.

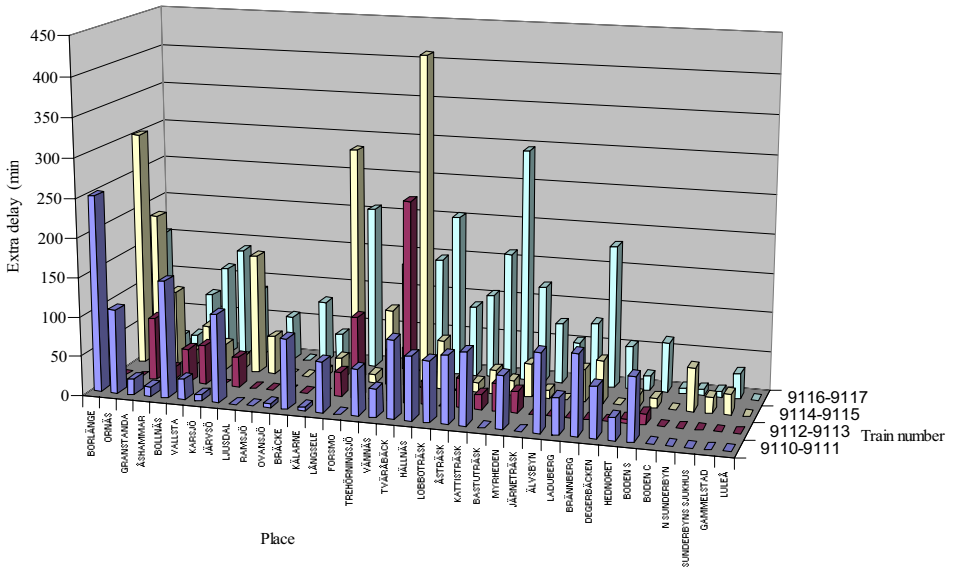


Figure 5. Delays caused by other trains to steel shuttles from Borlänge to Luleå (T02.1-T02.2).

4.10. M5: Damaged wheel

Statistics for the TFÖR delay code ‘detector alarm damaged wheel’ concern the track-side detector indicating damaged wheels during winter, i.e. November until February. The code ‘extra shunting’ was probably used interchangeably with the former code. ‘Detector alarm damaged wheel’ was found exclusively on trains from Luleå, i.e. loaded trains. The use of the ‘extra shunting’ code for trains approaching Luleå may conceal some damaged wheels, though the code probably describes the consequence of alarms from the hot box and faulty brake detector, located between Boden and Gammelstad. When a train with damaged wheels has passed, the track is to be inspected manually, thus delaying other traffic. Approximately 20 trains a year are coded as having damaged wheels.

4.11. M5: Overall identification of delay causes

The identification of causes of low availability and the identification of failures suffer from the fact that the root causes of delays are often not easy to identify uniquely (Nyström and Kumar 2003). Considering the infrastructure equipment, the investigation here is limited to a first glance at reliability in order to grasp the potential for improvement. Figure 6 shows the failure rates for some asset types.

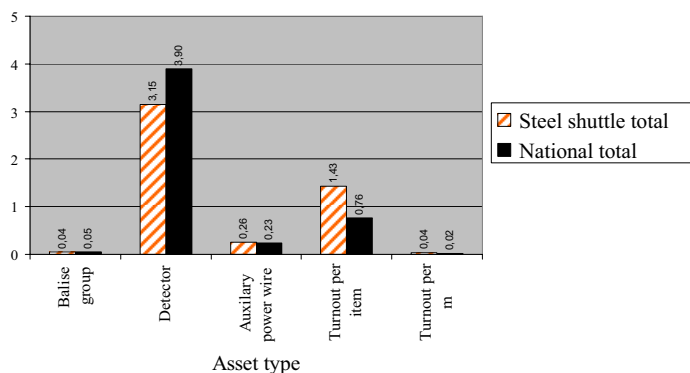


Figure 6. Failure frequency of certain infrastructure assets on the steel shuttle's line compared to the Swedish national average. The three leftmost asset types are given here as yearly failures per item, and the turnout as the yearly failures both per turnout and per metre of turnout. (The number of failures is for the year 2002, and the asset quantities are from 2005.)

4.12. M6: Establish the improvement potential

Due to the limited time span (the year 2002) the failure rates of Figure 6 must be seen only as indicative. Balises have about the same failure frequency on the steel shuttle's line as the national average. The difference from the national average for the detector failure rate may be explained by misclassifications, referring to the earlier discussion on damaged wheels. Which measure is appropriate for a certain type of equipment is sometimes not clear-cut, e.g. using the number of turnouts or using the turnout length. The latter may be more appropriate, as longer turnouts should have a higher failure rate, due to their greater number of components and higher train speeds. The turnout failure rate is about twice the national average. There may be room for improvement in this connection.

4.13. M7: Seek remedies with large improvement potentials

Remedies are discussed in the next section.

5. Discussion on the studied train

A small delay may be caused by varying conditions, while longer delays may be caused by malfunctioning equipment, e.g. turnouts. A possible cause of very long delays is bad weather. The different causes of delay have to be dealt with separately. Variation in the running time between stations makes scheduled times longer and also hinders others from using the track, even if such variation does not count as delays. Small variations may be due to minor engine problems, driving styles, a slight lowering of the overhead voltage because of simultaneous trains, the condition of the track, and the weather. The

on-going change-over to new wagons will probably result in fewer wagon failures as well as less variation of the travelling time.

SSAB uses the wagon circulation time as an indicator, with a goal of 3.5 days, which is a good measure of the capital employed for transportation. Green Cargo presents the percentage of trains on time plus tolerances to measure performance.

The statistics show train 9117 (scheduled arrival at Luleå Station at 13.15) to have the worst punctuality if the punctuality is calculated as the sum of negative deviations. However, this train is indicated to have the best punctuality, if the punctuality is regarded as the sum of the deviation minutes from the schedule. Which measure to choose depends on how one assesses arriving ahead of schedule. Fixed work shifts and the fact that iron ore trains also arrive at Luleå Station and SSAB probably advocate the former view of punctuality.

The steel shuttle's punctuality may in some cases be improved using technically simple means. An example is plainly to prevent trains or wagons not reported as ready from moving, in order to avoid problems later on.

Instead of discovering faulty wheels with the wheel damage detector when the shuttle is heading loaded from Luleå, it would be better to do so when the train is unloaded, i.e. when approaching Luleå. All the measurement data from the detector is registered in a database, but real-time access to this data requires awkward reprogramming of a dedicated computer. The detector system's built-in 'ratio high' alarm mode is probably worth employing next winter to allow the TOC to take wagons with faulty wheels out of service.

Better communication, e.g. better mobile telephone coverage and mobile phones not attached to the engine, would make the reporting of, e.g., damaged wheels swifter and allow the prompt deployment of assistance.

When the train is back in Luleå, de-icing is important to allow inspections of the wagons and hence to recognise faults. At present, trains may not be de-iced if, for example, they arrive close to a change of shift. The driver could be allowed already in Boden, when checking the train's braking, to decide whether to de-ice or not, in order to speed up the procedure.

The availability of infrastructure equipment differs between the route studied and the national average, but it remains to be investigated whether this reflects a hidden improvement potential or is entirely due to, e.g., a higher traffic load.

6. Discussion on the proposed procedure

Throughout this paper, different views of (un)punctuality have been illustrated, showing that care is needed when comparing punctuality statistics. A diverse and to some extent contradictory vocabulary, e.g. MAPS' term 'negative delay' (surely a delay is always

negative), makes it more difficult to discuss aggregated data, and one should definitely use a common terminology when making new software, reports etc.

Regarding the proposed procedure, the first step, identifying the train to investigate, is hindered somewhat by the current MAPS database layout, which does not couple exactly one train number to each train. On a macroscopic scale, this may result in misleading calculations of quality indices, e.g. by counting train numbers instead of trains.

The second step, calculating the unpunctuality cost as a function of the unpunctuality, needs such data on steel production runs as is not recorded by SSAB today.

The third step considers the relation between the unpunctuality and the deviations. Using extra delays as prioritising grounds for removing the causes of delay is relevant when trains in general are considered, but for individual train numbers this may not be relevant. The departure punctuality is good to use, as it affects the arrival punctuality. SSAB uses the circulation time for wagons as an indicator. This is a good measure of tied-up capital and punctuality, and it is important to calculate it.

The fourth step is to identify the impact from the timetable on the deviations. However, extra time (slack) is built into the timetable, e.g. on the southernmost part of the route regarding southbound trains. Therefore, it is a good idea to study the deviations from the scheduled travelling times between stations. The "TA" analysis tool of the infrastructure manager Banverket could be used to check the stability of the timetable, and its range of application should therefore be extended beyond the past three months of data.

The fifth and subsequent steps seek to eliminate the causes of low availability. Delay attribution codes keyed in wrongly, and inconsistent coding practices, as was exemplified in connection with the detector alarm, make the identification of causes more difficult and unreliable. One possible way to improve the uniformity is now proposed. When a cause of delay is fed into the database, it should become accessible to the persons at Banverket and the TOCs, respectively, who are responsible for the train, as well as to the punctuality analysts. When the person responsible for the train encounters a doubtful delay attribution for his/her train, this should be discussed with the analyst. The persons responsible for the train and a rule-based filter would save the analyst from information overload concerning the delay causes. This filter could detect possible anomalies, e.g. the use of a code reserved for passenger trains for a freight train or the attribution of delay to track work when the train has been late coming to the place of work. The initial set of rules would probably be best designed by a national group of analysts, who could discuss the delay attribution codes and their meanings. The persons responsible for the train would need tools to follow their train using near-real-time data, and would need to be able to present data for longer periods. It is proposed that the latter should be achieved in a standardised way and that following trains in order to establish causes of delay should be the day-to-day work of the person responsible for the train. As such personnel would be familiar with the dispatcher's screen, rerunning delay events in a simulator may facilitate the breaking-down of events to find the actual causes of delay.

Aspects to consider when performing a deeper comparison are the impacts on reliability from the asset model, modifications, traffic, age and maintenance. The performance of corrective maintenance, e.g. the time to repair, also needs to be investigated, in order to give a more complete picture of the availability.

Acknowledgements

The author would like to thank SSAB, Green Cargo and Banverket (the Swedish National Railway Administration) for making this study possible, as well as Per-Anders Akersten, Luleå University of Technology, for valuable comments on earlier versions of the manuscript. The author is also grateful for the financial support from Banverket and the EU Regional Fund.

References

- Granström, R. and Söderholm, P., Punctuality measurements effect on the maintenance process – a study of train delay statistics for the Swedish railway, in *Railway Engineering-2005*, ed M. C. Forde, Engineering Technics Press, Edinburgh, 2005 (CD).
- Harris, N. G. and Ramsey, J. B. H., Assessing the effects of railway infrastructure failure. *J. Operational Research Society*, 1994, **45**(6), 635-640.
- Higgins, A., Kozan, E. and Ferreira, L., Modelling delay risks associated with train schedules. *Transportation Planning and Technology*, 1995, **19**, 89-108.
- Kaas, A. H., Punctuality model for railways. *Advances in Transport*, 2000, **7**, *Computers in Railways VII*, 853-860.
- Nyström, B., Improved punctuality – a freight train study, in *Railway Engineering-2005*, ed M. C. Forde, Engineering Technics Press, Edinburgh, 2005 (CD).
- Nyström, B., Punctuality and railway maintenance. Licentiate thesis 2005:41, Luleå University of Technology, 2005.
- Nyström, B. and Kumar, U., Analysis of train delay information, in *Proceedings of the World Congress on Railway Research*, Edinburgh, 2003, 253-261.
- Rieckenberg, T., Telematik im Schienengüterverkehr – ein konzeptionell-technischer Beitrag zur Steigerung der Sicherheit und Effektivität. PhD thesis, Technische Universität Berlin, 2004 (in German).
- Yin, R. K., *Case study research – design and methods*. 3. ed, 2003 (Sage: Thousand Oaks, Calif.).

Paper IV

The Use of Availability Concepts in the Railway System

BIRRE NYSTRÖM*

Luleå University of Technology, Sweden

Abstract: This paper explores and describes indicators related to availability in the railway system. Examples of indicators include train punctuality and condition of track. The paper presents losses to the stakeholders, measured by the indicators. Indicators in Banverket (Swedish Rail Administration) and railway literature are presented. Indicators used in the electric power industry are also presented, as these offer good analogies from which to develop additional railway indicators. The indicators found in the literature, but not in Banverket, include travel time variation, slack and wagons. Indicators found by analogy to the power industry concern passengers, traffic work not delivered and how to subtract the effects of adverse weather. A classification of the availability indicators is also suggested.

Keywords: *availability, capacity, punctuality, electric power industry, indicators, maintenance, railway stakeholders*

1 Introduction

Train traffic in the European Union is increasing [1], i.e. more railway network capacity is being used. Simultaneously, there is a wish to improve punctuality [2], while more trains on the tracks might cause delays to spread more easily. This might be counteracted by increasing capacity, i.e. allowing for more train traffic, not just in terms of number but also in terms of length, weight and size, by upgrading existing railway lines or building new ones, and/or increasing railway availability; i.e., loosely put, by improving the dependability of existing lines. At the same time, several European railways are in a state of transition from one single governmental body responsible for the railway to several actors, both public and private. These circumstances call for good ways of communicating between railway stakeholders regarding capacity and availability. The analysis of railway capacity has a long tradition, dating back at least to Lardner's work in 1850 [3], while the interest in availability is of more recent origin.

In this paper, availability concepts that might be of use to railway stakeholders are illustrated. The paper begins with a description of the stakeholders in the Swedish reregulated railway. Thereafter, different railway capacity and availability concepts are described. Then, possible losses due to unavailability to railway stakeholders are described and indicators reflecting these losses found in Banverket (the Swedish Rail Administration) and in railway literature are given. Thereafter, a few industries are described, to which subsequent discussions on availability and related concepts refer. The electric power industry is extensively discussed and power industry indicators given, as well as possible analogies to the railway system. By means of a questionnaire, personnel at a centralised train traffic control office were asked to judge the relevance to their work tasks of the prospective indicators (found in Table 1, Table 2 and Table 5). The indicators that the largest number of respondents judged to be relevant are

* Corresponding author's email: birre.nystrom@ltu.se

indicated by a ‘☺’ in the tables, and the indicators that the largest number of respondents considered irrelevant or non-understandable, are indicated by a ‘☹’ [4]. The paper finishes with a classification of indicators related to availability, and conclusions with some suggestions for further research.

2 Railway organisation

The activities needed to produce any product are operation, maintenance and modification [5]. The exact borders between the activities might be disputed in specific cases, but they all help to transform input to output (performance). The output is then compared to the requirements, and one or more of the activities might then be adjusted so as to reduce the gap.

Railway assets often have a long life, so changes in operation are sooner or later likely to imply changes in maintenance. This is also the case with modification of an item, due to technical development, which might imply modifications of other items and changes to both operation and maintenance.

The most important stakeholders of the Swedish railway are shown in Figure 1. The trains might be run by any of the many Train Operating Companies (TOCs). Banverket is the Infrastructure Manager (IM) and also creates the timetables (i.e. constructs a set of train paths to be used by the trains) and is responsible for train traffic control (TTC). Outside the train traffic production are the customers and public subsidisers. The infrastructure maintenance is performed by any of the Maintenance Contractors (MCs).

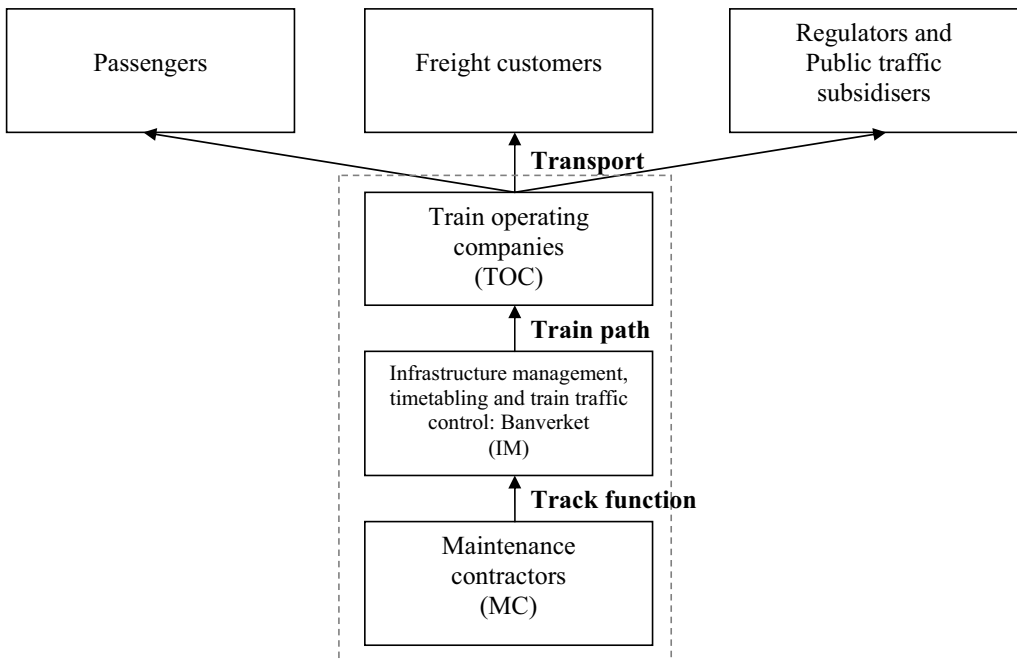


Figure 1: Railway organisation and the most important related stakeholders in Sweden.

The train traffic controllers base their work on the timetable and get input from, among others, train drivers, infrastructure personnel and the train positioning system. The timetable establishes a trade-off between different TOCs, as well as between train traffic and track works, for example preventive maintenance. When non-planned events occur, the controller makes trade-offs in real time, utilising different recovery strategies, such as rerouting and delaying certain trains in order to lower total delays, etc. [6]

3 Railway capacity and availability

The capacity of a railway line is affected by the timetable, train traffic control, signalling, top speed, stations and rolling stock. The ability of train traffic controllers affects the railway capacity when there are deviations. The capacity of railway terminals is affected by the number and characteristics of gates, track length and configuration, as well as loading, unloading and storage facilities [7]. So, the work of e.g. shunting personnel is also essential for sustaining capacity.

Generally, capacity is defined as the ability of an item to meet a service demand of given quantitative characteristics under given internal conditions. Internal conditions might refer to e.g. a specific combination of functioning and not functioning sub-items [8]. So, capacity might be used when discussing the difference between how much is demanded and how much is produced. UIC (International Union of Railways) defines theoretical capacity of a railway infrastructure as the number of trains able to run over it per unit of time, with the trains permanently running with minimum headways between them, i.e., the traffic is as dense as it could be and there are no delays. As running at minimum headway means that there are no margins to prevent disruptions from resulting in secondary delays (i.e. trains are delayed because of other trains hindering them), the theoretical capacity is not obtainable in practice. UIC defines practical capacity of an infrastructure as the number of trains able to run over it per unit of time with the degree of operating quality which corresponds statistically to the level desired, excluding major disruptions. Theoretical and practical capacities depend on the operating plan devised [9]. A more succinct way to define (practical) capacity of an infrastructure is as the ability to forward trains with acceptable punctuality [10]. Note that both these definitions of practical capacity require detailing, among other things, acceptable punctuality.

3.1 Capacity and the timetable

Consider different timetables for a part of a railway line, given by the time-space diagrams of Figure 2, where six trains travel upwards in each diagram. In (a), the trains have two different speeds, while in (b), they all have the same speed, which gives a shorter cycle time, i.e. the capacity is higher. Rearranging the trains of (a), e.g. as in (c), also shortens the cycle time, i.e. increases the capacity of the railway line. The speeds of trains affect the braking distance, and thus the headway, meaning that higher speed does not always give higher capacity. To summarise, the timetable has to be considered when discussing capacity.

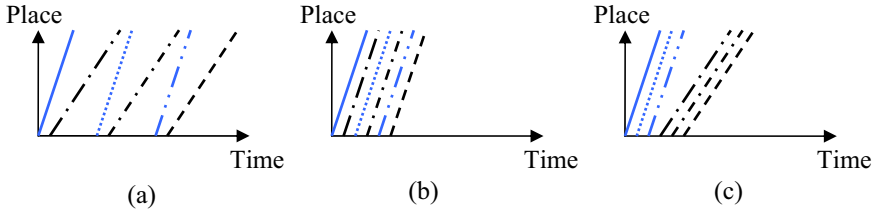


Figure 2: Three timetables (time-space diagrams), each showing six trains. The timetables are heterogeneous (a), homogeneous (b) and with similar train speeds bundled (c). Time is on the x-axis and place is on the y-axis, following Swedish convention.

One way to decrease the impact of the current timetable on the capacity calculations of a railway line is to theoretically compress the trains, i.e. to place the trains at minimum headways, as in Figure 3b. This is the method prescribed by the UIC 406 standard [11]. The output is a percentage; the share of time which the track is occupied by a train between the considered places (about 50% for the timetable in Figure 3). UIC 406 preserves the order of trains when compressing, while Banverket’s method rearranges the trains (similar to changing from Figure 2a to Figure 2c). Hence, the methods might give different figures, as discussed in [12].

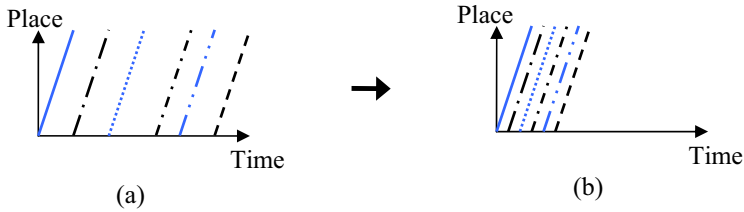


Figure 3: Timetable compression for calculation of capacity utilisation.

Another way is to reformulate the concept of timetable using the demand for train traffic, rather than a specific time-space diagram. This idea is elaborated in [13] and [14], in which a traffic pattern is defined as a set of train movements $\{x_1, x_2, \dots, x_n\}$ and a partial order on these, i.e. there are relations of the type $x_i \geq x_j + \text{time}$. The capacity of a railway network is then taken to mean the inverse of the time between two consecutive traffic patterns. In this way, one avoids calculation of capacity using timetables that were constructed considering railway network bottlenecks that have been removed. In [15], track capacity within station regions is investigated and capacity defined as the time needed to provide a train service intention in the considered region. Practical capacity utilisation might be higher than the theoretically calculated capacity utilisation, because trains are delayed (the space in-between a non-delayed and a delayed train is not possible to use for other trains) and because train routes are set early (that is, the space in front of a train is reserved well in advance). A case study at The Hague HS shows a considerable difference between theoretical (50%) and practical (70%) capacity utilisation [16]. When studying the relation between timetable and capacity, Dutch regulator Rained investigates the effects of prospective investments by constructing several timetables, consisting of both time-space diagrams and platform charts for stations [17].

One might even abandon the concept of absolute capacity, instead asking how much traffic might be added to the current timetable [18]. Each of the trains in the potential traffic increase may be detailed regarding acceleration & deceleration properties, departure time, etc.

To be able to cater for deviations, timetables include allowances, i.e. extra time in excess of the theoretical minimum travel time. Allowances might be allocated along the line or at stations. Margins are the extra time between two consecutive trains [19], i.e. the headway is longer than theoretically needed. Theoretically, allowances and margins lower the capacity of the network. In practice, they are added to reduce delays.

As track maintenance machines and personnel-carrying vehicles are needed to maintain the railway, their track occupation should be subtracted from the capacity of a railway line. This applies even to, e.g. fast Shinkansen measurement trains that traverse the network every tenth day [20].

3.2 Capacity utilisation trade-off

High capacity utilisation might cause delays to multiply more easily. On the other hand, high capacity utilisation allows an infrastructure manager to make gains from the economy of density. In railways, the maintenance cost elasticity relative to traffic volume (gross tonnes) is 0.2-0.3, i.e. a change in traffic of 10% gives a change of 2-3% in maintenance cost [21]. So, there is a trade-off between punctuality and IM's efficiency. However, it is hard to calculate the highest obtainable maximum perturbation tolerance as a function of capacity utilisation, according to Herrmann [22], who investigated station regions. This makes it hard to find an optimum.

3.3 A definition of availability

A general definition of availability (performance) is: the ability of an item to be in a state to perform a required function under given conditions at a given instant of time or over a given time interval, assuming that the required external resources are provided [8]. An item might be anything from a tiny component to an entire system, as long as it can be considered separately.

A definition of railway availability is:

Availability = Obtained capacity / Planned capacity,

calculated for a point in time or over a time interval. This definition treats availability as a function of the Obtained capacity and a constant Planned capacity. The Planned capacity might be set to the Theoretical capacity, the Practical capacity, or lower. One advantage of defining Planned capacity as the Theoretical capacity is that availability then takes on values between 0 and 1, not higher. The concepts are illustrated in Figure 4. The Obtained capacity consists of Required capacity and Spare capacity.

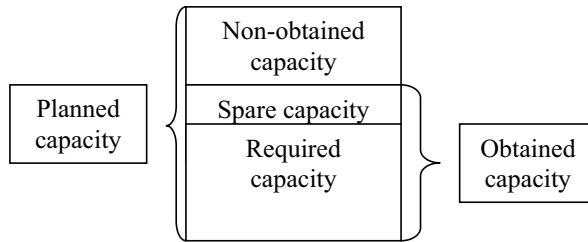


Figure 4: Availability might be defined as the quotient $\text{Obtained capacity} / \text{Planned capacity}$.

Spare capacity might absorb variations from day to day or a future traffic increase [23]. It is understood that spare capacity on an individual section might absorb variations in travel times there and in its vicinity, while spare capacity along entire lines is needed to establish new train traffic. As seen earlier, no commonly acknowledged definition of the capacity of a railway line exists. The same is true for availability.

4 Performance of the railway

Figure 5 shows how performance of the railway might be considered. The performance experienced by the passengers is the right-most arrow in Figure 5. The control is considered to affect the infrastructure and rolling stock, but not the passengers, as illustrated by the arrows. Passengers affect performance, or seen the other way around, losses. Examples are how fast they board the train and how tight connections they choose. The performance of the infrastructure (including track works), the performance of rolling stock (including personnel) and the performance perceived by the passengers, might all be named 'performance'.

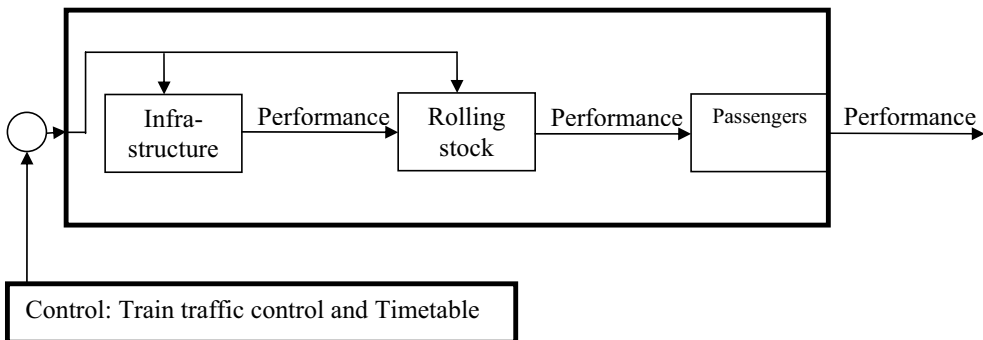


Figure 5: The performance of the railway might be considered to be the performance of the infrastructure, the rolling stock performance or the performance perceived by the passengers (or freight customers), dependent on where one considers the border of the railway system to be.

Note that one cannot calculate the total availability as the product of the availabilities of the items infrastructure, rolling stock and passengers. This is because the items might be available at different times and an item might be considered available, despite its receiving no input and therefore not being able to perform [5]. As several items might be unavailable consecutively or simultaneously in time, the problem of apportioning responsibility for the total unavailability arises. A procedure to solve this for delays in construction projects is suggested

in [24]. According to the procedure, the schedule is iteratively updated and the change in project duration apportioned. The procedure also accounts for time-shortened activities.

5 Indicators related to unavailability in Banverket and in the railway literature

Losses related to unavailability (i.e. lack of availability), encountered in Banverket and in the railway literature, are listed in Table 1 and Table 2, respectively. Most of the indicators can be detailed further and, directly or indirectly, affect all of the stakeholders. Table 1 and Table 2, show that indicators encountered in the railway literature, but not in Banverket, deal with travel time variation, slack (more time in the timetable) and wagons.

5.1 Availability concepts in Banverket

Although having no formalised availability concept, Banverket employees use the term availability and there are related indicators in its strategic plan [26]. Some of the indicators, as signified in Table 1, are readily associated to a delay code in Banverket's delay attribution system, TFÖR [25]. The indicators reflect availability, although the requirements in Table 1 are formulated in other words.

Table 1: Losses associated with unavailability, found within Banverket. In the left-most columns, the stakeholders to whom the indicators are most relevant, are indicated by ‘x’. The starred indicators were not explicitly defined by Banverket. The relevance to train traffic control personnel is indicated by ☺ or ☹. The rows with thick vertical borders are further described in the text.

P a s s	T O C	T C	I M	M C	Requirement	Indicator	Loss scenario
x					Punctuality	Fraction of departures within _ min; fraction of arrivals within _ min; sum of arrival delay minutes (TFÖR) ☺	Delay may incur costs; hindering of other trains; lost connections; worried passengers.
	x		x		Punctuality	# unplanned speed restrictions [26] ☺	Delays.
	x	x			Timeliness	Fraction of train path applications granted [26]	Demand is not fulfilled.
	x		x		Timeliness	# changes to the Network statement [26] ☹	Planned traffic might not be carried out.
	x				Passengers / goods on embarking / loading place	# late departure from freight terminal (TFÖR)	Delays; more work for yard workers; time losses due to extra transfer or change of transport.
	x				Clearance gauge	Fractions of different clearance gauges*; # not reported changes to encroachments when a track work has been carried out*	Too small clearance gauges restrict traffic; too large might mean costs to IM; unreported encroachments might cause accidents; costs for finding clearance information.
		x	x	x	Maintenance slot	# time exceeded for track work (TFÖR)	Too much or too little time has been planned.
			x		Availability	# faults divided on type of asset and cause code [26]	Failure causes unavailability, which causes repair costs.
		x	x		Availability	# disturbances resulting in train disturbances [26]	Failure causes unavailability, which causes train delay.
			x		Availability	Disabled time (several related indicators are in [26])	Unavailability.
			x		Condition, Availability	# inspection remarks broken down by urgency [26]	Infrastructure in too bad condition means expenses for corrective maintenance for IM, losses of comfort for TOC, etc. Too much preventive maintenance means expenses for IM and lower availability for TOC.

Table 1 shows that the indicators range from changes in the Network statement (which, for a certain timetable period, describes the infrastructure, specifies order of capacity allocation and requirements on TOCs [27]) via punctuality to disabled time and the condition of the railway infrastructure.

The indicators in the first and second rows consider punctuality. The punctuality indicators reflect the effect of unavailability, for example the ‘Fraction of arrivals within 5 minutes from timetabled time’ (for which data is found in the TFÖR system). The ‘Number of unplanned speed restrictions’ tells that availability of track is lowered, which might cause trains to

become unpunctual. Timeliness, i.e. the extent to which a journey or transport occurs when desired, is reflected by two types of indicators in Table 1. The ‘Fraction of train path applications granted’ tells to what extent the demands for track time of the TOCs are fulfilled by the timetable. In other words, to what extent the supply of train paths covers demand. However, a low fraction of applications granted does not necessarily mean that there is a high capacity utilisation – it might well mean that the TOCs, taken together, demand train paths for their traffic that are not possible. The ‘Number of changes to the Network statement’, on the other hand, shows when and where the planned capacity has not been reached.

The ‘Number of time exceeded for track work’ is, to some extent, to be found in TFÖR, in which occasions when trains get delayed hereby are recorded. For Disabled time of the railway infrastructure, the second last row of Table 1, it is possible to get only approximate data from Banverket’s infrastructure failure database. The symptoms and causes of failure, however, are detailed, making it, to some extent, possible to subtract various externalities so as to obtain more comparable indicators of availability. External causes of unavailability might include some instances of bad weather, animals on track, trees fallen on track, objects struck on level crossings and sabotage.

5.2 Availability concepts in the railway literature

Indicators reported in the literature or constructed by the author are listed in Table 2.

Table 2: Losses associated with unavailability, found in railway literature or constructed by the author (starred). In the left-most columns the stakeholders to whom the indicators are most relevant are indicated by ‘x’. The relevance to train traffic control personnel is indicated by ☺ or ☹. Each indicator is further described in the text.

P a s s e n g e r s	T O C	T T C	I M	M C	Requirement	Indicator	Loss scenario
x					Punctuality	# non-delayed passengers / # delayed passengers	Delayed passengers.
x					Punctuality	Travel time variation.	Delays; lost connections; increased waiting; less train traffic fits into the railway network.
x					Punctuality	# last minute platform change; # dwell delay, i.e. delay due to extra passengers embarking	Departure delay; crowding; bad traffic information.
x					Availability	Fraction of trains that have planned standard for all wagons	Too few or wrong wagons, which causes delays and/or lower comfort.
	x				Timeliness	$A = 1 - \text{Slack time} / \text{Total travel time}^*$ ☹	Longer travel time.
		x			Maintenance information	Fraction of planned but not used track work time*	Worse condition of track; lower capacity available for trains.
		x			Traffic/train information	Fraction of not accurately registered train data* ☺	Lower registered than factual train weight, renders the IM lower charges and trains might get stuck on slopes.
		x	x	x	Availability	Formula for availability given punctuality, maintenance, line and traffic volume	Suboptimal use with respect to current traffic, maintenance and assets.
			x		Punctuality	Actual delays / Simulated delays*	Availability is not allocated to the best places possible in the network.
x	x	x	x	x	Availability	Multi-dimensional availability: loading, embarking, allowed trains and speeds	Lower performance than expected.

Table 2 shows that the indicators range from timeliness to punctuality and provides information related to availability such as data on cancelled track works.

The indicator on the first row of Table 2, Number of non-delayed passengers / Number of delayed passengers (in [28] called ‘Service reliability’) has the disadvantages of being unbound and being sensitive to changes to the threshold when a train is to be considered as delayed, thus making it difficult to make historical comparisons. On the second row, the travel time variation shows how stable the traffic process is and thus how far the practical capacity is under the theoretical capacity. Travel time variation is studied in [29]. Dwell delay, i.e. delay due to extra passengers embarking, is an effect from train delays, but also passenger behaviour and passenger information.

The fraction of trains that have a planned standard for all wagons (in [30] called ‘comfort dependability’) is dependent not only on rolling stock, but also the ability of the railway infrastructure to convey the wagons to the place where they are needed.

The amount of slack time relative to total travel time is an entity often encountered in papers on railway operations research, e.g. [31]. This inspires the author to define availability as:

$$A = 1 - \text{Slack time} / \text{Total travel time}$$

for a specific train, for a specific TOC as well as for all TOCs. This is an indicator of track time losses and hence also of lack of timeliness.

The indicator ‘Fraction of planned but not used track work time’ tells how often the maintenance organisation fails to correctly assess how much time it will need on the track, i.e. how much capacity that is needed. It is cumbersome to calculate, since data on track work time is currently recorded only on paper. The fraction of planned, but not used, track work time should be put in relation to the time exceeded for track work (Table 1).

Fraction of not accurately registered train data (to the appropriate database) gives the unavailability of one kind of information needed to control train traffic appropriately. The problems associated with this may be easily remedied by denying departure to trains not registered.

Krueger [32] gives a formula for the delay volume as a function of railway line layout, maintenance, traffic composition and traffic volume, in his study on the Canadian National Railway. Simulation involving characteristics of a line (e.g. how uniformly sidings are distributed), traffic (heterogeneity and how peak hour traffic relates to average traffic) and maintenance (e.g. track work time) were used to construct a delay estimate. Key parameters, such as the ones exemplified, make up each of the factors in the approximation:

$$\text{Delay volume} = \text{Factor line} \cdot \text{Factor traffic} \cdot \text{Factor maintenance} \cdot e^{k \cdot \text{Traffic volume}}$$

Now specify the delay volume (i.e. acceptable punctuality). The Factor traffic might then be regarded as the availability (to train traffic) of the investigated line, and is then given by the equation, when the line, maintenance regime and traffic volume are known. The author notes that this way of defining availability is analogous to how capacity is defined in [10]. One way to measure heterogeneity is given in [33]. In [33], the Sum of Shortest Headway Reciprocals (SSHR) is defined as $SSHR = \sum_{i=1}^n (h_i^-)^{-1}$, with h_i^- the shortest headway between trains i and $i+1$, train n is followed by train 1.

The quotient Actual delays / Simulated delays might be used as an indicator. It shows how well availability is allocated among items of the same type, e.g. turnouts. One way to calculate it is as follows. Assign each turnout the same probability of functioning as the average turnout. Simulate the train traffic along the railway line. A small quotient then indicates a good allocation of availability, i.e. availability is where it is most needed, in order to prevent delays.

5.2.1 Multi-state and multi-dimensional availability

The last row of Table 2 concerns multidimensional availability. When an item can be in more states than an up-state or a down-state, it is said to be able to attain multiple states of availability. When, for example, a light bulb fails, it changes to a complete down-state. However, this puts the railway network in only a partial down-state, as trains still might be allowed to pass the signal at reduced speed.

The EN 50126:1999 Railway Standard [34] defines System Availability, A , as:

$$A = \text{Mean Uptime} / (\text{Mean Uptime} + \text{Mean Downtime}),$$

which seems to require items to be in either a down-state or an up-state (0 or 1 system).

However, EN 50126 also categorises failures in different classes:

- Significant (immobilising failure): Prevents train movements or causes a delay and/or cost over specified thresholds
- Major (service failure): Must be rectified to achieve specified performance, do not cause delay or cost over thresholds specified
- Minor: Does not prevent a system to achieve its specified performance

This categorises some train-stopping failures causing delays just longer than the threshold, e.g. ice in a turnout as Significant, while chronic speed-restricting failures causing delays just under the threshold, e.g. signalling failures lowering the speed, are categorised only as Major. So, this categorisation does not take into account multiple failures. One way to consider these is to aggregate them on a higher level, e.g. by calculating availability as one minus the sum of all downtimes on a line, or as the product of all section availabilities. However, the availability figures calculated in this way for a line will differ considerably, even over short distances of about ten sections (about 100 km).

Availability might be lowered with respect to the dimensions:

- Loading and unloading for road and water
- Embarking and disembarking
- Kinds of trains that might traverse a route: axle load, track regularity, length, cross section, electric power requirements, train dynamics
- Speed.

5.2.2 More on availability-related concepts

EN 50126 [34] defines System Availability using planned Non-Availability and unplanned Non-Availability. One might assess unplanned downtime as worse than planned downtime. Another assessment regards the length of the downtime, indicated by the probability of not getting failures that take longer than a specified time t_r to repair (called ‘mission availability’ in [35]), as well as in the uptime distribution function $F_U(t)$ [36], i.e. the probability that an uptime will be less than, or equal to, t .

A way of directly using punctuality to define availability is given by the contract of the high speed line south of Amsterdam. The Dutch define availability of this railway line as:

$$A = 1 - \text{Delay time caused by infra provider} / \text{Planned travel time.}$$

The delay time in the formula is not the real delay, but calculated, in order to omit the effects of the behaviours of others, e.g. trains delayed from Belgium [37].

A high capacity utilisation might also make it harder for the traffic to recover from delays. In [38], recovery potential is defined as the time for an initial delay of 60 minutes to each train to

diminish. Andersson & Berg [39], among others, define recovery potential as the difference between initial delay and final delay, divided by initial delay. Metro traffic is investigated in [40], which defines recovery as the time needed to regain the planned time interval between trains. Although specific trains are delayed, passengers are not delayed. The author notes that a recovery time cannot be the sole indicator, as the consequences of different recovery strategies, such as turning some trains prior to their timetabled terminus, vary between railways.

A concept related to capacity is resilience. Engineering resilience is defined by Gunderson & Pritchard [41] as the speediness back to normal after a perturbation, while Ecological resilience is taken to mean the disturbance that can be absorbed by the system before other variables and processes take control of the system. The railway's Ecological resilience might be considered to be exceeded when extreme weather limits the train traffic or when rerouting to road is necessary, due to a major fault.

6 Availability concepts in other industries

A few industries are studied in order to be compared to the railway: the military mission, computing, electric power, house painting and manufacturing. The industries might be classified according to their relation between production and consumption, with respect to time and space, as in Table 3. For the purpose of this paper, useful analogies may be drawn from the electric power industry. Therefore, it is described in greater detail.

Table 3: Different products have different time and space relations between production and consumption.

	Same location	Different location
Same point in time	Train transport, The military mission, Computing	Electric power
Different point in time	House painting	Torch (Manufacturing)

6.1 Electric power

The electric power industry has moved from integrated utilities into companies generating the power and companies distributing the power. An electric power network is built in a hierarchy of power grids, with transformers between grids of different voltages. Interesting to customers are the interruptions (i.e. loss of electric power), not the outages (i.e. faulty state of an item) in higher voltage grids or in power generation, because grids closer to the customers often are redundantly fed from several higher voltage grids [42]. For reasons of varying demand and the chance of power generation outages, the power industry relies on prognoses and the possibility to buy more power. A primary power reserve, achieved e.g. by only part-loading a water-turbine generator, allows for rapid increase in production, but might be expensive. One might note that the same is true, when lack of (perceived) availability is compensated by customers who buy backup power equipment, e.g. batteries or diesel generators. The voltage varies due to demand variations. The voltage quality is regulated in SS-EN 50160 [43], including maximum and minimum voltages, rapid changes in voltage and flicker.

6.1.1 Availability and related measures

Power 'System reliability' might be divided into 'System adequacy', meaning to what extent the energy supplied is enough to cover demand, and 'System security', meaning the power system's ability to deal with disturbances inside itself [51].

The availability of a power plant might be measured with respect to time; ‘Time used factor’ is the fraction of total time the power plant produces power, regardless of the effect generated, called ‘on-stream availability’ by Aven [44] and the Z-016 oil and gas industry standard [45]. The continuous nature of oil & gas transports motivates the use of ‘Throughput availability’, defined as the expected throughput as a fraction of the demand. However, as the production of railways is not continuous and 100% throughput availability might be the case even in the case of much unpunctuality, this is not a suitable indicator for railway traffic, at least not for longer periods of time, as it is only affected by cancelled trains.

The ‘Time availability factor’ is defined as the fraction of total time the plant produces power or is on standby (i.e. ready to produce power). The ‘Energy availability factor’ is calculated for a certain time interval and denotes the quotient

$$\text{Energy availability factor} = \frac{\text{Maximum energy producible using the current available power}}{\text{Maximum energy producible using stated maximum power}}$$

So, this indicator relates the energy that the power plant was able to produce to the energy that the power plant should have been able to produce [46, 47].

The severities of interruptions are calculated as ‘Power disconnected’ (kW) and ‘Energy not supplied’ (kWh). In order to estimate the energy not supplied, Swedish Vattenfall uses the day’s average temperature, and also takes into account e.g. weekly fluctuations [48]. Estimating ‘Energy not supplied’ is not straightforward; it is calculated differently in e.g. the different Nordic countries [49]. When factories start their processes after an interruption, the energy not delivered due to the start, using less power than during regular operation, is not included in ‘Energy not supplied’ [50]. It is unclear how ‘Energy not supplied’ should be calculated for a higher voltage grid, when lower voltage grids are redundantly fed from other grids.

The most commonly used availability-related indicators all calculate averages with respect to the customers. These include System Average Interruption Frequency Index (SAIFI), System Average Interruption Duration Index (SAIDI), Customer Average Interruption Duration Index (CAIDI) and Average Service Availability Index (ASAI). The calculation of the indicators is straightforward. SAIFI is calculated as the number of interruptions per time unit per customer. SAIDI is the average length of interruptions for customers served during a year. CAIDI concerns the lengths of the interruptions for the affected customers (customers who suffer several interruptions is counted as multiple customers). ASAI is the proportion of time when power is available. Another indicator is Customer Average Interruption Frequency Index (CAIFI), which is defined as the average number of interruptions for the affected customers during a certain period of time. Customer Total Average Interruption Index (CTAIDI) is similar to CAIDI and is the average length of interruptions for affected customers (customers who suffer multiple interruptions are counted as one customer). Average System Interruption Frequency Index (ASIFI) is a load-based index, based on the subscribed effect (kVA). It is calculated as interrupted effect divided by connected effect. In the case of large differences in customer size, it gives a more accurate picture than SAIFI does. Average System Interruption Duration Index (ASIDI) is calculated analogously and includes the duration of the interruption [51, 52]. Examples of values are given in Table 4. The data covers more than 90% of the customers in Sweden during 2004. Interruptions of more than three minutes in length, planned and unplanned, are included [53].

Table 4: Indicators for grids with different voltages in Sweden 2004 [53].

Grid	SAIFI	SAIDI	CAIDI	ASAI	Total number of interruptions	Total number of customers affected
	Interruption frequency - all customers	Interruption length - all customers	Interruption length - affected customers	Availability (fraction of time)		
	#/year	Minutes/year	Minutes/year			
24 kV	0.197	14.248	72.362	0.999973	2827	1011406
12 kV	0.540	55.543	102.779	0.999895	13925	2640403
<10 kV	0.006	0.245	37.881	1.000000	105	31643
0.4 kV	0.038	6.047	159.520	0.999989	23048	185199

SAIFI, SAIDI and CAIDI give different orderings of the worst causes of interruptions in Canada, e.g. SAIFI points out the cause ‘Defective equipment’ to be the worst, while SAIDI gives ‘Loss of supply’ and CAIDI ‘Adverse weather’ [54]. Considering SAIFI and SAIDI, ‘Adverse weather’ ranks seventh and third respectively. This example illustrates that a tacit valuation is made by the selection of indicator. From the customer’s point of view, it is better to experience many interruptions at short time intervals than the same number of interruptions separated by days or weeks, a phenomenon which, as argued in [54], affects the suitability of indicators.

Large interruptions, caused by e.g. extreme weather, might be treated separately when evaluating. The IEEE P1366 standard [55] defined ‘major reliability events’ as 10% of the customers being affected in a 24-hour period, while IEEE 1366-2003 [52] employs daily SAIDI to statistically define the outlying ‘major event days’ by calculating a SAIDI threshold.

Two different indicators are given by dividing interruptions by the value of the energy delivered respectively by dividing interruptions by the length of the power lines. The first quotient takes the consumer’s perspective and the second more of the producer’s, as some of the network might consist of failure-prone air wires, with fewer redundant feedings, to sparsely populated, remote areas. Another indicator is the recovery ability, which might be given as the number of customers reconnected per day during a large interruption.

For the Swedish power market, the Network Performance Assessment Model [42] calculates the regulation price. The regulated maximum price is computed by constructing a virtual, ideal network. The locations of the customers, nominal voltages and energy bought are inputted into the model. By dividing income from sales by network performance including a quality bonus, a quotient is calculated. Based on this quotient, the price is regulated. Thus, a higher availability allows for a higher price.

6.2 Availability concepts in other industries summarised

A characteristic of the power industry is that power generators and grids on different voltages are all needed to deliver energy to the customer, but redundancies in generation and higher voltage grids allow for unavailability of equipment, without affecting the delivery. This comes with the cost of having more generators, more expensive energy generation and more wires. Availability, or rather unavailability, might be measured as the fraction downtime, effect lost or amount of energy not delivered, the latter being somewhat difficult to estimate.

How to consider the availability of a distribution grid that is not producing (e.g. due to a breakdown in a higher voltage grid) is also an issue. One needs to be aware that different indicators might give different rankings of causes of failure, as well as how one chooses to assess planned interruptions, compared to unplanned interruptions. How to get indicators that give fair performance figures despite external impact is also an issue. One way to measure performance is to compare a virtual network to the real network, regarding interruptions.

For the military mission, the performance of, e.g. an aircraft might be measured by its System effectiveness, in [56] defined as the product of three probabilities:

$$\text{System effectiveness} = \text{Operational readiness} \cdot \text{Mission reliability} \cdot \text{Design adequacy}$$

‘Operational readiness’ is the probability that the item has begun operating within the time stated. ‘Mission reliability’ is the probability that the item will continue to operate, given that it did so at the beginning of the mission. ‘Design adequacy’ is the probability that the item finishes its mission, given that it operates within its design specifications, thus the similarity to power ‘system adequacy’ is understood.

For the computer industry, faulty specifications are a major concern, probably because the environment and users’ requirements change fast. The rapid pace makes an issue of a quantification of maintainability (in a broad sense, including performance during updates and changes) of different configurations, e.g. with different numbers of redundant components (concepts are found in [57] and [58]).

For house painting, short interruptions in the work are not important and the capacity might be increased by more personnel working simultaneously. Quality is hard to assess objectively.

For manufacturing, operations to produce, e.g., a torch might include turning, painting, assembling and packaging, possibly with buffers between the operations. The finished product is stored and variations in demand are thereby handled at the cost of the capital tied up and the risk of the product not being in demand in the future. The manufacturing speed has an upper limit, but often, machines are not run at top speed, because e.g., cutting tools might wear uneconomically fast. So, defining a specific top speed might be hard to justify. The speed of the production process might not reveal how much is produced, as faulty pieces may be reworked in one or several operations.

In the presented industries, it is seen that availability might alternatively be viewed as a frequency, probability or as the product of three probabilities; start mission, complete mission and design allows success. None of the encountered indicators answer how a scattered uptime should be described in terms of availability, compared to a continuous uptime. It is preferable to measure the benefit (e.g. energy) lost during downtime rather than downtime hours. The production profile varies between different industries; from producing when needed, as in the military and in power, to producing according to a timetable, as in the railway system, to producing constantly at a maximum speed, as in manufacturing. Therefore, availability, defined as a quotient between capacities, means different things in different industries, and the figures should therefore not be compared.

7 Indicators related to unavailability in the electric power industry and their analogies to the railway

Analogies can be drawn between the industries discussed in Section 6 and the railway. It is desirable to have indicators for each of the stakeholders on each of the levels: passenger, train, network, route and asset. From the computing field, it was found that some analogies can be drawn on all levels. In the military, the stakeholder structure is different to that of the railway, which makes it difficult to draw analogies. Finding relevant analogies in house painting was difficult. Regarding manufacturing, analogies were found, but as its products are possible to store, it is difficult to find analogies on the level of its customers (corresponding to passengers). A characteristic of transport is that postponed production in itself not only causes the losses of delayed delivery and less production over a certain time interval, but also causes a product that is faulty in itself. Unlike manufacturing, it is not possible to reject the product at a final check and rework it.

The analogies found from electric power span at least the levels passenger, train, network and route. Electric power also has a similar stakeholder structure to that of the railway; power distributor (corresponds to the IM), power generator (TOC) and power consumer (passenger). A difference is that electric power production does not follow a timetable, as demand varies. Electric power indicators that might supplement the existing railway indicators are listed in Table 5, from which it is seen that many indicators are relevant to passengers. The table does not aspire to give all possible analogies. The drawn analogies include power (kW) equals speed (km/h) and energy (kWh) equals travelled distance (km). The railway indicators constructed might be further specified, e.g., speed might be measured as km/h, passenger-km/h or tonne-km/h.

Table 5: Electric power performance indicators and the loss scenarios they reflect and possible analogous indicators in the railway system. In the left-most columns the railway stakeholders to whom the performance indicators are most relevant, are indicated by ‘x’. The relevance to train traffic control personnel is indicated by ☺ or ☹. The rows with thick vertical borders are further described in the text.

P a s s e n g e r s	T O C	T C	I M	M C	Indicator	Loss scenario	Possible railway indicator and/or loss scenario
x	x	x	x		ASAI - Average service availability index (i.e. fraction uptime at outlet)	No power in outlet.	Fraction uptime as the time the passengers (train) move with planned speed; fraction uptime of a certain location on the line.
x	x	x	x		SAIDI - System average interruption duration index	Loss of power, total customer base (time).	Downtime per passenger, during a given time interval.
x	x	x	x		SAIFI - System average interruption frequency index	Loss of power, total customer base (#).	# trains hindered
x	x	x	x		MAIFI - Momentary average interruption frequency index (maximum 5 minutes).	Short interruptions are severe, too.	# failures resulting in maximum 5 minutes downtime, calculated with respect to passengers, trains or IM's items. ☺
	x	x			MAIFI _E - Momentary average interruption frequency index	Interruptions in rapid sequence are together considered as bad as the last, longest, interruption.	# gradual failures such as obstructing wagon door which finally fails, turnout needs gradually more attempts to be thrown and finally fails. ☺
x					CAIDI - Customer average interruption duration index	Loss of power and loss of energy, per affected customer.	How much each passenger on a journey has been delayed.
x					CAIFI - Customer average interruption frequency index	Loss of power and loss of energy, per affected customer.	# each passenger has been delayed
x					CTAIDI - Customer total average interruption duration index	Loss of power and loss of energy, per affected customer.	How much each passenger that has been delayed any time, has been delayed.
x					Value of lost energy to customer.	Damage to equipment, property or production.	Passenger's delay cost.
x					CEMI _n - Customers experiencing multiple interruptions	Frequent loss of power for some customers.	Fraction of passengers that are delayed more than a certain number of times per week or year. ☹
	x				ASIDI - Average system interruption duration index (i.e. energy not delivered).	Energy is interrupted for power consumers during a certain period of time.	Traffic work not delivered, calculated as the product downtime · planned speed
	x				ASIFI - Average system interruption frequency index	Interrupted effect / Connected effect.	1 - Speed / Planned speed
	x	x			Regulation price model.	Low quality of electric power lowers price.	The extent to which prices of journeys and train paths reflect quality.
	x				Time availability factor - the fraction of time production is possible.	Loss of the ability to deliver power will make it more difficult to sell.	Fraction of time that trains might travel the tracks.
	x	x	x	x	Major event days.	Some days are excluded when calculating indicators.	Exclude delays due to adverse weather or a specific event.

On the third row of Table 5, the SAIFI analogy, ‘Number of trains hindered’, is relevant to passengers, but also to TOCs, TTC and IM, as they might want to know the number of trains hindered by, e.g. turnout failures.

The indicator MAIFI_E reflects the fact that outages often cause one another to happen in a rapid sequence. An analogue indicator in the railway system might consider gradually failing items that finally become totally unavailable, e.g. a turnout, or delayed trains that cause secondary delays. Analogue indicators might be taken to be the number or length of long-lasting total unavailability of an item and the number or length of secondary delays, respectively.

Another railway indicator found by analogy is the fraction of passengers that suffers more than a specific number of delays per year, calculated e.g. for popular commuter routes.

‘Traffic work not delivered’ is probably best calculated per hour or day, as delayed traffic is carried out later. In order always to sum to at least zero, the indicator should only consider trains that are to travel during the considered time period, according to the timetable. Cancelled traffic is better reflected when this indicator is calculated for longer periods of time, e.g. a month or a year. ‘Traffic work not delivered’ might be easier to calculate more accurately than energy not delivered, as demand for train traffic follows a timetable, whereas demand for power does not.

In Table 5, the last row concerns extreme events. To exclude delays due to severe weather, railway stakeholders might consider adapting the statistical methodology in IEEE 1366 [52] or simply when 10% of the passengers are affected during a day.

One might define availability as the fraction of the planned capacity that is obtained for a specific period of time (Section 3.3). This implies that what is not produced during that period has no value (as that production is postponed to the end of the life-length of e.g. a power grid, power plant or oilfield, depreciation makes the production nearly worthless). In the railway system, on the other hand, a slightly delayed journey might still have a considerable value.

As opposed to the contemporary technology employed during interruptions in the power industry, the railway might consider its customers individually. The passenger buys a trip, maybe involving several trains, so one needs to define whether passengers, trips or trains are considered. Practical measurement problems include passengers who do not buy tickets for each trip, as they have a season ticket. Yet another problem on the border between passengers and trains concerns the existence of replacement buses for cancelled trains. Train traffic controllers disagree whether to report the train or the passengers (i.e. the bus) to the delay attribution system [59].

Banverket measures delays (of five minutes or more) and failures, in analogy to the power industry, which measures interruptions (five minutes or less are termed momentary) and outages. Today, the current state of train traffic is often graphically illustrated on screens showing the number of delay minutes for different trains – a more up-to-date picture may be the train speeds relative to timetable.

8 A classification of availability indicators

From the indicators presented, it is suggested to label availability indicators with respect to their scope:

- Stakeholder: stakeholder and level relevant to the indicator. The level might range from availability of the entire system, of concern to e.g. end customers, to availability of a tiny component, of concern to e.g. IM's component procurers.
- External disturbances: to which extent the effects of disturbances external to the system, e.g. sabotage, are included. In the military, such disturbances are included in the Design adequacy factor.

The indicators might be further classified with respect to how they are calculated as belonging to one of the categories:

1. As a function of punctuality or delay.
2. As the quotient Obtained capacity / Planned capacity. As far as the railway system is concerned, where production and consumption occur simultaneously, it might be beneficial to define the Obtained capacity with respect to the Planned capacity or vice versa. Downtime during Required time is defined in, inter al., [5].
3. As the product Operational readiness · Mission reliability · Design adequacy. Analogue factors in the railway system might be Departure punctuality, Underway punctuality and Arrival punctuality, the last two defined as conditional probabilities.
4. As a frequency. As opposed to indicators of categories 1-3, such an indicator does not usually lie in the range [0, 1]. An example is the number of inspection remarks per year (in Table 1).

The majority of the indicators in Table 1, Table 2 and Table 5 belong to category 1 or 2.

A psychological attribute of an indicator is whether it is positively or negatively correlated to performance, i.e. whether a higher figure is better than a lower or vice versa. This differs between the indicators given in Table 1, Table 2 and Table 5.

In Table 6, a hierarchy of availability indicators is given, indicating the most concerned stakeholders on each level, together with possible methods of translating between the indicators on different levels. Note that the indicators in Table 6 only are examples and might be specified in several variants; for example, passenger delay minutes might be defined so as to include or not include the effect of a train delay causing lost connections in other transport modes.

Table 6: Examples of indicators related to availability on different levels and methods of translating between them. For each level, the most concerned stakeholders are indicated.

Level	Concerned stakeholders				Indicators
	Pass.	TOC	IM	MC	
Passenger	x	x			Traffic information, crowding, delay minutes
▲	<i>Train delay minutes to Passenger delay minutes: Calculations on passenger counts</i>				
Train	x	x	(x)		Delay minutes, seat utilisation, vehicle utilisation
▲	<i>Vehicle utilisation: Travel time with top speed divided by Actual travel time</i>				
Network	(x)	(x)	x	X	Delay minutes, recovery potential
▲	<i>Allowances and margins to Recovery potential: Discrete event simulation</i>				
Route	x	x			Punctuality, delay minutes, allowances and margins
▲	<i>Asset availability to System availability: Reliability block diagram</i>				
Asset			x	X	# failures per passage of train (by failure mode)

9 Conclusions

Making analogies from the electric power industry is rewarding, since stakeholder structure and production characteristics are similar to those of the railway system. Many power industry indicators directly concern the power customer, from which passenger-related indicators might be adapted. As different indicators might give different rankings of causes for unavailability, the choice of indicators is important.

Availability indicators encountered in the railway literature, but not applied in Sweden, concern slack, travel time variation and wagons. New railway indicators found by analogy to the power industry include indicators that concern passengers to a large extent, traffic work not delivered and how to subtract the effect of adverse weather. Many indicators are of concern to several stakeholders.

Most of the availability indicators are calculated based on punctuality or on capacity. In order to be able to compare availability, the definitions in both cases need to be further standardised. The degree of abstraction of the timetable is important. The fact that unpunctual trains have a value, but not as high as punctual trains, advocates availability being defined relative to punctuality.

It is suggested to categorise availability indicators with respect to concerned stakeholder (e.g. TOC), inclusion or not of external disturbances (e.g. effects of heavy snowfall) and how they are calculated (one of the alternatives 1-4 given above).

To summarise, there are numerous possible railway availability indicators. Some of the less suitable uses, as well as the better ones, have been exemplified in this paper. It remains to calculate numerical examples from real-life data in order to get a sense for the magnitudes and variations of the indicators. The choice of indicators might be steered by the desire to maximise the correlation between pairs of indicators at different levels and possibly to minimise it on the same level. Also, prospective indicators need to be tested for clarity by stakeholder individuals. For train traffic control, availability of information is an important issue, for which indicators might need to be further developed.

Acknowledgements

The author would like to thank Banverket and the EU regional fund for financial support. The author is grateful to several persons for useful suggestions on earlier versions of the manuscript. First of all, Professor Per Anders Akersten at Växjö University. Also, Dr Peter Söderholm and Professor Uday Kumar, both at Luleå University of Technology. Finally, Professor Ulf Sandberg, BU Services, Vattenfall, for sharing his knowledge in electric power.

References

-
- [1]. EU, *Energy, transport and environment indicators, Data 1991-2001*, Office for Official Publications of the European Communities, Luxembourg, 2004.
 - [2]. The Swedish Ministry of Enterprise, Energy and Communications, *Regleringsbrev för budgetåret 2006 avseende Banverket m.m. inom utgiftsområde 22 Kommunikationer (Appropriation directions for the fiscal year 2006 to the Swedish Rail Administration, etc.,*

within expenditure area 22 Communication), 2005. <http://webapp.esv.se/statsliggaren> (07-05-28). (In Swedish.)

[3]. Lardner, D., *Railway Economy: a treatise on the new art of transport its management, prospects, and relations, commercial, financial, and social. With an exposition of the railways in operation in the United Kingdom, on the continent, and in America.*, Taylor, Walton, and Mabe, London, 1850.

[4]. Nyström, B., *Enkät svar från DLC Boden (Answers from Train Traffic Control Boden)*, Diariennr 02889-2007, Luleå University of Technology, 2007. (In Swedish.)

[5]. SS 441 05 05:2000, *Tillförlitlighet – Ordlista / Dependability – Terminology. Utgåva 3.*, SIS, Swedish Standards Institute, Stockholm.

[6]. Hofman, M., L. Madsen, J. J. Groth, J. Clausen & J. Larsen, *Robustness and Recovery in Train Scheduling – a simulation study from DSB S-tog a/s*, ATMOS 2006 6th Workshop on Algorithmic Methods and Models for Optimization of Railways, 2006. <http://drops.dagstuhl.de/opus/volltexte/2006/687> (07-10-03).

[7]. Weigel, M. L., *A railroad intermodal capacity model*, In: Tew, J. D., S. Manivannan, D. A. Sadowski & A. F. Seila (Eds.), *Proceedings of the 1994 Winter Simulation Conference*, Orlando, FL, pp. 1229-1232, 1994.

[8]. IEC 60050(191), *International Electrotechnical Vocabulary*, International Electrotechnical Commission, <http://std.iec.ch> (07-10-03).

[9]. UIC, *Code 405 – Links between railway infrastructure capacity and the quality of operations, 1st edition*, UIC (International Union of Railways), Paris, 1996.

[10]. Kaas, A., *Metoder til beregning af jernbanekapacitet (Methods for Calculating Railway Capacity)*, Dissertation, Institut for planlægning, DTU, Technical University of Denmark, Rapport 6, 1998. (In Danish.)

[11]. UIC, *Code 406 – Capacity, 1st edition*, UIC (International Union of Railways), Paris, 2004.

[12]. Almkvist, P., *En jämförelse av två linjekapacitetsmodeller för järnvägstrafik (A Comparison Between Two Line Capacity Models for Rail Traffic)*, Linköping University of Technology, 2006. (In Swedish.)

[13]. Ekman, J. & P. Kreuger, *En analytisk metod för utredning av kapacitet vid signalprojektering (An Analytical Method for Evaluation of Capacity when Projecting Signalling)*, SICS Technical report T2003:11, SICS Swedish Institute of Computer Science, 2003. <http://www3.banverket.se/raildokuffe/pdf/MC0857.pdf> (07-06-30). (In Swedish.)

[14]. Forsgren, M., *Computation of Capacity on Railway Networks*, MSc Thesis, KTH, Stockholm, 2003.

[15]. Burkolter, D. M., *Capacity of Railways in Station Areas using Petri Nets*, Dissertation ETH, Zürich, 2005. <http://e-collection.ethbib.ethz.ch/show?type=diss&nr=16254> (07-10-01).

-
- [16]. Yuan, J. & I. A. Hansen, *Optimizing capacity utilization of stations by estimating knock-on train delays*, In: Allan, J., C. A. Brebbia, R. J. Hill, G. Sciuotto & S. Sone (Eds.), *Computers in Railways IX*, Vol. 15, WIT Press, Southampton, pp. 593-602, 2004.
- [17]. Peeters, L., *Cyclic Railway Timetable Optimization*, PhD Thesis, Erasmus University of Rotterdam, 2003. <http://www.irim.nl> (07-06-30).
- [18]. Curchod A. & L. Lucchini, *CAPRES General Description of the Model*, SBB CFF FFS Swiss Federal Railways & École Polytechnique Fédérale de Lausanne, 2001. http://www.fasta.ch/capres/capres_description_english.pdf (07-06-30).
- [19]. Rudolph, R., *Allowances and margins in railway scheduling*, Proceedings of the World Congress on Railway Research, Edinburgh, pp. 230-238, 2003.
- [20]. Morimura, T. & M. Seki, *The course of achieving 270 km/h operation for Tokaido Shinkansen – Part 1: technology and operations overview*, Journal of Rail and Rapid Transit, No. F1, pp. 21-26, 2005.
- [21]. Andersson, M., *Empirical Essays on Railway Infrastructure Costs in Sweden*, Dissertation, KTH, Stockholm, 2007.
- [22]. Herrmann, T., *Stability of Timetables and Train Routings through Station Regions*, Dissertation, ETH, Zürich, 2005. <http://e-collection.ethbib.ethz.ch/cgi-bin/show.pl?type=diss&nr=16398> (07-10-01).
- [23]. Ceylan, H. & M. G. H. Bell, *Reserve Capacity for a Road Network under Optimized Fixed Time Traffic Signal Control*, Intelligent Transportation Systems, Vol. 8, pp. 87-99, 2004.
- [24]. Kim, Y., K. Kim & D. Shin, *Delay Analysis Method Using Delay Section*, Journal of Construction Engineering and Management, Vol. 131, No. 11, pp. 1155-1164, 2005.
- [25]. Banverket, *Orsakskoder i TFÖR (Cause codes in TFÖR)*. Editors: T. Johansson/Å. Lundgren, Banverket, Borlänge. Dated 1999-06-14. (Pamphlet in Swedish.)
- [26]. Banverket, *Måttbilaga för strategiska mål i SSP 2007-2009 (Appendix of measures for strategic goals in the general strategic plan 2007-2009)*, Banverket, Borlänge, 2006. (In Swedish.)
- [27]. Banverket, *Network Statement 2008*, Banverket, Borlänge, 2007. <http://www.banverket.se> (08-02-05).
- [28]. Walters, M., *Performance measurement systems – a case study of customer satisfaction*, Facilities, Vol. 17, No 3/4, pp. 97-104, 1999.
- [29]. Noland, R. B. & J. W. Polak, *Travel time variability: a review of theoretical and empirical issues*, Transport Reviews, Vol. 22, No. 1, pp. 39-54, 2002.
- [30]. Uher, R. A., *Transportation System Assurance: Terminology, Performance Indices, Failure Assessment and Maintenance Management*, Carnegie Mellon University, 1975.

-
- [31]. Goverde, R. M. P. & M. A. Odijk, *Performance evaluation of network timetables using PETER*, In: Allan, J., E. Andersson, C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Eds.), *Computers in Railways VIII*, WIT Press, Southampton, pp. 731-740, 2002.
- [32]. Krueger, H., *Parametric Modeling in Rail Capacity Planning*, In: Farrington, P. A., H. B. Nembhard, D. T. Sturrock & G. W. Evans (Eds.), *Proceedings of the 1999 Winter Simulation Conference*, pp. 1194-1200, 1999.
- [33]. Vromans, M. J. C. M., R. Dekker & L. G. Kroon, *Simulation and railway timetabling norms*, In: Allan, J., C. A. Brebbia, R. J. Hill, G. Sciutto & S. Sone (Eds.), *Computers in Railways IX*, Vol. 15, WIT Press, Southampton, pp. 685-694, 2004.
- [34]. SS-EN 50126, *Järnvägsanläggningar – Specifikation av tillförlitlighet, funktionssannolikhet, driftsäkerhet, tillgänglighet, underhållsmässighet och säkerhet (RAMS) / Railway specifications – The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)*, SIS, Swedish Standards Institute, Stockholm, 1999.
- [35]. Sherwin, D. J. & A. Bossche, *The Reliability, Availability and Productiveness of Systems*, Chapman & Hall, London, 1993.
- [36]. IEC 61703:2001-09, *Mathematical expressions for reliability, availability, maintainability and maintenance support terms*, International Electrotechnical Commission.
- [37]. Zoeteman, A. & P. R. de Weijts, *The HSL South Infraprovider Contract: contents, process and results*, In: *Innovations for a cost effective Railway Track*, ProM@in, No. 3, pp. 33-37, 2003. <http://www.promain.org> (08-02-12).
- [38]. Lindbäck, B., *Kapacitetsanalys av tågtrafik mellan Boden – Luleå (Capacity Analysis of Train Traffic Between Boden and Luleå)*, MSc Thesis, Luleå University of Technology, 2000. (In Swedish.)
- [39]. Andersson, E. & M. Berg, *Järnvägssystem och spårfordon (Railway Systems and Railway Vehicles)*, Järnvägsgruppen, KTH, Stockholm, 2001. (In Swedish.)
- [40]. Schmöcker, J.-D., S. Cooper & W. Adeney, *Metro Service Delay Recovery: Comparison of Strategies and Constraints across Systems*, *Transportation Research Record*, Vol. 1930, pp. 30-37, 2005.
- [41]. Gunderson, L. H. & L. Pritchard, *Resilience and the behavior of large-scale systems*, Island Press, Washington, 2002.
- [42]. Larsson, M. B.-O., *The Network Performance Assessment Model: a new framework of regulating the electricity network companies*, Licentiate thesis, Department of Electrical Engineering, KTH, Stockholm, 2005.
- [43]. SS-EN 50160:2000, *Spänningens egenskaper i elnät för allmän distribution / Voltage Characteristics in Public Distribution Systems*, SIS, Swedish Standards Institute, Stockholm, 2000.
- [44]. Aven, T., *Availability Evaluation of Oil/Gas Production and Transportation Systems*, *Reliability Engineering*, Vol. 18, pp. 35-44, 1987.

-
- [45]. *NORSOK Standard Regularity Management & Reliability Technology Z-016 Rev. 1*, Norwegian Technology Standards Institution, Oslo, 1998. <http://www.nts.no/norsok> (07-06-30).
- [46]. Pettersson, L. & G. Sfiris, *Förstudie om uppföljning av tillgänglighet inom värme- och kraftvärmeteknik (Pre-study for Availability Follow-up Within Heating and Combined Power and Heating Technology)*, Värmeforsk, Stockholm, 1997. (In Swedish.)
- [47]. Alvi, J. S., *Availability analysis of three different types of combined cycle power plants*, Dissertation, Chalmers, Göteborg, 1993.
- [48]. Hautanen, U., *Utvärdering av Vattenfalls nätinformationssystem (Evaluation of Vattenfall's Network Information System)*, Institutionen för Infrastruktur, KTH, Stockholm, 2002. (In Swedish.)
- [49]. Nordel, *Driftstörningsstatistik – Fault statistics. 2003.*, Nordel, c. 2004. <http://www.nordel.org> (06-06-30). (In Swedish.)
- [50]. Nordel, *Nordels riktlinjer för klassificering av driftstörningar, 2. utg. (Nordel's Policies for Classifying Disturbances)*, 2002. <http://www.nordel.org> (06-06-30). (In Swedish.)
- [51]. Billinton, R. & R. N. Allan, *Reliability assessment of large electric power systems*, Kluwer Academic, Boston, 1988.
- [52]. IEEE Std 1366-2003, *IEEE Guide for Electric Power Distribution Reliability Indices*, IEEE Power Engineering Society, New York, 2004.
- [53]. Svensk Energi, *Elåret 2005 (The Power Year 2005)*, Stockholm, 2006. <http://www.svenskenergi.se> (06-06-30). (In Swedish.)
- [54]. Chowdhury, A. A. & D. O. Koval, *Considerations of relevant factors in setting distribution system reliability standards*, IEEE Power Engineering Society General Meeting, Vol. 1, pp. 9-15, 2004.
- [55]. IEEE P1366-1998, *IEEE Trial-Use Guide for Electric Power Distribution Reliability Indices*, IEEE Power Engineering Society, New York, 1999.
- [56]. Kececioglu, D., *Maintainability, availability and operational readiness engineering handbook. Vol. 1*, Prentice Hall, Upper Saddle River, N.J., 1995.
- [57]. Avižienis, A., J.-C. Laprie, B. Randell & C. Landwehr, *Basic Concepts and Taxonomy of Dependable and Secure Computing*, IEEE Transactions on Dependable and Secure Computing, Vol. 1, No. 1, pp. 11-33, 2004.
- [58]. Zhu, J., J. Mauro & I. Pramanick, *Robustness Benchmarking for Hardware Maintenance Events*, Proceeding of the International Conference on Dependable Systems and Networks, June 22-25, San Fransisco, CA, pp. 115-122, 2003.
- [59]. Nyström, B., *Consistency of Railroad Delay Attribution in Sweden – Measurement and Analysis*, submitted to an international journal, 2007.

Paper V

A Methodology for Measuring the Quality of Deviation Reporting – Applied to Railway Delay Attribution

Abstract

Purpose – The purpose of this paper is to describe a methodology based on vignettes and its application to measure the consistency of railway delay attribution, in order to report on experiences of using the methodology in this context.

Design/methodology/approach – A survey is used to ask personnel performing delay attribution about how they would report the delays described in the vignettes of the survey. The development, application and analysis of the survey are thoroughly described in this paper.

Findings – The methodology proved useful for measuring the consistency of delay attribution.

Research implications – The methodology also supports a further investigation of the accuracy of delay attribution systems.

Practical implications – A survey similar to the one presented in this paper can be used by railways to estimate the accuracy of their delay attribution systems, as well as to continuously improve them. Changes to computer software and training will improve the delay attribution system under study.

Originality/value – By investigating current delay attribution practices using surveys based on vignettes, drawbacks of the current delay attribution system can be identified and remedied. The methodology is applicable to a wide class of deviation attribution applications.

Keywords delay attribution, deviation classification, vignette, survey, railway, punctuality

Paper type Research paper

Introduction

Several European railways are in a state of transition from a single governmental body responsible for the railway to several bodies, governmental, otherwise publicly funded or private. This new operating environment, together with the wish to improve punctuality, calls for new and better ways to support the improvement efforts. The requirements on a delay attribution system vary among different stakeholders, including Train Operating Companies (TOCs), the railway Infrastructure Manager (IM) and Maintenance Contractors (MCs), see Nyström and Karlsson (2006). The Swedish Rail Administration, Banverket, is the IM. Apart from specific requirements, the stakeholders consider it important for a delay attribution system to be consistent. This means that identical delays are reported using the same delay attribution code, regardless of who is reporting. High consistency in a delay attribution system makes it easier to find the causes of delays and develop suitable actions to avoid them in the future.

In general, organisations apply tools and methodologies that use deviations from the specified operation as a basis for improvements, e.g. Failure Reporting Analysis and Corrective Action Systems, FRACAS, (Hallquist and Shick, 2004), Reliability-Centered Maintenance, RCM, (Moubray, 1997), Total Productive Maintenance, TPM, (Nakajima, 1988), Condition-Based Maintenance, CBM, (Williams *et al.*, 1994) and Six Sigma (Raisinghani *et al.*, 2005).

However, the usefulness of these tools and methodologies relies on the quality of the input data. Before detailing the railway environment of this study on delay attribution, it is noted that the consistency of deviation reporting is also an issue for activities other than the railway and that the proposed methodology might be useful there as well. An example is manufacturing, where machine operators are required to attribute deviations – a work task that

might be seen as secondary to the operators' 'real work' and therefore not subject to management attention, training or control. Another example is the description of accidents at work or traffic accidents, which happen in many different places, and therefore can not be immediately attributed by a single person or body. Yet another application area of the methodology might be the classification of reports to the police. Obviously, courts do the ultimate classification of crimes, but the reports, received by police officers, might be useful for calculating the proportions of different types of crimes resulting in a conviction or for allocating resources to districts.

For a system with a long life, such as the railway, it is essential to know "what went wrong" and learn from that, as well as to benchmark, i.e. to find and learn from the best performer. Measurements regarding railway infrastructure costs, failures, customer satisfaction and delays are carried out on a regular basis. One important source of information for improvement work is the existing database of attributed delays. For the Swedish railway, delay attribution normally has to be performed when the extra delay, that is, the increase in delay, exceeds five minutes between two stations.

Banverket controls the trains from eight regional Centralised Train Traffic Control offices (CTTCs) and many Local Train Traffic Control offices (LTTCs), administratively divided into six geographical areas. The numbers of front line personnel at the train traffic control offices range from a few at LTTCs for lines with sparse traffic for only part of the day, to more than 140 at the Stockholm CTTC. Each train traffic controller (dispatcher) is responsible for managing the train traffic and track works in a specific part of the railway network. The controller does so by monitoring the train traffic on computer screens and through telephone contact with train drivers and track maintenance teams. When a train is delayed, it might hinder other trains, which then also become delayed. The controller tries to minimise the spread of delays to other trains and thus has to decide which trains (or track works) have to be delayed or cancelled. The computer system notifies the personnel when a train has been delayed, where it has been delayed and by how many minutes. The personnel then have to perform a delay attribution, i.e. describe the delay using a code. This is illustrated in Figure 1. This might be done immediately after the delay has occurred or later, e.g. at the end of the work shift. The controllers and/or, at some CTTCs, other personnel, report the delays into the TFÖR database, either via the TFÖR text-based user interface or the Basun graphical user interface. The data can later be accessed by the delay attributors themselves or by analysts at TOCs, Banverket and MCs.

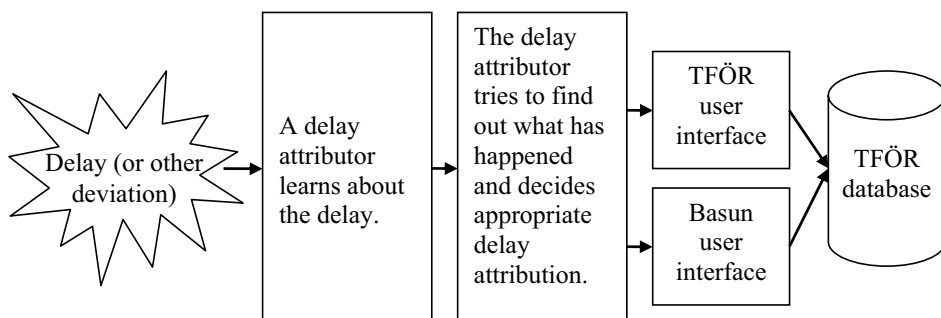


Figure 1. How a delay gets into the TFÖR delay attribution database of Banverket. The two different user interfaces are shown.

In the railway sector, there is no generally agreed rationale as regards delay attribution, but reporting delays serves several purposes. The stakeholders have talked about delay attribution for several years, often discussing which code should be attributed to a certain kind of delay. Delay attribution is needed to calculate the penalties that shall be paid by the stakeholder accountable for a delay. Also, analysis of delay data forms part of the basis for maintenance prioritisation, often aimed at improved punctuality. Therefore, the quality of the delay attribution should be high, including consistent delay attribution. But, as with language in general, local ways of describing situations might evolve, which, over time, results in the use of different expressions to describe the same phenomena and ultimately leads to misunderstanding. This is likely if people are spatially separated. Within a distributed organisation such as Banverket, where delay attribution data is used a long time after it has been created, this difficulty is even more pronounced. Furthermore, delay attribution is not the main task of the personnel performing it. Therefore, a multitude of practices has emerged.

When the methodology presented in this paper was developed, Banverket was in the process of replacing the TFÖR text-based user interface to the delay attribution database with the Basun graphical user interface. Currently, a new system for delay attribution, Giant, is being developed, to which this study is meant to contribute.

This paper describes the development and application of a methodology for investigating the consistency of attribution of railway delays. At the core of the methodology is a survey. The survey asks for delay attributors' answers to designed vignettes, i.e. stories describing different delay situations, as detailed later. All delay attributors in the Swedish railway system were asked to indicate how they would report the delays presented in the vignettes.

The outline of this paper is as follows. First, the purpose is given and the consistency concept is discussed. The subsequent section describes the proposed methodology. Thereafter follows a section which discusses findings regarding the methodology, gained by applying it in the Swedish railway. A separate section presents some major findings regarding consistency of delay attribution, from this application of the methodology. The next section outlines a procedure, based on the methodology, to estimate the accuracy of delay data. The paper ends with discussion and conclusions.

Purpose and the consistency concept

In this paper, a methodology to measure the consistency of a railway delay attribution system, consisting of a collection of delay attribution codes and the work of the delay attributors, is presented, applied and discussed. This is done in order to provide information about the experiences of using the methodology in this context. Before describing the methodology, the concept of consistency is discussed in the remainder of this section.

To be effective, a delay attribution system should be consistent among delay attributors and over time. The following hypotheses and associated sub-measures of consistency were created, reflecting the need for consistency among delay attributors.

1. The delay attribution of a single delay attributor has high precision. This is measured as the agreement of attributions of different vignettes, checking the same underlying principle.

2. The delay attribution within a CTTC has high precision. The precision is measured as the correlation of attributions of similar vignettes between different delay attributors, working at the same CTTC.
3. The delay attribution between CTTCs has high precision. The precision is measured as the correlation of attributions of similar vignettes between different delay attributors, working at different CTTCs.
4. Delay attributors judge the importance of performing a certain delay attribution correctly, higher for the vignettes where the delay might be attributed to either railway infrastructure or a TOC, than for the other vignettes. The importance of the attribution of delays which might be attributed to either IM or TOC is stressed by the vertical breakdown of the railway sector into several bodies with different owners. This sub-measure is motivated by the assumption that if personnel consider the delay attribution at hand to be important, it is more likely to be consistently performed.

‘Precision’ should not be understood merely in terms of low variability of delay attributions, but also as the similarity of the codes attributed. What is said above with reference to CTTCs also applies to LTTC areas.

Proposed methodology

In this section, the proposed methodology is presented. The section begins by giving the reasons for using a survey as the preferred method, continues with the survey content and study population, discusses the concept and application of vignettes, extensively discusses the content of the questionnaire and ends with concerns about response rates and reliability & validity.

One might measure the consistency of the delay attribution and examine how well it corresponds to reality in several alternative ways. It might be done by observation of delay attributors’ work, and from this inferring how the delay attribution relates to the delays encountered during the observation period, i.e. reality. However, a generalisation to how delays, other than the ones encountered during the observation period, would be attributed is not possible. Instead, one might pose hypothetical delays and present them to the delay attributors. Such an experiment might be performed in a natural setting with realistic delay scenarios simulated on delay attributors’ computer screens. However, this approach is very cumbersome and difficult to standardise, as e.g., the dispatching affects which delays will occur. As delay attribution is performed in many places, a large sample or a total population survey is desirable, making interviews and observations burdensome methods. For these reasons, a survey was selected.

The proposed and applied methodology is outlined in Figure 2. The three phases of the methodology are questionnaire design, data collection and corroboration & analysis. In this paper, the questionnaire design phase (steps 1-3) is detailed. Experiences drawn from applying the remaining phases, relevant for future surveys, are also presented.

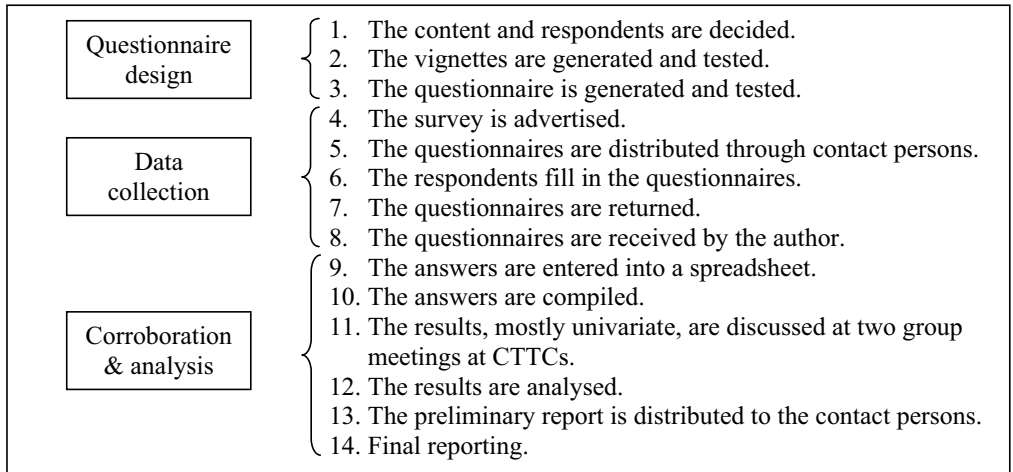


Figure 2. A survey is at the core of the proposed methodology. The questionnaire design (steps 1-3) has an iterative character, data collection is performed in steps 4-8, corroboration & analysis is performed in the remaining steps, 9-14.

Survey content and study population

The survey aims to cover most kinds of delays, but focuses on infrastructure-related delays. The questions range from simple (rather obvious delay attribution) to difficult (hard to tell which code to choose), in order to be able to measure consistency.

The smallest CTTC has 30 delay attributors. There are four different versions of the questionnaire, as motivated in the section “The questionnaire”. This implies that each questionnaire version is completed by seven or eight persons (in the case of no attrition). Each respondent’s answer thus represents up to $1/7 = 14\%$ of the population at a CTTC. The author wishes to be able to detect a delay attribution code that is used by not considerably more than these 14% of the delay attributors at a CTTC. Therefore, it was decided to carry out a total population survey. Furthermore, a total population survey makes it more likely that a way of delay reporting, which is used by only a small proportion of the delay attributors at a small CTTC, will be represented by more than one respondent. This is desirable, as an answer given by only one person is more likely to be a slip or lapse. Thus, it was decided to carry out a total population survey.

Vignettes

A vignette is a story that describes a hypothetical situation. Vignettes, describing occurred deaths, were used by Smith *et al.* (1992) to investigate how death certificates were written in the U.S. Their interviews showed that doctors issuing death certificates often did not understand, and therefore mixed up, the concepts used on the death certificate form; ‘mode of dying’, ‘immediate cause’, ‘underlying cause’ and ‘other significant conditions’.

Another use of vignettes was developed by Martin *et al.* (1991), who read vignettes to informants in order to develop the questionnaire of the U.S. Current Population Survey. They investigated whether the informants considered the persons depicted in the vignettes to be ‘working’. They grouped answers and tried to pin down the underlying heuristic of each informant, e.g. that the informant considered direct cash payment to be the factor deciding whether the person portrayed in the vignette was ‘working’ or not.

Generation and test of the vignettes

Generation and test of vignettes were performed by interviewing train traffic controllers and other personnel at two CTTCs with different kinds of traffic. The purpose was to spawn new vignettes and check credibility and interpretability of vignettes. About 100 vignettes were authored (Nyström, 2006d).

Pre-testing was performed by cognitive interviewing, where advice given in Willis (2005) was useful. Instead of performing think-aloud interviews, which might interfere with thinking, probing was used. Probing, i.e. asking for rationales, was performed after each vignette, e.g. “Please restate the question”, “Is this instruction clear or confusing?”, “What does the word ... mean to you?”, “How do you know?”, “Are you sure?”, “Is this question relevant to you, or it is irrelevant?”, “Does this happen often, seldom or almost never?”, “What does ‘correct’ delay attribution mean here?”. The interviewees gave many useful suggestions regarding the vignettes’ realism, relevance, jargon and applicable rules & regulations. One change was from denoting the trains by letters to applying the convention from training material for train traffic controllers, using 01, 02,... for trains. Each vignette was tested by at least two different interviewees.

The vignette approach and its validity

The delay attributor only perceives fragments of reality, improving his or her understanding, for example, by asking train drivers by phone and discussing with colleagues. Then, the delay attributor uses the computer to perform the delay attribution. However, in this study, the delay attributor is presented to a number of facts, fixed on paper, which are selected from (a hypothetical) reality. Figure 3 depicts the situations with two respondents, defined by subscripts 1 and 2. An event (A) in reality is at the bottom level, while different descriptions of this event are at the two topmost levels. B_1 & B_2 are descriptions in TFÖR, directly generated from reality A. The vignette C is a description of A and D_1 & D_2 are descriptions in TFÖR generated from the vignette C.

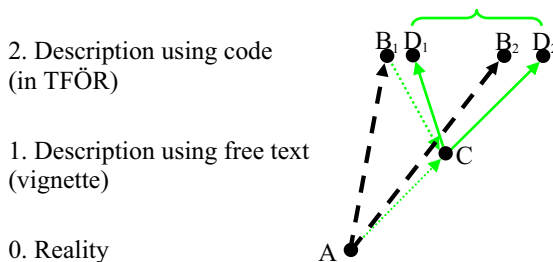


Figure 3. A train delay (A on level 0) and different descriptions of it (found on the levels 1 and 2). The two thick dashed lines show how two delay attributors attribute the A delay when they encounter it in reality. The solid lines show how the two delay attributors attribute the A delay when it is described by vignette C. The brace shows the distance, i.e. the difference between answers that is measured using the C vignette. The dotted lines, from A and B_1 to C, symbolise the use of interviews to generate the C vignette.

In Figure 3, a short distance between B_1 and B_2 implies high consistency of describing delay A. In this study, this is not measured directly. Instead, D_1 and D_2 are used for describing C, thus the distances $B_1 - D_1$ and $B_2 - D_2$ are a measure of the lack of validity of the study. So, a short distance between D_1 and D_2 means high consistency, as the Ds then very precisely describe the same vignette C, which in turn describes exactly one reality, A.

As understood from the above, external validity is a difficult issue. By creating realistic vignettes (C on level 1 in Figure 3) in a wording familiar to the delay attributor, external validity is increased. Situations from reality, as well as TFÖR codes, were used to generate the vignettes.

Testing the questionnaire

The time to fill in the questionnaire is to be kept short to lessen the possibility of slips, lapses or mistakes due to fatigue, as well as for resource reasons. From tests of full-sized questionnaires, the number of vignettes to answer in half an hour – maximum one hour – was set to 14.

To test the questionnaire, interviewees were asked to fill it in and were interviewed afterwards. The interviewees used the TFÖR pamphlet code list (Banverket, 1999) as an aid. Some interviewees spontaneously mentioned the names of codes which no longer exist. The TFÖR codes have been altered over the years, see Cervin *et al.* (1987) for an old edition.

Furthermore, delay attribution codes that were to be anticipated as answers to the vignettes (according to the interviews, documents and judgment of the author) were taken into consideration when distributing vignettes among the four different versions of the questionnaire. The vignettes cover as many different subjects in each questionnaire version as possible, in order to avoid the effect, pointed out in Willis (2005), of the respondent thinking that the answer to this question shouldn't be identical to the preceding one.

The questionnaire

The layout of the questionnaire is now described. The questionnaire consists of some introductory questions, questions on the vignettes and finally questions on opinions. Practical advice on questionnaire layout is found in Bourque and Fielder (1995). See Nyström (2006a) for the four versions of the questionnaire.

The front page instruction tells the respondent, for instance, to have the TFÖR pamphlet handy as a reference.

The first questions, numbers 1-6 of the questionnaire, deal with the respondent's working position and how delays are reported.

In question numbers 7-20, vignettes are given. In order to allow for large content, while keeping the questionnaire reasonably brief, four different sets of vignettes were formed. The resulting four different versions of the questionnaire were randomly distributed to the respondents.

The final questions, numbers, 21-25, are on the respondent's general opinion about the importance of delay attribution. Finally, space is left for general comments.

The introductory questions

Questions 1-6 ask about the respondent's work position, number of years of experience in current position and in the railway sector. Controllers at LTTCs are instructed to answer one extra question, in order to classify the workplace. Respondents are also asked how frequently and via which software they report delays (whether TFÖR and/or Basun).

The vignette questions

Questions 7-20 consist of one vignette each, with accompanying questions. The respondent writes one of TFÖR's delay attribution codes and sometimes writes free text. See an example in Figure 4.

20. Train 01 waits for train 03 because of their scheduled exchange of drivers. Train 03 is 20 minutes late due to a signalling fault (this has previously been coded into TFÖR), causing 01 to leave 18 minutes late from the exchange. How do you code this?

Train 01:

Code: _____ Causing train number: _____ Event report: _____

How important do you judge the correctness of this coding? Mark the most appropriate box.

Not important at all Rather important Important Very important Don't know

Comment: _____

Figure 4. Example of a vignette with accompanying questions. The questions concerning each vignette are all stated in the same way. The vignettes involving several trains might have more questions. (The vignette is shown shrunk from A4-wide and translated from Swedish.)

The stems for each vignette are on:

- Code: how one or several of the trains in the vignette are to be coded into the TFÖR database. In practice, this is a closed question, although the delay attributor writes the code. A TFÖR code consists of one letter and two digits. For example, 'L21' stands for 'Meeting', which means that the delayed train has suffered from waiting for another, oncoming, train. This happens frequently on single-track railway lines, forcing one train to stand still at a station before continuing its journey. 'L21 Meeting' is one of about 100 TFÖR codes.
- Causing train number: is to be filled in when the train is hindered by another train, resulting in a 'secondary delay'. Other delays are called primary.
- Event report: free text might be written in an event report.
- Importance: the importance of correct coding in this case.

The respondent, although discouraged by the limited time and space available, also has the opportunity to comment on the question. In three of the vignettes, the respondent is also asked to state the lengths of the delays.

The vignettes span several dimensions (presumably underlying variables). Each dimension is investigated by using one or several vignettes. The dimensions in the four different versions of the questionnaire are given in Table I. For example, it is seen that the dimension 'Root cause vs symptom' spans eight vignettes in questionnaire versions A and C, while it spans six vignettes in the B and D versions, and the dimension 'Allocation on several causes' is investigated using the three vignettes listed lowermost in version B. Table I also shows that the dimension 'Cancelled train' is studied using only one vignette, in version B of the questionnaire (the delay attribution code 'O29 Cancelled train', however, shows up as an

answer elsewhere). The order of vignettes within each version of the questionnaire is randomised, except for the last three vignettes of the D version of the questionnaire, where the respondents are asked to also state the lengths of the delays.

Table I. The contents of the vignettes in the four different versions of the questionnaire (denoted A, B, C and D). The arrows denote identical vignettes in the versions. The vignettes of B18 and D20 are also stated identically, but in D20, the respondent is also asked to provide the number of delay minutes. There are thus 51 different vignettes altogether. Within each version of the questionnaire, the order of the vignettes is not as schemed here, but randomised. Questions 1-6 and 21-25 are identical in all versions of the questionnaire.

Dimension	Vignette position	Questionnaire version			
		A	B	C	D
Root cause vs symptom	7	Infra - rolling stock	Infra - rolling stock	Infra - rolling stock	Infra - rolling stock
	8	Contact wire - pantograph	←	Contact wire - pantograph	←
	9	Wheel - rail	←	Wheel - rail	←
	10	Infra - infra	Infra - infra	Infra - infra	Infra - infra
	11	Rolling stock - rolling stock	Rolling stock - rolling stock	Rolling stock - rolling stock	Rolling stock - rolling stock
	12	Infra fault and to some extent Info you have or do not have	Personnel and Secondary delays	Maintenance	Track work
Next dimension for B and D versions of the questionnaire	13	Root cause - symptom and Not Info you have or do not have	Track work and Secondary delays	Root cause - symptom and Not Info you have or do not have	Terminal station and Maybe computer related
	14	Root cause - symptom and Not Info you have or do not have	Track work and Secondary delays	Root cause - symptom and Not Info you have or do not have	Passengers
Final dimensions of interest	15	Track work and Secondary delays	Train driving (e.g. train turn)	Train driving (e.g. train turn)	Passengers
	16	Track work and Secondary delays	Secondary delays	Train driving (e.g. train turn)	Rolling stock and Personnel
	17	Track work and Not Secondary delays	Cancelled train	Train driving (e.g. train turn)	Train driving (e.g. train turn)
	18	Track work and Not Secondary delays	Allocation on several causes	Secondary delay attributed directly to cause	Delay minutes manually
	19	Maintenance	Allocation on several causes	Secondary delay attributed directly to cause	Delay minutes manually
	20	Maintenance	Allocation on several causes	Secondary delay attributed directly to cause	Delay minutes manually

The questions on opinions

Questions 21-25 ask for opinions on delay attribution:

21. How important do you consider the delay reporting to be in relation to other work tasks? (closed question)
22. How important do you consider the follow-up of the performed delay attribution? (closed and open questions)
23. How important does Banverket consider the follow-up of the performed delay attribution? (closed question)
24. What are the advantages and disadvantages of TFÖR? (open question, asking for maximum three advantages and three disadvantages)
25. What are the advantages and disadvantages of Basun? (open question, asking for maximum three advantages and three disadvantages)

Distributing and receiving the questionnaires

Delay attributors are distributed among few CTTCs with many employees each and many LTTCs with few employees each. Therefore, it was not feasible to let the author distribute and collect the questionnaires individually. For that reason, contact persons, at least one at each CTTC and LTTC area, distributed, collected and sent in the questionnaires.

Data analysis

An ordinary spreadsheet was used to compile and analyse the data (Nyström, 2006c). The analysis began with checking for anomalies, such as respondents answering that they had held their positions for a longer time than they had worked within the railway sector. The vignettes on delays allow the respondent to jot down only a brief code, e.g. 'L21', as an answer. Respondents might occasionally write a code that is irrational or nonexistent. When at least two respondents gave the same answer, it was always considered unlikely that they had written it by mistake.

If a set of codes is given as different answers to a question, it indicates that these codes are used in a similar way. If this set of codes appears in many questions, it deserves extra attention. Two methods of finding such sets of codes are described here. A simple method is to consider all sets of two delay attribution codes, and count the number of questions they have in common. Another method starts with representing the relationship between questions and answers using a binary matrix. There, a '1' element denotes that at least two respondents gave that answer, a '0' element that this was not the case. The elements of the matrix are then grouped by rearranging rows and columns (i.e., questions and answers). This was done employing production flow analysis, which aims to create contiguous areas of '1' elements (Burbidge, 1963). Thereby, the codes that were used interchangeably, for groups of vignettes, were distinguished. The '1' areas in the matrix show which delay attribution codes were used interchangeably (and for which questions) by the respondents as a collective. For the reason detailed in the section "Analysis of interchangeably used delay attribution codes", the former method was used.

A good representation would facilitate the discussion on the codes concerned. One might respond to each of the code sets so attained, either by arguing that the borders between the codes need to be clarified or that the codes should be joined.

The free-text answers to the questions on the opinions on TFÖR and Basun were analysed using the following procedure. A random sample of 50 questionnaires, which also had free-text answers to the questions or comments concerning the questionnaire itself, was taken from the questionnaires received within a certain time after distribution. Using this sample, the author constructed categories for the opinions on TFÖR and Basun, respectively. As the focus was on finding improvement possibilities, a differentiation was made between positive and negative comments when constructing the categories. In order to improve objectivity, another person also classified the answers of the questionnaires in the sample to the categories constructed. Subsequently, the free-text answers of all questionnaires were categorised.

The compiled answers, highlighting the peculiarities of respective CTTCs and LTTC areas, were discussed at two group meetings at CTTCs. A preliminary report (Nyström, 2006b) was thereafter distributed to the contact persons.

Response rate

A high response rate makes it more likely that the answers of the returned questionnaires reflect the whole population. There are many ways of investigating the extent to which the respondents are representative for the delay attributor collective. These include:

- Whether the answers are markedly different on questionnaires returned at different points in time.
- Whether the answers from CTTCs/LTTC areas with low response rates show markedly different answers to the questions on opinions, compared to other CTTCs/LTTC areas.
- Whether current work position, number of years in current work position and sex of the respondents markedly differ from the delay attributor collective.

In this study, the number of persons in different work positions and number of years in current work positions were not known for the delay attributor collective. The questionnaire did not ask for the respondent's sex. Since some CTTCs have a very low number of women, such a question might therefore have caused some respondents to doubt the anonymity of the survey.

Reliability and validity

As the questions are documented and retrievable by other researchers, the reliability of the study is improved. The validity is increased by using congruent questions (identical questions with different words). General introductions to survey methodology, found useful, are in Trost (2001) and Scheuren (2004).

Of the vignette questions (7-20), some are identical in the different versions of the questionnaire, but placed in a different order. The answers to the questions should not be dependent on the version of questionnaire in which they appear; if they were, this would imply a non-random distribution of the versions of questionnaires to respondents or that prior (or subsequent) questions affect the answers.

Findings regarding the methodology

In this section, findings regarding the methodology, gained by applying it, are presented. In order to investigate the consistency of the delay attribution in the Swedish railway system, the methodology described in the preceding section, "Proposed methodology", was applied.

Delay attribution is a work task for approximately 800 persons employed by Banverket all over Sweden. The survey was distributed to all of them. Possible lessons to be learned from this are discussed here.

Response rate and quality of answers

The study's overall response rate was 50%. Tests indicate that the completed questionnaires are representative of the delay attributor collective, except for the largest CTTC, where the response rate was only 5%. Delay attributors say that the high attrition might to a degree be due to other surveys, which have discredited the use of surveys among some of the delay attributors. To use contact persons to administer the survey worked well in most cases. Having the author on site might have increased response rate as well as guaranteed that respondents were given one hour to fill in the questionnaire.

The survey asked "How important do you judge the delay attribution to be, compared to your other work tasks?" The respondents marked 'Not at all important' (7%), 'Quite important' (34%), 'Important' (42%) and 'Very important' (14%). The distribution of answers, showing that delay attributors have quite a positive attitude, indicates that the delay attributors took the questionnaire seriously and are committed to improving delay attribution.

Many respondents left a blank answer when the vignette described a cancelled train. So, when interpreting the data from the completed questionnaires, one has to account for the possibility that a blank answer means 'O29 Cancelled train'. This indicates that cancelled trains, to a large extent, are not reported. Therefore, cancelled trains, as well as deviations other than "mainstay delays", should be included in delay attribution surveys.

The vignettes received between 0% and 15% irrational answers. For some vignettes, this is explained by their lack of relevance to the traffic situations of some respondents. Obviously, this applies to vignettes involving Automatic Train Control (ATC) when the line does not have ATC installed, and vignettes concerning contact wires when the line is not electrified. This correlation is revealed when comparing the answers to such vignettes to the respondents' work positions and CTTC/LTTC areas. Therefore, it was useful to write the CTTC/LTTC areas on the questionnaires prior to distribution.

Analysis of interchangeably used delay attribution codes

When two different delay attribution codes show up as answers to the same vignette, these codes are used in a similar way in some respect. This might be due to the fact that they are used synonymously, that it is unclear or disputed which of them should be used and/or that the vignette is on the borderline between codes. To work out the extent to which different codes suffer from these drawbacks, one might count the number of questions that different sets of codes share. As there are about 100 codes in TFÖR, the numbers of pairs, triplets, quadruples, etc., to look for are numerous. On top of the list of pairs is found the set 'F28 Vehicle damage/faulty wagon another train' & 'Ö13 Other cause', with nine questions common to the codes.

An alternative way of finding sets of two or more codes is offered by Burbidge's (1963) grouping algorithm. However, the algorithm was to little avail in identifying the sets of codes with most common questions, as the most common set (a pair) identified in this way has only five common questions. The algorithm may assist identification of sets larger than two codes.

Reliability and validity check

The answers to the questions that appear in several versions of the questionnaire do not vary more than may be explained by chance. This means that the answers are not influenced by the questionnaire in which they appear, which indicates that the designs of the questions and the questionnaire are satisfactory.

The survey measures how a delay situation is reported by a delay attribution code. Using database studies, one might go in the opposite direction, i.e. check a certain code's true-life application area. According to Nyström (2005), half of the factually reported 'O61 Locomotive personnel missing' represented a train delayed by another driver (train) being late, 1/3 that the driver overslept and the like, and the remainder other causes. This observation presents a way of checking the validity of the compilation of vignettes chosen for the survey, i.e. the extent to which the survey represents the kinds of delays occurring in reality. In the study, 42% of the factually reported 'O61 Locomotive personnel missing' represented a train delayed by another driver (train) being late. This might give an indication of the validity of the survey.

Findings regarding the consistency of delay attribution

In this section, a few findings from applying the methodology in the Swedish railway are presented.

The survey shows that there are quite large differences in how a certain delay is attributed, i.e. the consistency is low. The vignette questions received up to 17 different codes as answers. For most questions, one, two or three codes dominated nationally. However, the low response rate of one CTTC makes conclusions regarding that workplace difficult.

The results of the survey suggest that delay attributors do not position themselves in a specific position in the cause – symptom dimension. This means that, if one wants to take measures in order to improve the consistency of delay attribution, it is not advisable to assume that delay attributors employ such a scale.

A vignette that describes a train arriving late at the location of planned track work is coded in different ways. The delay attribution indicates hesitation whether to report the explanatory cause (i.e. the cause of the train being delayed) or the necessary cause (i.e. track work being done). The survey also shows that certain delay attributors always report train delays due to planned track works as 'I10 Time exceeded', while another group chooses among three other codes.

Another vignette, involving replacement buses, shows that delay attributors disagree on whether to report delayed trains (56%) or delayed passengers (16%). The remainder of the delay attributors do not clearly subscribe to any of these views. A nationwide policy should be established for delay attribution in connection with replacement buses.

A procedure to estimate the accuracy of the delay data

In this section, a procedure to obtain a more accurate picture of the delays, recorded in the delay attribution database, with the aid of the results of the survey, is presented. The

procedure is illustrated by Table II – Table V. To keep the tables small, the codes and the input data (Table II and Table III) are fictive.

First, the actually reported delays are retrieved from the delay attribution database. See the bordered area of Table II.

Table II. Reported delays (number of delays). The values are fictive.

Reported code	Reported delays (number of delays)
Code 1	1234
Code 2	2345
Code 3	456
Total	4035

Secondly, the answers of the survey are used, see Table III. In order to keep the illustration small, there are only three questions and three alternative codes in this example. Assume that the correct answer to the first question is Code 1, to the second question Code 2 and to the third question Code 3. For example, Table III shows that, when the correct answer was Code 1, 70% of the respondents answered Code 1, while 0% answered Code 2 and 30% answered Code 3. Each row adds up to 100%. The procedure requires the number of different codes to be equal to the number of questions. Therefore, if the employed survey has several questions with identical correct answers, it is suggested that an average be calculated.

Table III. Matrix to adjust the delay statistics using the answers of the survey. Vertically, the correct answers to the vignette questions are given, horizontally the distribution of answers to each of these questions. The values are fictive.

	Distribution of codes, according to the survey →			
↓ Correct code	Code 1	Code 2	Code 3	Sum
Code 1	0.70	0.00	0.30	1.00
Code 2	0.00	0.90	0.10	1.00
Code 3	0.00	0.20	0.80	1.00

Then, the columns of Table III are normalised so each column adds up to 1, see Table IV.

Table IV. The matrix of Table III after normalisation of the columns.

	Distribution of codes, according to the survey →			
↓ Correct code	Code 1	Code 2	Code 3	
Code 1	1.00	0.00	0.25	
Code 2	0.00	0.82	0.08	
Code 3	0.00	0.18	0.67	
Total	1.00	1.00	1.00	

The bordered matrix of Table IV is then multiplied by the vector of actually reported delays (Table II). The resulting product, Table V, is an estimation of the distribution of the delays, adjusted using the answers of the survey.

Table V. Adjusted delays (number of delays).

Cause according to code	Adjusted delays (number of delays)
Code 1	1348
Code 2	1959
Code 3	728
Total	4035

This procedure assumes, for example, that the vignettes of the survey used to construct Table III are representative for the delays (and other deviations) that arise in reality. Furthermore, respondents (i.e. delay attributors) experiencing only few delays, for example because of little traffic, affect the figures in Table III as much as respondents with more traffic. From this, one understands that although the adjusted distribution of delay codes is more correct than the reported delay codes, the obtained result is still an approximation of reality. The reported delays of Table II might be compared to the resulting point estimation of the true delays, Table V. The differences between the tables give estimations of the accuracy of the respective delay attribution codes. Comparing Table II and Table V, it is seen that Code 3 has the largest relative difference. So, Code 3 is considerably more common in reality than is seen from the original reporting. When deciding the amount of effort used to avoid the Code 3 delays, the new, higher, estimated number of delays should be considered.

One might use another measure instead of number of delays, e.g. aggregated length of delays, in the input delay data (Table II). The procedure of adjustment remains the same.

Discussion and conclusions

The purpose was to consider consistency of delay attribution, i.e. that similar delays are reported in similar ways. This is essential in order to give decision-makers accurate information, e.g. concerning maintenance issues, a topic which is investigated in Nyström and Söderholm (2008). A detailed description of the results from the survey and their implications for improvement of delay attribution is given in Nyström (2006b) & Nyström (2007).

The methodology applied was useful for investigating differences between delay attributors, although the low response rate of one CTTC makes conclusions regarding that CTTC difficult. The analysis of the survey was rather straightforward, although the corroboration by discussing results with groups and distributing a preliminary report gave relatively little new information. This is probably because of the extensive testing of the vignettes and the fact that the scope of the group meetings was limited to the results from the questionnaires. Applying a grouping algorithm (Burbidge, 1963) did not work out well, probably because the codes and/or questions were too numerous. The final questions, on opinions, asked for absolute judgement on the delay attribution. Therefore, their answers gave little information, which probably would have been retrievable by instead formulating comparative questions like “Who holds delay attribution to be most important, you or Banverket?”.

It was not possible to find out how delay attributors considered the importance of correctly answering the vignettes where the delay might be attributed to either railway infrastructure or a TOC, because the answers regarding their importance differed too little between the vignettes. An approach analogue to the one suggested above on the final questions, on

opinions, might be used in the future to better discriminate the vignettes, if any, that respondents consider most important to code correctly.

The different stakeholders of the railway have different requirements regarding the reporting of delays and other deviations. These include measuring performance, to prevent or reduce the symptom or the cause of the deviation by different means, by using short-term or long-term remedies. Therefore, the reporting system should deal with deviations, not just delays.

In order to learn from country-wide experiences, it is important to standardise the reporting. By including reporting as a part of the vocational training, giving feedback and system support, the consistency of reporting will improve. The majority of the respondents judged delay attribution to be 'Important' or 'Very important', so the majority of the delay attributors are motivated to improve the quality of the reporting.

The methodology was applied to measure consistency among the delay attributors, as of today, but not changes over time or between different delay attribution systems. The route to measure the consistency among delay attributors has been detailed. Measuring changes over time is simply performed by distributing the survey again and comparing the results obtained. The relations between two different delay attribution systems might be investigated by asking the delay attributors to report each vignette according to both delay attribution systems. Future work might also include delving deeper into the relation between delay attributions made in the survey and those that have been factually reported. Also, it remains to apply the procedure to adjust the delay data, given in the preceding section, "A procedure to estimate the accuracy of the delay data". This might be done, for example, in order to find under-reported delay causes that are comparatively undemanding to reduce in frequency.

From this study, it has been seen that the vignette approach was fruitful. The experiences have been described in the section "Findings regarding the methodology". Continuous use of the survey would make it possible to follow the changes in delay attribution over time. This is especially important when changes are made in the delay attribution system, whether to the software or otherwise. Hence, the proposed methodology can be applied in order to evaluate improvement efforts aimed at enhanced consistency of the attribution system. The need to use data that has been created at different times and at spatially separated workplaces is universal. The methodology might therefore be of use to any industry, where deviations are manually attributed.

Acknowledgements

The author would like to thank Banverket and the EU regional fund for financial support. Thanks also to the staff at Banverket, especially the interviewees, whose participation was essential to the study. In addition, the author is grateful to Stefan Karlsson for discussions on research design and performing some of the interviews, Fredrik Johansson for discussion on questionnaire design, as well as Lina Helgesson for categorising free-text answers, all at Luleå University of Technology. Thanks also to the colleagues at the Division of Operation and Maintenance Engineering for useful suggestions on the paper.

References

Banverket (1999), *Orsakskoder i TFÖR (Cause codes in TFÖR)*, Banverket, Borlänge (pamphlet in Swedish).

Bourque, L.B. and Fielder, E.P. (1995), *The survey kit: TSK 3, How to conduct self-administered and mail surveys*, Sage, Thousand Oaks, CA.

Burbidge, J.L. (1963), "Production flow analysis", *Production Engineer*, Vol. 42, pp. 742-752.

Cervin, T., Kannerby, G. and Rydell, O. (1987), *Analys av störningsorsaker för SJ:s interregionala persontrafik (Analysis of causes for disturbances of the interregional passenger traffic of SJ)*, MSc thesis, KTH, Stockholm (in Swedish).

Hallquist, E.J. and Schick, T. (2004), "Best Practices for a FRACAS Implementation", Proc. 2004 Annual Reliability and Maintainability Symposium, IEEE, New York, pp. 663-667.

Martin, E.A., Campanelli, P.C. and Fay, R.E. (1991), "An Application of Rasch Analysis to Questionnaire Design: Using Vignettes to Study the Meaning of 'Work' in the Current Population Survey", *The Statistician*, Vol. 40 No. 3, pp. 265-276.

Moubray, J. (1997), *Reliability-centred maintenance: RCM II. 2. ed*, Butterworth Heinemann, Oxford.

Nakajima, S. (1988), *Introduction to TPM Total Productive Maintenance*, Productivity Press, Cambridge.

Nyström, B. (2005), *Punctuality and railway maintenance*, licentiate thesis, Luleå University of Technology. Available <http://epubl.ltu.se/1402-1757/2005/41/index.html>

Nyström, B. (2006a), *Enkät om rapportering av förseningar (Questionnaire on reporting of delays)*, Luleå University of Technology (in Swedish). Available <http://epubl.ltu.se/1402-1536/2006/06>

Nyström, B. (2006b), *Kort rapport – Enkät om rapportering av förseningar 2006 (Short report – Survey on reporting of delays 2006)*, Diariennr 03345-2006, Luleå University of Technology (in Swedish).

Nyström, B. (2006c), *Data – Enkät om rapportering av förseningar 2006 (Data – Survey on reporting of delays 2006)*, Diariennr 03485-2006, Luleå University of Technology (in Swedish).

Nyström, B. (2006d), *Frågebaser – Enkät om rapportering av förseningar 2006 (Question database – Survey on reporting of delays 2006)*, Diariennr 03485-2006, Luleå University of Technology (in Swedish).

Nyström, B. (2007), "Consistency of Railroad Delay Attribution in Sweden – Measurement and Analysis". Submitted for publication.

Nyström, B. and Karlsson, S. (2006), "Description of railway delays – a study from the views of the stakeholders", Kumar, U., Parida, A. and Rao, R. B. K. N. (ed.): Proc. of the 19th International Congress, Luleå, Sweden, 12-15 June 2006: Condition Monitoring and Diagnostic Engineering Management (COMADEM 2006), Luleå University of Technology, Luleå, Sweden, pp. 701-710.

Nyström, B. and Söderholm, P. (2008), "Selection of maintenance actions using the Analytic Hierarchy Process (AHP): decision-making in railway infrastructure". To appear in *Structure and Infrastructure Engineering*.

Raisinghani, M.S., Ette, H., Pierce, R., Cannon, G. and Daripaly, P., (2005), "Six Sigma: concepts, tools and applications", *Industrial Management & Data Systems*, Vol. 105 No. 4, pp. 491-505.

Scheuren, F. (2004), "What is a Survey?". Available <http://client.norc.org/whatisasurvey>

Smith, A.F., Mingay, D.J., Jobe, J.B. and Weed, J.A. (1992), "A Cognitive Approach to Mortality Statistics", *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 812-817.

Trost, J. (2001), *Enkätboken (The Survey Book)*, Studentlitteratur, Lund (in Swedish).

Williams, J.H., Davies, A. and Drake, P.R. (1994), *Condition-based maintenance and machine diagnostics*, Chapman & Hall, London.

Willis, G.B. (2005), *Cognitive interviewing: a tool for improving questionnaire design*, Sage, Thousand Oaks, CA.

Paper VI

Consistency of Railroad Delay Attribution in Sweden – Measurement and Analysis

Birre Nyström*

Abstract: This paper presents an analysis of how consistently delays are reported in the railroad, i.e. to which extent different persons report similar delays in similar ways. For railroad infrastructure managers, train operating companies and infrastructure maintenance contractors, a high degree of consistency in the delay reporting creates a firm basis for improving the quality of the traffic. A questionnaire has been distributed to all delay attributors in the Swedish railroad. In the questionnaire, vignettes tell fictive, but realistic, stories about train delays. The delay attributors are asked to write how the events described in the vignettes would be categorized when reported into the delay database. The consistency appears to be rather low, as e.g., delay attributors at different sites report considerably differently. The analyses of different kinds of delays are detailed in the paper. It should be possible to improve the consistency of the delay attribution through training and changes to computer aids.

Keywords: Delay attribution; Punctuality; Railroad; Train; Vignette; Survey; Deviation classification

Introduction

Several of the European railroads are in a state of transition from one single governmental body responsible for the railroad to several bodies. Simultaneously, train traffic is increasing (EU 2004), the speeds, axle loads, etc. are increased, and there is a wish to improve punctuality (the Swedish Ministry of Enterprise, Energy and Communications 2005). These factors call for new and better ways to support the improvement efforts. One important source of information for improvement work is the existing database of attributed train delays. Nearly 300 000 delays in the Swedish railroad are reported each year into the TFÖR database. Statistics on the delays might be used for several purposes. This is a general quality indicator (Ackermann 1998) and the losses due to different delays might be quantified (Rietveld et al. 2001). By careful design of the schedule (timetable), the risk of delays might be lowered (Chandesris 2003; Herrmann 2006). To calculate the impact on passengers from different kinds of infrastructure failures in the London underground, Harris and Ramsey (1994) used network simulation. However, in order to efficiently increase availability of the different items in the infrastructure, one also needs to know the frequency of failures and have a good understanding of the causes of the failures. In many cases, however, the underlying causes of a certain delay are not readily found in the data from the delay attribution system (Nyström and Kumar 2003).

The requirements on a delay attribution system vary among different stakeholders, including Train Operating Companies (TOCs), the railroad Infrastructure Manager (IM) and Maintenance Contractors (MCs), see Nyström and Karlsson (2006). Apart from specific requirements, the stakeholders consider it important for a delay attribution system to be

* Doctoral candidate, Division of Operation and Maintenance Engineering & JVTC Luleå Railway Research Centre, Department of Civil, Mining and Environmental Engineering, Luleå Univ. of Technology, SE-971 87 Luleå, Sweden. E-mail: birre_nystrom@hotmail.com

consistent. This means that identical delays are reported using the same delay attribution code, regardless of who is reporting. High consistency of a delay attribution system facilitates identification of causes of delays and suitable actions to avoid them in the future. The delay attribution system might also constitute the foundation for calculating penalties, whereby each stakeholder gets an incentive to prevent delays. The stakeholders have discussed the reporting of delays for several years, often regarding which code should be attributed to a certain kind of delay.

In Sweden, Banverket (the Swedish Rail Administration) is the IM. Banverket also controls the trains from eight regional Centralized Train Traffic Control offices (CTTCs) and many Local Train Traffic Control offices (LTTCs) administratively divided into six geographical areas. The numbers of front-line personnel at the train traffic control offices range from a few at LTTCs for lines with sparse traffic over only a part of the day, to more than 140 at the Stockholm CTTC. Each dispatcher (train controller) is responsible for managing the train traffic and track works, e.g. maintenance, in a specific part of the railroad network. The train dispatcher does so by monitoring the train traffic on computer screens and via telephone contact with engineers (train drivers) and railroad track maintenance crews. When a train is delayed, it might hinder other trains, which then also become delayed. The dispatcher tries to minimize the spread of delays to other trains and thus has to decide which trains (or track works) have to be delayed or cancelled. The computer system tells when a train has been delayed, where it has been delayed, and by how many minutes it has been delayed. Personnel then have to make a delay attribution, i.e. describe the delay using a code. This might be done immediately after the delay has occurred or later, e.g. at the end of the work shift. The dispatchers and/or, at some CTTCs, other personnel, report the train delays into a database called TFÖR, either via the TFÖR user interface or the newly developed Basun user interface. A pamphlet describing the TFÖR codes (Banverket 1999) is available as an aid to reporting delays. The data can later be accessed by the delay attributors themselves or by analysts at TOCs, Banverket, and MCs. Since Banverket is a distributed organization, it is likely that a multitude of delay attribution practices has emerged.

When the study presented in this paper was undertaken, Banverket was in the process of replacing the text-based TFÖR user interface to the delay attribution database with the Basun graphical user interface. Currently, a new system for delay attribution, Giant, is being developed, to which the study presented in this paper seeks to contribute.

The purpose of this paper is to measure the consistency of train delay attribution and, if possible, suggest actions to improve it. It aims to facilitate the improvement of existing delay attribution systems of railroads, as well as the implementation of new ones. At the core of the employed methodology is a survey. The survey asks for delay attributors' answers to designed vignettes, i.e. stories describing different delay situations. All delay attributors in Banverket were asked to indicate how they would report the delays presented in the vignettes.

The outline of this paper is as follows. First, the research questions are given. Next, the research approach is presented and the layout of the questionnaire is described. An analysis of results from the survey follows. The paper concludes with a summary and conclusions section. Appendix A gives the TFÖR delay attribution codes relevant to the paper.

Research questions

In this paper, the consistency of a train delay attribution system, consisting of a collection of delay codes and the work of the delay attributors, is measured. As has been described, high consistency is important for many activities in the railroad sector. In order to answer the research questions;

- How good is the consistency of the delay attribution?
- How can consistency of delay attribution be improved?

different facets of consistency are first discussed. To be effective, a delay attribution system should be consistent between different attributors and over time, i.e. identical delays should be attributed in the same way. Also, different delays sharing an underlying principle should be reported in a consequent way.

The aim of the delay attribution should be to facilitate the understanding of what has happened and its consequences in order to avoid a reoccurrence and be able to take appropriate measures. This might be performed by looking for a fault (of item or individual), considering where more control is needed, or by investigating to determine whether system-wide changes are needed. These different ways of looking at the purpose of the deviation reporting system facilitate remedies of the types; improve reliability of item/individual, introduce more controls against mistakes, and trace the variations in system performance, respectively.

As an example, consider a train which has been held up by a sabotaged light signal and is thus delayed on arrival at the terminal. The train arrives in the midst of the change of work shifts, and de-icing and loading are performed slowly, adding to the delay. As the train is delayed, it misses its assigned train path and its departure is even more delayed. Here, an underlying cause is sabotage and contributing causes are the change of work shifts and that the train is not immediately assigned a train path. When all of this is known, the improvement work can include each of the three types of remedies mentioned.

As will be shown later, there are cases where some delay attributors undoubtedly report incorrectly. However, the author believes that it is more important to be consistent than to be correct, in most cases. This is because high consistency means that the vast majority agrees on what a certain delay code describes, although the delay attribution is not performed 'by the book'.

Research Approach

Delay attribution is performed at many sites and there are only a few delay attributors working at most of these sites. Therefore, a large sample or a total population survey is desirable, which would make interview and observation burdensome to apply in this study. For these reasons, a total survey was selected. Out of about 833 persons performing delay attribution within Banverket, 417 completed the questionnaire, thus giving a response rate of 50%. The response rates vary widely between different CTTCs and LTTC areas. Answers that seems irrational to the author account for between 0% and 15% of the answers to each question. For some vignettes, this is explained by their lack of relevance to the traffic situations of some respondents.

The initial questions, in positions 1-6 of the questionnaire, deal with the respondent's working position and how delays are reported.

In positions 7-20, vignettes are given. In order to allow for a large content, while keeping the questionnaire reasonably brief, four different sets of vignettes were formed. The resulting four different versions of the questionnaire were randomly distributed to the respondents. Some of the vignettes are identical in the different versions of the questionnaire, but placed in different positions. The preceding (or the following) questions might affect the answers given to a certain question. However, an analysis shows that the differences in the answers between the different versions of the questionnaire are no larger than is explained by chance.

The final questions, in positions 21-25, are on the respondent's general opinion of the importance of delay attribution.

See Nyström (2006a) for the four versions of the questionnaire. See Nyström (2006b) or Nyström (2007) for more about the research approach, including the interviews carried out with delay attributors in order to construct the vignettes, as well as regarding the content of the survey. The data from the survey are in Nyström (2006c).

Results

In this section, a selection of the answers to the vignettes is described. First, the most frequently interchangeably used delay codes are described. Then, the cause – symptom dimension and the answers given to specific vignettes are studied. Comments are given on the content of the TFÖR pamphlet, and the questionnaire's final questions, on opinions, are analyzed.

The Answers to the Questions on Vignettes

The questions on vignettes received up to 17 different codes as answers. For individual CTTCs and LTTC areas, the answers were normally dominated by one or two alternative delay attribution codes. However, these delay codes were not always the same at each CTTC/LTTC area, although in most cases, one, two or three codes dominated nationally. This indicates that delay attribution at individual CTTCs/LTTC areas is reasonably consistent, but the consistency over the whole delay attributor collective is considerably lower.

In this section, the answers to the vignettes are analyzed from different perspectives and specific problems are detailed. The first subsection deals with the TFÖR codes that are most often used in the same situation. Subsequent subsections deal with the overall preferences for reporting delays. Then, beginning with absolute delay and cumulative delay, quantification of delays is investigated. Finally, the question whether deviations, or only delays, are reported is answered.

Interchangeably Used Delay Codes

When two different delay codes show up as answers to the same vignette, these codes are used in a similar way in some respect. This might be due to the fact that they are synonymously used, that it is unclear or disputed which of them should be used and/or that the vignette is at the border between the two codes. The pairs of codes with the highest numbers of vignettes in common are the ones that are most often interchangeably used.

For all sets of two delay codes, the ones with the highest number of questions in common are found in Table 1. In Table 1, the number of common questions ranges from 9 down to 7. The number of questions in common is quite high for many sets of two delay codes, e.g. there are 36 pairs that have five or more questions in common. This makes it difficult to clearly

demarcate pairs that are more frequent than the rest. It is seen that ‘Ö13 Other cause’ appears frequently in the pairs, although this code says nearly nothing about what has happened. Parts of the explanation for the high ranking of the code pairs including Ö13 are probably that the respondents thought that the associated vignettes were intricate and that they are located towards the end of the questionnaire. ‘O29 Cancelled train’ is used as an alternative to other codes, although it is possible to attribute two codes in the computer system. The terms employed in Table 1 are explained and all TFÖR codes relevant to this paper are listed in Appendix A.

Table 1. The sets of two delay codes with the top-ten-highest numbers of common answers to vignettes (ranging from 9 down to 7 for the pairs shown in the table).

Codes used interchangeably		The author’s comment on the vignettes concerned	Possible implied overall impact on delay statistics
F28 Vehicle damage/faulty car another train	Ö13 Other cause	These two delay codes were both given as answers to vignettes on the exchange of a faulty train set and renumbering of train. It was also used for a few vignettes where the respondents probably judged the cause to be unclear.	Rolling stock gets a lower share of the delays, compared to the infrastructure.
L50 Incident/accident	L56 Other mishap	In the TFÖR pamphlet, the descriptions of these codes are located under the heading “Incident/accident” and this contradiction probably confuses the respondents. The vignettes about fire by the track and a moose collision are reasonably answered by these codes, but not the vignettes on locomotive problems, catenary wire torn down and grade crossing accident (which should be coded ‘L54 Grade crossing accident’).	The number of accidents is not entirely reliable.
F10 Damage pulling vehicle	Ö13 Other cause	A vignette about a train heard passing with a probable damaged wheel, causing visual inspection of the track, received these codes.	Rolling stock gets a lower share of the delays, compared to the infrastructure.
L56 Other mishap	Ö13 Other cause	There is no obvious explanation for why vignettes about tower (signal box) problems and catenary wire torn down, received any of these codes.	Not possible to identify improvement possibilities in infrastructure or rolling stock.

F14 Pantograph	I40 Catenary fault	This pair shows a well-known conflict of how to code when a catenary wire is torn down. This might be attributed to the train (the locomotive's pantograph) or the infrastructure (the catenary).	The shares of catenary wire tear-downs that are attributed to pantograph and catenary fault, respectively, are not entirely trustworthy.
P10 Track work according to KTRAV	P11 Track work according to BUP	Delay attributors often do not use the information on which sort of plan (long or short term) governs a specific track work.	Less is known about the impact of the track maintenance planning on delays.
L50 Incident/accident	O29 Cancelled train	'O29 Cancelled train' is a consequence or symptom of what L50 describes.	Information on the cause is lacking.
I40 Catenary fault	O29 Cancelled train	'O29 Cancelled train' is a consequence or symptom of what I40 describes.	Information on the cause is lacking.
Ö13 Other cause	Not to be coded	The vignettes are about, e.g., a passenger train that passes a station without making the stop it is supposed to and a train passing a fire by the track. These <u>deviations</u> do not result in delays to the trains.	Information on trains which do not make the scheduled stopovers is lacking.
F10 Damage pulling vehicle	F28 Vehicle damage/faulty car another train	The vignettes of this pair bring up the issue of whether to code the damage to another train as primary delay (F10) or secondary delay (F28).	No bias, as the codes reflect two different ways of saying the same thing.

Table 1 shows which codes delay attributors most often disagree upon. These disagreements lead to information missing or the delay statistics being distorted. This might result in the wrong remedial actions being prioritized. Therefore, the delay codes in Table 1 need to be clarified or redefined.

The Cause – Symptom Dimension

The two or three most popular ways of coding a delay are often cause and symptom, to different extents. Figure 1 shows the answers to a vignette which illustrates this. In the vignette, an engineer has a pain in his stomach and asks to be relieved. In the answers, the code 'O61 Locomotive personnel missing' (26% of all respondents) describes the cause as a circumstance, while the most frequently used code, 'O64 Urgent change of personnel' (60%), describes the cause as an action. Another used code, 'O15 Extra stop-over', is more of a symptom or consequence. The other answers, 'L42 Train traffic control personnel missing', 'Ö12 Police/illness', and 'Ö13 Other cause' are, or are close to being, irrational, and cannot be positioned along the cause-symptom dimension.

In Figure 1, the different colors represent the results from different CTTCs and LTTC areas. It can be seen that the answers of LTTC area Göteborg (hatched) and CTTC Göteborg (chequered) differ substantially, i.e. the consistency is low. From Figure 1 it is seen that for LTTC area Göteborg, O64 is a more frequent answer than is O61, while for CTTC Göteborg, it is the other way around.

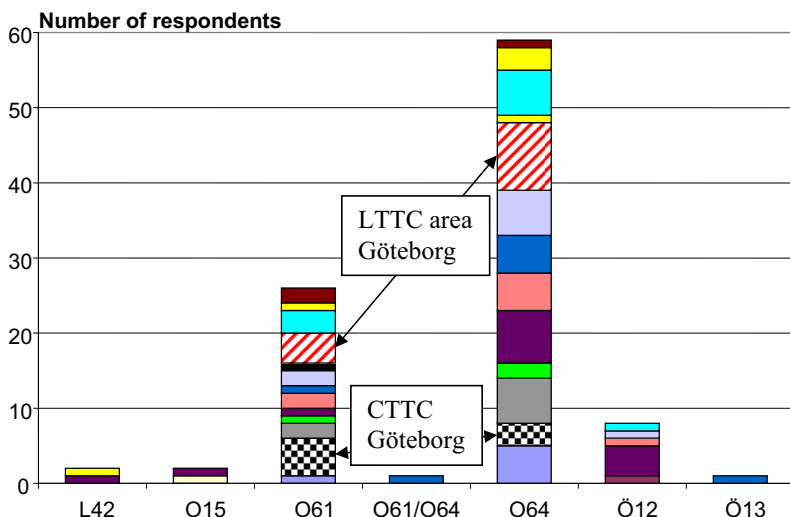


Figure 1. The delay attribution of a vignette where an engineer has a pain in his stomach and asks to be relieved. Most of the answers from the LTTC area Göteborg were ‘O64 Urgent change of personnel’ and a minority ‘O61 Locomotive personnel missing’, while it was the other way around at CTTC Göteborg. The arrows point to these answers. The other answers were ‘L42 Train traffic control personnel missing’, ‘O15 Extra stop-over’, ‘Ö12 Police/illness’, and ‘Ö13 Other cause’. One delay attributor answered that the appropriate code is either O61 or O64.

In this study, it has not been possible to show that the delay attributors prefer to report the delays at a certain place along the cause – symptom dimension, specific to each individual. This might imply that the delay attributors do not consider the cause – symptom dimension, use other dimensions, or, most probably, use heuristics (rules of thumb) that are specific to the type of deviation considered. It is probable that it is extra difficult to understand ‘two-piece’ codes, such as ‘F13 Exchange of locomotive because of damage’, which incorporates both effect (exchange of locomotive) and cause (damage). This might also be why the different codes in connection with wheel damage are often interchanged, even for uncomplicated vignettes. Of all the CTTCs and LTTC areas, only one CTTC demonstrably reports wheel damage correctly.

In vignettes where information needed for proper delay attribution is missing, the performed attribution varies with the situation described in the vignette. For some of these vignettes, a likely delay code is attributed; for others, ‘Ö14 Unknown cause’, is used. For example, a vignette on a ‘ghost train’ (a train is indicated on dispatcher’s computer screen although there

is no train) was coded as either tower (signal box) fault or de-shunting, although it is impossible to tell which delay attribution is correct until it is investigated on site.

The results of the survey suggest that delay attributors do not position themselves in a specific position in the cause – symptom dimension. This means that, if one wants to take measures in order to improve the consistency of delay attribution, it is not advisable to assume that delay attributors employ such a scale. In situations where information is missing, the author recommends reporting what is known at the moment and subsequently adding the result from further investigation.

Different Codes have Different Supporters

When a train is hindered by another, faulty train, the delay attributors might report the delay of the hindered train as either ‘F28 Vehicle damage/faulty car another train’ or ‘L23 Train ahead’. The choices made between reporting F28 or L23 reveal that each code has its supporters. Furthermore, from the survey, it is also clearly seen that the big cities prefer ‘L26 Scarcity of track’ to ‘L23 Train ahead’, probably because it is difficult to point out a specific blocking train at their large stations. These delay attribution codes are equally correct, as L26 refers to the circumstance “too few tracks for the trains”, and L23 the circumstance “too many trains for the tracks”. Another possible explanation is that the big cities are accustomed to reporting ‘L26 Scarcity of track’, due to track maintenance works at their large stations.

A train that arrives late at the location of a planned track work is coded in different ways (according to falling frequency: P10, F10, P11, L15). These TFÖR codes are found in Appendix A. The delay attribution indicates hesitation whether to report the explanatory cause (i.e. the cause for the train being delayed) or the necessary cause (i.e. that the track work is carried out). The survey also shows that certain delay attributors always report train delays due to planned track works as ‘I10 Time exceeded’, while another group chooses among P10, P11 and P12.

As stated earlier, the individual delay attributor does not favor a specific location along the cause – symptom dimension. Here, it is seen that different delay attributors do prefer different codes.

Absolute Delay and Cumulative Delay

Absolute delay refers to the deviation from the schedule. Cumulative delay refers to the increase in absolute delay between two places. Thus, absolute delays are measured at a place, while cumulative delays are measured over a distance. For the Swedish railroad, delay attribution normally has to be performed when the cumulative delay between two stations exceeds five minutes.

Regarding the vignettes for which the respondents were asked to also state the length of the delay, about 90% of the respondents gave the cumulative delay (which is correct, according to the rules) and 10% the absolute delay (which is wrong, according to the rules).

Secondary Delay

A train being hindered by another train is said to suffer a secondary delay. Other delays are called primary. Secondary delays might be reported in several ways. The least informative

way is to report each secondarily delayed train in the same way as the primarily delayed train, i.e. attributing the same primary delay code to each train, e.g. 'I40 Catenary fault' when one train has been delayed by such a fault. The most informative way to report a secondarily delayed train is to attribute the relevant secondary delay code, e.g. 'L23 Train ahead', and include information on the train that immediately delayed it, i.e. the number of the train that was ahead. This makes it possible to identify the train that started the chain of delays. Which of the ways of reporting secondarily delayed trains is chosen is decided more by the situation than by who is performing the delay attribution. Secondary delays caused by infrastructure are more often coded by giving the primary delay code (e.g. 'I40 Catenary fault') than are delays caused by trains.

Sometimes, in delay statistics, the volume of delays is given as only including the primary delays. Assume the volumes of delays caused by the infrastructure and the delays caused by rolling stock are compared by using primary delay codes. Delays caused by infrastructure will then have a larger share of the total delays in the delay statistics than they do in reality. Therefore, the volume of primarily caused delay is a blunt measure for comparing different causes of delays.

Thus, secondary delays are not reported in the same way by different delay attributors. So, caution is needed when interpreting and analyzing primary and secondary delay data.

Train Turn

A vignette on a train that arrives delayed at the terminal station, is turned, gets loaded and departs even more delayed (now with a new train number), results in different delay attributions. The delay attributors' answers differ regarding code (according to falling frequency: O21, O30, O20, O11, O35, O60). Their answers also differ concerning the length of the occurred delay. The lengths of delays written by the delay attributors show that some of them view the departing train as a new train and therefore report absolute delay (the delay relative to the schedule), while others consider the train delayed upon arrival and therefore report only the cumulative delay which was caused by the turning and loading. If one reports the departure delay relative to schedule, it implies that a train that arrives 10 minutes after schedule and departs 10 minutes after schedule generates cumulative delays of 20 minutes in total. This happens, despite that only one delay of 10 minutes occurred. Alternatively, if one reports the cumulative delay for the departure, it implies that the train might leave 10 minutes after schedule and continue to lag that amount of time, without a cumulative delay being reported. So, both ways have their drawbacks.

The author's recommendation is to report deviation from the schedule, as customers suffer in both directions of travel, at least in the case of long-distance passenger trains. A single-commodity train on the other hand, loses only ten minutes of production time. In analogy to this recommendation, renumbering of train (which, like train turn, gives the train a new train number) should mean that the renumbered train does not inherit the delay of the former train number. This was also the result from the survey.

Delays or Deviations

The vignette about a passenger train that passes a station without making the scheduled stop-over and the vignette about a turnout that needs several attempts to be thrown, with the successful attempt just before the train passes, are two stories about deviations. The vignettes do not involve delayed trains, although both describe deviations from what is planned. In the

vignette first mentioned, the passengers are hindered from alighting from the train, in the second vignette, the turnout fails. The answers to the vignettes show that most delay attributors report delays and not deviations. An exception is 'I24 ATC-fault'. The TFÖR pamphlet explicitly says that such an Automatic Train Control fault is to be reported, regardless of whether a delay results or not. Nevertheless, up to 38% do not report this.

A vignette describes that train 03's pantograph gets caught in the catenary wire. A lot of catenary wire is torn down and the track remains closed for eight hours. When the catenary has been repaired, the train is assigned a new train number, 13, and continues its journey. After this, the train suffers no more delays. The most popular answers were 'I40 Catenary fault' (42%) and 'F14 Pantograph' (20%). Remarkably, 2% of the respondents chose not to report. They justified it by the fact that the old train number (03) did not receive a delay, as the train was assigned a new train number (13). As the train got a new, designated, train path, the new train number might not be considered as delayed. The distribution of delay codes for train 03, on different CTTCs and LTTC areas, is shown in Figure 2.

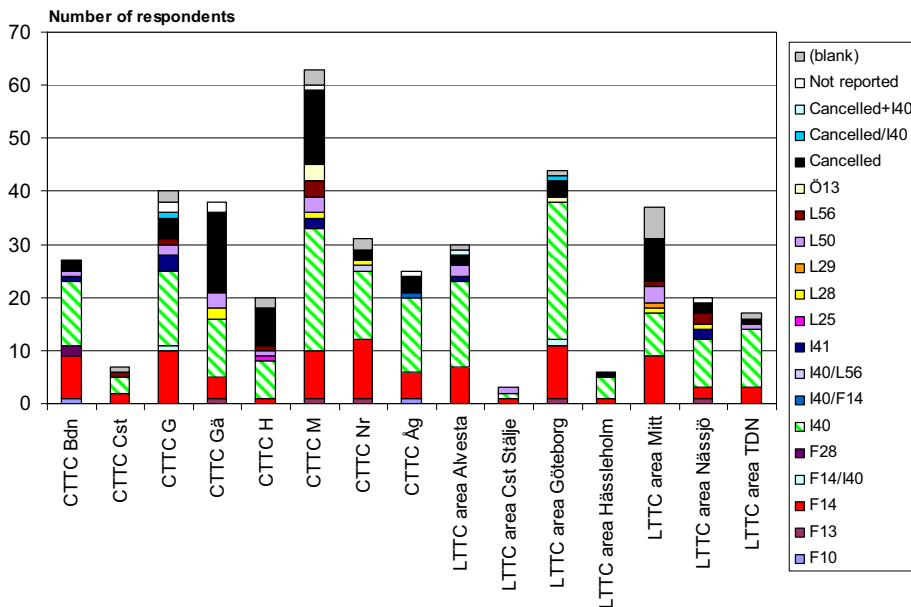


Figure 2. The distribution of the delay codes to a train that is involved in a catenary wire tear-down. Most of the delay attributors chose one of 'I40 Catenary fault' (42% of all respondents, varying from 22% for LTTC area Mitt to 67% for LTTC area Hässleholm) and 'F14 Pantograph' (20%), while 2% (7 respondents) explicitly stated that this is not to be reported in TFÖR. Similar vignettes were given in all versions of the questionnaire, resulting in 407 readable answers, including 19 blanks and one non-existing code (L29).

In order to more easily identify possible differences in delay attribution practice, excerpts of Figure 2 might be made. See Figure 3 for an example. In Figure 3, the answers from CTTC Malmö (the southernmost CTTC) and the LTTC area TDN (the northernmost LTTC area) are shown. Along the y-axis the percentage of answers is shown. Inside the bars, the numbers of answers are stated. It is seen that the delay attributors of CTTC Malmö are more divided than the ones of LTTC area TDN. The most common answer for both CTTC Malmö and LTTC

TDN is 'I40 Catenary fault', but the percentages differ considerably. It is also seen that CTTC Malmö attributes 'Cancelled' to a greater extent than does LTTC area TDN.

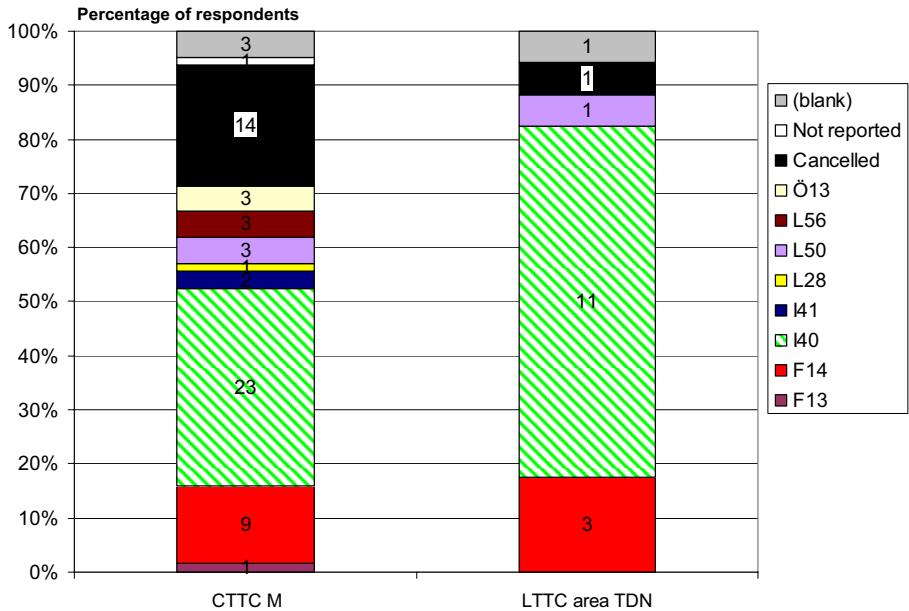


Figure 3. Here, differences in delay attribution practice between CTTC Malmö and LTTC area TDN are seen. For example, the share of respondents attributing the delay to infrastructure (by using the codes I40 and I41) differs considerably. The percentage of respondents is shown along the y-axis and the numbers of respondents are stated inside the bars, summing up to 63 and 17, respectively.

Another vignette, involving substitution buses, shows that delay attributors disagree on whether to report delayed trains (56%) or delayed passengers (16%). The remainder of the delay attributors do not clearly subscribe to any of these views. A nationwide policy should be established for the reporting of substitution buses.

Trains that are cancelled on a part of their scheduled route were coded in many different ways by the respondents, which probably reflects hesitation as to how this should be done. Above all, many missed the opportunity to supplement the code 'O29 Cancelled' with another code. This is possibly because the questionnaire did not explicitly state in each vignette that it was permissible to write two codes, e.g. "O29+I40".

All deviations from the schedule and the requirements on infrastructure and on trains should be reported. The recommendation is based on the fact that a deviation that did not cause any delay might have done so, e.g. if the traffic had been denser.

The TFÖR Pamphlet and the Answers

When reading the description for a code in the TFÖR pamphlet, one often gets the impression that it fits with the delay at hand. Therefore, it is probable that the delay attributors use the code that they encounter first. However, in order to attribute more correctly, one has to read the descriptions for several other TFÖR codes. This phenomenon is seen in the survey, as the respondents have answered with a code whose headline, but not the text under it, fits well with the vignette, or is in a salient spot, or is in the vicinity of the correct code. The text under a headline describes the application area of the code by telling when to use it, when not to use it (and referring to another, applicable, code) or that the code applies in cases when certain other codes do not apply. A general difficulty is that it is sometimes not clear which one of two codes is superior to the other. An example is whether a derailment of a vehicle carrying maintenance personnel, en route to the place of a planned track work, should be coded according to the activity (maintenance) or event (derailment alternatively accident). This is reflected in the answers given, including ‘I10 Time exceeded’, ‘L52 Derailment’, and ‘L50 Incident/accident’.

The Answers to the Questions on Opinions

The questions in positions 21-25 of the questionnaire were on opinions. Most delay attributors gave identical answers to the questions about how important they think the delay attribution is (21), how important its follow-up is (22), and the importance Banverket places on the follow-up (23). This means that the delay attributors believe that they hold the same opinion as Banverket. How important the respondents consider the delay attribution to be (21) is given in Figure 4. The answers to 21-23 follow almost identical distributions, also when each CTTC and LTTC area is considered separately.

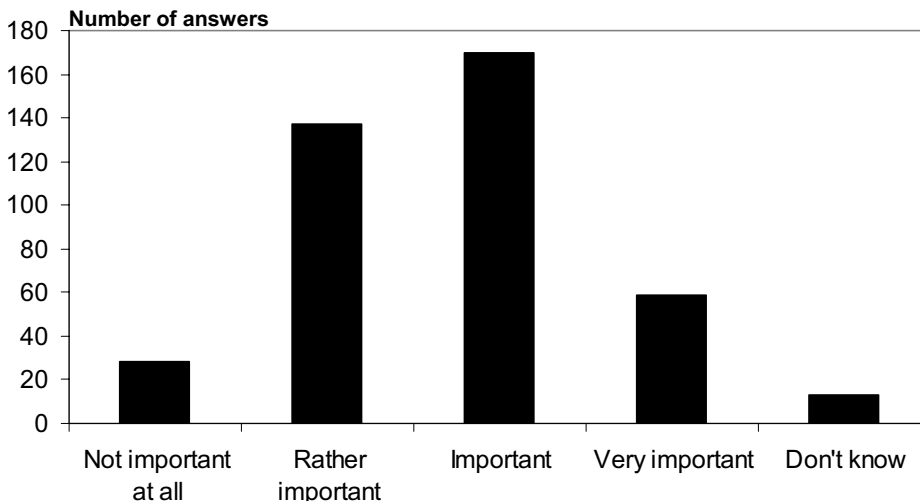
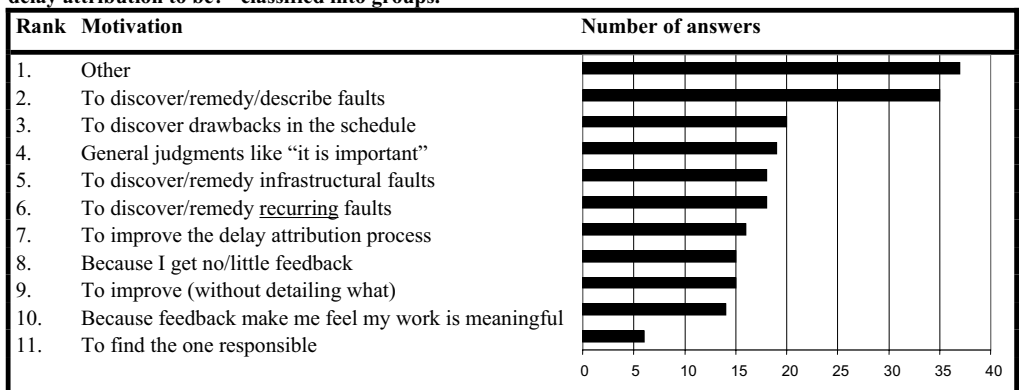


Figure 4. Answers to “How important do you judge the delay attribution to be, compared to your other work tasks?” The majority of the delay attributors answered ‘Important’ (42%) or ‘Very important’ (14%).

Question 22 reads “How important do you judge the follow-up of the performed delay attribution to be?” and gives the respondent the opportunity to write motivations for the

answer. The motivations were classified in groups and the groups were ranked according to the number of answers; see Table 2. Note the multitude of different answers, as many answers belong to the ‘Other’ group. This indicates that the purposes of the delay attribution system are not entirely agreed upon among the delay attributors. The most frequent motivations are “close and immediate” to the respondents in their roles as delay attributors and dispatchers (ranked 2, 3 and 5 in Table 2). Notable is that the motivations ranked 6 and 7, were given to a greater extent by respondents who judged delay attribution to be more important than the average respondent. The answer ranked 8 is a motivation for answering that delay attribution is not so important, being the “opposite” of the one ranked 10. That so few answered that the purpose of delay attribution is “To find the one responsible” (ranked 11) might be considered surprising, as this aspect has recently been increasingly stressed by managers.

Table 2. The motivations of the answer to “How important do you judge the follow-up of the performed delay attribution to be?” classified into groups.



Questions in positions 24 and 25 were about TFÖR and Basun, the interfaces to the delay attribution database. These open-ended questions asked for at most three advantages and three disadvantages of TFÖR and Basun, respectively.

TFÖR (the old mainframe-based text interface) is considered to be equally easy to use regardless of how long the delay attributor has held his/her position. Concerning Basun, ‘easy to use’ is not mentioned as an advantage as often as for TFÖR. That Basun is easy to use is more often written by delay attributors with fewer years in their current position. This is probably because these younger persons are more computer literate (reflecting that Basun is a mainstream Windows program).

The complicated login procedure was a rather common complaint on TFÖR, as the passwords to the computer and the TFÖR program have to be changed by the operator at different intervals. Besides general comments on delay codes or being old-fashioned, spurious complaints concern specific functions such as the event report.

Many respondents stated the low availability of the Basun program, as well as its bugs, as drawbacks. This probably also explains why the respondents say that the availability of TFÖR is high, although delay attributors complained during the interviews about its low availability; e.g., that it is down for planned maintenance each night between Saturday and Sunday. The most appreciated characteristics of Basun are that it is easy to use the program in general, to overview it and to attribute delays, although there were respondents who disagreed. The

advantages mentioned include that scrollbars make it easy to pick a delay code, it is easy to follow several trains simultaneously and right-clicking the train's number shows its telephone number (if the telephone number exists in this system). Disadvantages include difficulty to change an incorrect passage time reading for a specific train.

So, the majority of the delay attributors have a positive attitude towards the delay attribution, although they complain about the software used.

Summary and Conclusions

The stated research questions; “How good is the consistency of the delay attribution?” and “How can consistency of delay attribution be improved?”, considered consistency of delay attribution, i.e. that similar delays are reported in similar ways. This is essential in order to give decision-makers correct information, and thus to reach good decisions, e.g. concerning maintenance issues. The study shows that there are quite large differences in how a certain delay is attributed in different parts of the delay attribution organization, i.e. the consistency is low. However, the low response rate of one CTTC makes conclusions regarding that CTTC difficult.

For most vignettes in the questionnaire, the respondents do not agree on the delay attribution. Instead, they choose, in roughly equal numbers, two, three or four different delay attribution codes. In some cases these codes are identical at different CTTCs/LTTC areas, in other cases, they are not. Very often, about 10% of the respondents give a delay attribution that differs markedly from the rest, which doubtlessly reflects a lack of consistency in the delay attribution. For some of the vignettes, some of the delay attributions are not rational. Such answers are sometimes explained by their lack of relevance to the traffic situation of the respondent. An example is a vignette about Automatic Train Control (ATC), as some delay attributors do not have ATC on their railroad lines.

The survey indicates that the delay attributors use heuristics (rules of thumb) that are specific to the particular kind of situation presented in each vignette, rather than using general rules, as it has not been possible to show that individual respondents choose to answer in a certain range along, e.g. the cause – symptom dimension. Measured in this way, the consistency of the individual delay attributor is low.

The survey respondents believe that the purpose of the follow-up of the completed delay attribution should be more than just “to find the one responsible”. Furthermore, the majority of the respondents judged delay attribution to be ‘Important’ or ‘Very important’. Hence, the majority of the delay attributors are probably motivated to improve the quality of the reporting.

One scope for improvement is the TFÖR pamphlet. Certain delay attribution codes are difficult for the delay attributors to interpret unambiguously, because of the coverage of the codes or the texts describing them in the TFÖR pamphlet. The delimitations of the codes should be made clearer and it should be stated which code has precedence over the other.

Another improvement concerns the feedback given to delay attributors. Currently, the only feedback that is given to the delay attributors at all CTTCs is the percentage of the delays that are coded. This is a weak measure, as the lengths of the delays and the quality of the reporting are not incorporated. The on-the-job training should include regular and individual feedback, e.g. concerning the last work shift or week. The computer system should facilitate immediate

reporting of what is known, provide simple reasonability checks such as that the code ‘Other cause’ always has an event report, that an even (odd) train number has met an odd (even) train number, etc. The system might also assist by suggesting codes based on the current traffic situation.

In order to improve by learning from country-wide experiences, it is important to standardize the reporting. One way to do that would be to include delay attribution as a part of the vocational training for dispatchers and other personnel whose tasks include delay attribution. This, taken together with improved feedback and support from the computer system, will probably lead to greater consistency.

Another possibility for improvement is based on acknowledging that different stakeholders have different requirements regarding the reporting of delays and other deviations. These include measuring performance to prevent or reduce the symptom or the cause of the deviation by different means using short-term or long-term remedies. Therefore, the reporting system should deal with deviations, not just delays.

Future work might delve deeper into the relation between delay attributions made in the survey and those that have been factually reported. Also, delay data has yet to be refined. This might be done by applying the procedure, given in Nyström (2006b), to the data in the TFÖR database and the results from the survey. Such a calculation might reveal, for example, under-reported delay causes that are comparatively undemanding to reduce in frequency.

From this study, it has been seen that the vignette approach was fruitful. Continuous use of the survey would make it possible to track the changes in delay attribution over time. This is especially important when modifications are made to the delay attribution system, whether to the software or otherwise. Hence, the proposed methodology can be applied in order to evaluate improvement efforts aimed at enhanced consistency of the attribution system. The need to use data that has been created at different times and at spatially separated workplaces is universal. The methodology might therefore be of use to any organization in which deviations are manually attributed.

Acknowledgements

The author would like to thank Banverket and the EU regional fund for financial support. Thanks also to the staff at Banverket, especially the interviewees, whose participation was essential to the study. In addition, the author is grateful to Stefan Karlsson for discussions on research design and for performing some of the interviews, Fredrik Johansson for discussion on questionnaire design, as well as Lina Helgesson for categorizing free-text answers, all at Luleå University of Technology. Thanks also to colleagues at the Division of Operation and Maintenance Engineering for useful suggestions on the paper.

References

Ackermann, T. (1998). *Die Bewertung der Pünktlichkeit als Qualitätsparameter im Schienenpersonenverkehr direkten Nutzenmessung [Assessment of punctuality as a quality parameter in rail passenger transport through direct measurement of benefit]*, PhD thesis, Universität Stuttgart (in German).

- Banverket (1999). *Orsakskoder i TFÖR [Cause codes in TFÖR]*. Editors: T. Johansson/Å. Lundgren. Dated 1999-06-14. (Pamphlet in Swedish).
- Chandesris, M. (2003). "Statistical methods for the evaluation of timetable robustness" *Proc., World Congress on Railway Research (WCRR 2003)*, Edinburgh, Scotland, UK, 28 September – 1 October, 637–643.
- EU (2004). *Energy, transport and environment indicators. Data 1991-2001*. (Luxembourg: Office for Official Publications of the European Communities).
- Harris, N. G., and Ramsey, J. B. H. (1994). "Assessing the effects of railway infrastructure failure" *J. Operational Research Society*, 45(6), 635–640.
- Herrmann, T. M. (2006). *Stability of timetables and train routings through station regions*. PhD thesis, Diss. ETH No. 16398, Swiss Federal Institute of Technology, Zurich.
- Nyström, B. (2006a). *Enkät om rapportering av förseningar [Questionnaire on reporting of delays]*. Teknisk rapport 2006:06. Järnvägstekniskt forskningscentrum, Luleå University of Technology. <<http://epubl.ltu.se/1402-1536/2006/06>> (in Swedish)
- Nyström, B. (2006b). *Kort rapport – Enkät om rapportering av förseningar 2006 [Short report – Survey on reporting of delays 2006]*. Diariernr 03345-2006, Luleå University of Technology (in Swedish)
- Nyström, B. (2006c). *Data – Enkät om rapportering av förseningar 2006 [Data – Survey on reporting of delays 2006]*. Diariernr 03485-2006, Luleå University of Technology (in Swedish).
- Nyström, B. (2007). "A methodology for measuring the quality of deviation reporting – applied to railway delay attribution". Submitted to an international journal.
- Nyström, B., and Karlsson, S. (2006). "Description of railway delays – a study from the views of the stakeholders" *Proc., 19th International Congress: Condition Monitoring and Diagnostic Engineering Management (COMADEM 2006)*, U. Kumar, A. Parida, R. B. K. N. Rao, eds, Luleå, Sweden, 12–15 June, 701–710.
- Nyström, B., and Kumar, U. (2003). "Analysis of train delay information" *Proc., World Congress on Railway Research (WCRR 2003)*, Edinburgh, Scotland, UK, 28 September – 1 October, 253–261.
- Rietveld, P., Bruinsma, F. R., and van Vuuren, D. J. (2001). "Coping with unreliability in public transport chains: a case study for Netherlands" *Transportation Research Part A*, 35(6), 539–559.
- The Swedish Ministry of Enterprise, Energy and Communications (2005). *Regleringsbrev för budgetåret 2006 avseende Banverket m.m. inom utgiftsområde 22 Kommunikationer [Regulation letter for the fiscal year 2006 concerning the Swedish Rail Administration, etc. within expenditure area 22 Communication]*. <<http://webapp.esv.se/statsliggaren>> (May 2007) (in Swedish).

Appendix A: TFÖR delay attribution codes

The TFÖR delay attribution codes relevant to this paper are listed in Table 3. The complete TFÖR pamphlet (Banverket 1999) has about 100 entries, of which about 80 were used by the respondents. Most of the codes not used are relevant only to some delay attributors or in special situations. Examples of such codes are ‘Late from abroad’, applicable to trains arriving late from abroad, and ‘Bridge opening’, applicable when a bridge has been kept open too long for boats, thereby delaying trains. The abbreviations in Table 3 are ATC (Automatic Train Control), a safety system, KTRAV, an agreement between Banverket and a TOC (Train Operating Company), which includes the BAP (a plan for track works). BUP is a plan for track works, on a shorter term than BAP.

Table 3. TFÖR delay attribution codes relevant to this paper.

F10 Damage pulling vehicle
F13 Exchange of locomotive because of damage
F14 Pantograph
F28 Vehicle damage/faulty car another train
I10 Time exceeded
I24 ATC-fault
I40 Catenary fault
I41 Zero voltage catenary
L15 Train remains due to KTRAV/BUP work
L21 Meeting
L23 Train ahead
L25 Rerouting
L26 Scarcity of track
L28 Renumbering of train
L42 Train traffic control personnel missing
L50 Incident/accident
L52 Derailment
L54 Grade crossing accident
L56 Other mishap
O11 Express/luggage
O15 Extra stop-over
O20 Train link
O21 Circulation, train turning
O29 Cancelled train
O30 Late departure from freight terminal
O35 Delivered, TOC does not collect
O60 Stationary personnel missing
O61 Locomotive personnel missing
O64 Urgent change of personnel
P10 Track work according to KTRAV (agreement)
P11 Track work according to BUP (plan)
P12 Track work in excess of KTRAV and BUP
Ö12 Police/illness
Ö13 Other cause
Ö14 Unknown cause

Paper VII

Investigating Recurring Faults - A Case Study of Oxmyran Signal Box

Stefan Niska, Birre Nyström

Abstract

Failure reporting systems in the railway industry are reliant on correct reporting into the system, so that correct information is available to the users. This information is needed for correct decision-making in the maintenance process. To base decisions on the fault reporting systems, the remedying technician should have an accurate report. The Swedish Rail Administration has a problem in a signal box, where faults are reported frequently. The fault is electromagnetic interference in the infrastructure and is the target in this case study. The popular belief that lightning is often the cause of this failure is disproved. The results show that the fault is quite difficult to diagnose.

Keywords: Electromagnetic interference, Failure reporting system, Railway signalling, Delay, Reliability

1 Introduction

It is important that the failure reporting systems in any industry supply the correct information to the user [1]. The maintenance managers' decisions are based upon the information that has been entered into the system by the maintenance technicians out in the field or other maintenance staff.

Banverket (the Swedish Rail Administration) has a failure reporting system, Ofelia, to note the numbers of faults and the repair times. This reporting system has been expanded during the period it has been in use. The maintenance management personnel have started to investigate different failures according to the information in the Ofelia database. Insufficient fault diagnosis (i.e., fault recognition, localization and cause identification) give an untrue picture about the behavior of the item of the infrastructure [2]. That this information is as accurate enough is essential if the correct decisions are to be made in the maintenance management.

To make decisions on the basis of data from the failure reporting system, the technician has to be precise in his reporting, to verify that the precise problem out in the process is reported into the failure reporting system. Accuracy in reporting make it possible that the management can make decisions that are relevant to the problems and determine how well items function in the railway system. Therefore, it is important that the managers and the staff consider the complex relationship among items and between departments when they report failures [3].

The railway infrastructure is an old system that is undergoing renewal. This is proceeding slowly and many of the new items are electronic. This is a common development in other industries as well [4]. Where electronic items are used, this makes the system more sensitive to electromagnetic (EM) disturbance [5, 6]. Banverket personnel must therefore learn about EMC (Electromagnetic Compatibility). This is defined as “the ability of a device, unit or system to function satisfactorily in its electromagnetic environment without introducing intolerable electromagnetic disturbances to anything in that environment” [7]. Similarly, contractors must have fundamental knowledge of EMC.

In the Swedish railway, a certain signal box, at the station Oxmyran, see Figure 1, has a component that fails remarkably often compared to similar signal boxes, for example, Öre Älv north of Oxmyran. This incurs unnecessary costs for exchanging the faulty component, as well as train delays. The failures are suspected to be caused by electromagnetic interference. By using the 0felia database, Oxmyran was identified and chosen as a case study and Öre Älv as a reference construction, because it is a similar signal box and the corresponding component has not failed.

Oxmyran signal box and Öre Älv signal box were built in 1995 on the main railway line in northern Sweden. The closest bigger station is Vännäs and the closest town is Umeå, see Figure 1.

The old line made a big turn from Oxmyran via Nyåker to Öre Älv, see Figure 2. The lowest location is Öre Älv and the old track via Nyåker has a high slope. In 1995 Banverket built a new main line direct from Oxmyran to Öre Älv to get a lower slope on the track, but kept the old line, effectively resulting in a double track. Brattsbacka is the adjacent station south on the line, after Oxmyran. The distance between the stations Öre Älv and Oxmyran is 5 km on the new route.

In the study a multitude of possible failure mechanisms are investigated using databases, e.g. of maintenance and traffic, as well ocular inspections, interview with maintenance support management as well as measurements on site, as presented in this paper.

Means of improving the reliability of the signal box are suggested, as



Figure 1: The location of Oxmyran, at the end of the arrow, on a map of Sweden with some of the railway lines and towns.

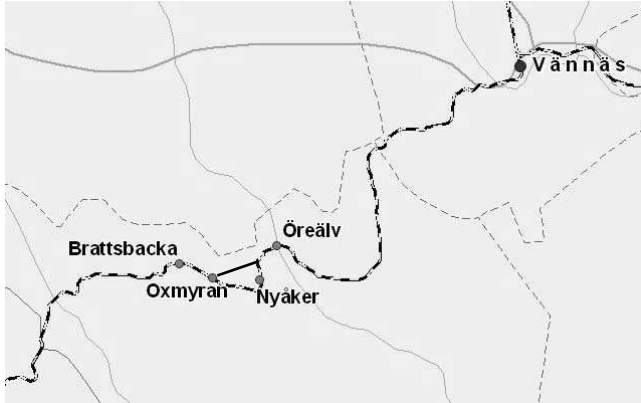


Figure 2: The main railway line from Brattsbacka via Oxmyran, Öre Älv and north to Vännäs. This track was built in 1995 between Öre Älv and Oxmyran, marked as the straight black line between Öre Älv and Oxmyran. The old track over Nyåker remains.

well as ways to improve experiential feedback. The definition of *reliability* in this paper is as follows: “The ability of an item to perform a required function under a given time interval” [8]. The technical personnel must have a good knowledge of EMC if they are to accurately report the faults and sources of the faults. In general, knowledge of the process is becoming increasingly important for the maintenance staff and management, when equipment becomes complex, i.e., electronics [4].

2 Method

The method that has been used is to identify, using Banverket’s fault reporting system Ofelia, a place where similar failures occur frequently due to electromagnetic (EM) interference. Analyze every fault and the specific sources of this faults and the involved technicians are interviewed.

The specific fault is investigated further and a cause-and-effect diagram, see Figure 3, for structuring the possible causes of the fault is made [9].

The authors judge that every factor in the diagram could be a reasonable cause of EMC failure in the infrastructure and in the signal box. After thorough investigation of the faults, possible causes of failure were excluded one by one.

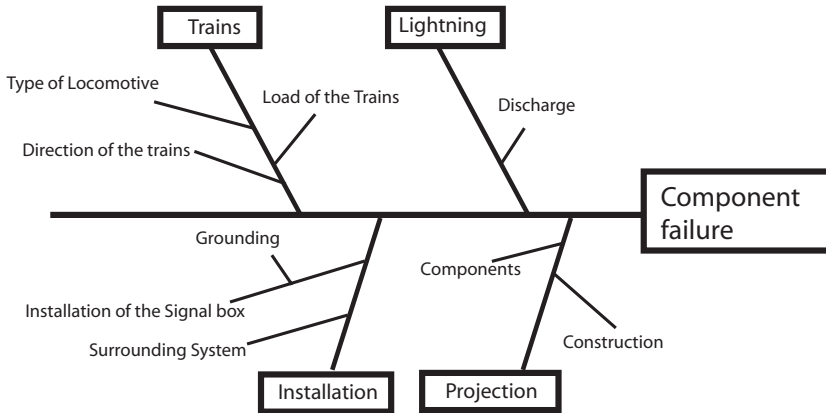


Figure 3: Cause-and-effect diagram over possible causes.

3 System function

The signalling system in the railway infrastructure controls operation and safety; e.g., the light for the locomotive driver, the turnout position, indicates train position on the track, etc. The train positioning system's task is to report where the trains are so as to prevent collisions.

A station is where there are one or several turnouts. On a single-track there are stations to make it possible for trains to meet and overtake, see Figure 4.

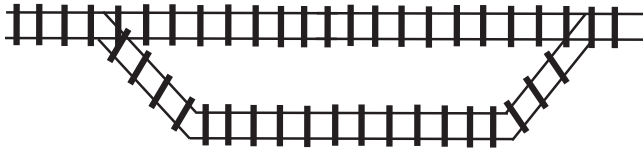


Figure 4: A station is used for meeting and overtaking of trains on a single-track.

Most of the equipment for the signal system is housed near the track in signal boxes. These boxes house signalling equipment and other equipment that is needed by the track or the station. In the signal box there are three types of equipment: electrical, telecommunication and signalling systems. The electrical system delivers power to the other systems, and the telecommunication systems are for communication, e. g. telephone, modem, etc. From the signal boxes, wires go out to the signalling lamps, turnouts and positioning systems in the track. The signals control the safety and inform the

locomotive driver and the train traffic controller, who is in a centralized train traffic control (CTTC) office, of the situation on the track. The cables placed across the track have to send the right information and get correct signals from every position along the track. The signalling system communicates the signals to and from the CTTC office.

3.1 Train positioning system

The train positioning system shows the CTTC office where the trains are on the track. The two rails of a track help to indicate where the trains are. One rail has gaps and there is a potential between the rails when there are no trains on the line, see Figure 5. Every second I-rail (information rail) section has a positive potential of 7 V and the other sections have a negative potential of -7 V. The gaps are an insulation between the I-rail sections. This design ensures that there are no connections over the gaps, between the sections. The other rail is a common rail (S-rail) and is grounded. Therefore, when a train moves into a new section, over a gap to a new rail section on the I-rail side, the potential drops to zero, because the wheels from the train connect the two rails and the section is grounded. The relay connected to this section indicates where the train is in the train positioning system.

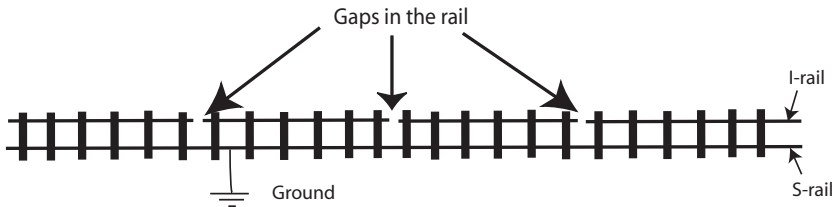


Figure 5: The train positioning system, where one rail, the information rail (I-rail), has gaps to split the rail sections. The other rail, common rail (S-rail), is continuous. Every second rail section has either a negative potential of - 7 V or a positive potential, +7 V. The other rail (S-rail) is grounded. The wheels from the trains connect the two rails and ground the potential and the CTTC office sees in which section the trains are on the track.

3.2 System description

Signal boxes are placed along the track protect the equipment that must be located close to the track. Signals are sent to the signal box via shielded twisted pair cables, see Figure 6. In the signal box there are electrical components, computer systems like PLC (Programmable Logic Controller), and

electrical relays for the equipment in the area controlled by the signal box. Between the signal boxes the electrical signals are sent through the MOK (MellanOrtsKabel, bigger cable between places). This cable, the MOK, is a belted and shielded cable with 27 twisted pairs of copper wires that lead over to the next station. All electrical, telecommunication and signalling systems, use the MOK cable for communication. A bigger signal box (station), in this case Mellansel (MSL), receives the signals, and relays them to the CTTC office, Ånge (ÅG) in this case, via an optical cable, see Figure 7.

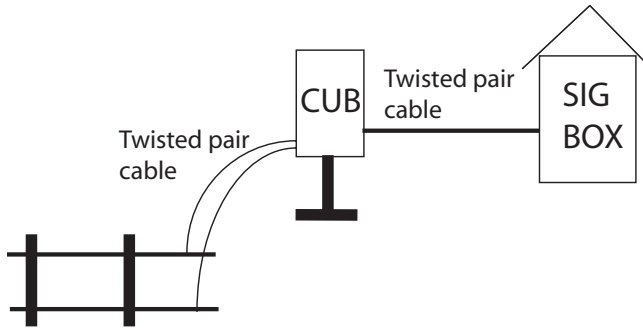


Figure 6: Signals from the track are sent via twisted pair cables over to a trackside box (CUB) which houses the equipment that must be close to the track, and from there to the signal box (SIGBOX) via a twisted pair cable.

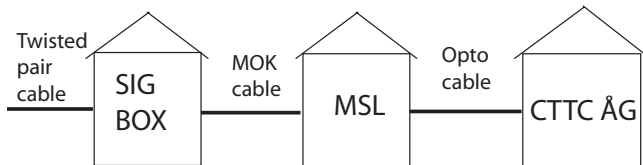


Figure 7: The path of the signals via shielded twisted pair cables from the trackside equipment box to the signal box (SIGBOX). Between different stations there is a belted and shielded cable, with 27 twisted pairs of wires, called the MOK. In a bigger station like Mellansel (MSL) the signals are sent over optical cables to the centralized train traffic control (CTTC) office in Ånge (ÅG).

The train positioning system in the Oxmyran station uses several relays to detect when a train is in the station. When one of the S1a, S1b, or S1c relays gets an indication that there is a train in the station, the RS1-relay gets an indication, see Figure 8. This verifies that there is a signal that indicates that there is a train in the station and makes the system fail-safe.

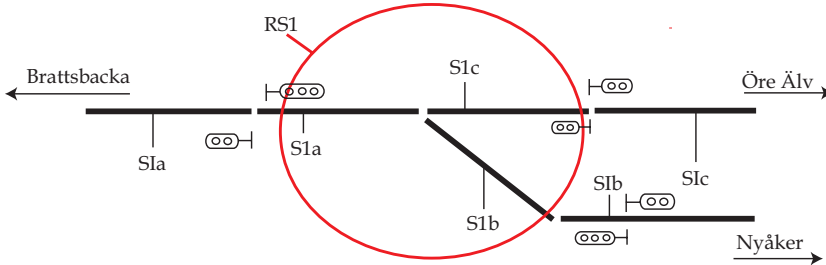


Figure 8: All relays to indicate the train is in the station. The S1a, S1b and S1c relays detect if a train is in the station and the RS1 relay repeats the signal from any one of the S1 relays.

When the train leaves the station, the last axle from the last wagon has passed the section and no axle is grounding the section, the voltages is back to its normal level again and the S relay switches back, indicating that there is no train left in this section. The RS relay also switches back and cuts the power to the chain after the RS relay. All the subsequent relays in the chain need power to switch back. This power, needed by the relay for a short time, is delivered by the component, which has this as its sole function. The component is in Figure 9.

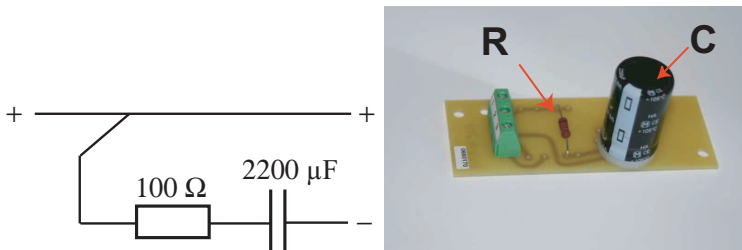


Figure 9: The component is a RC circuit and has a resistance, marked R in the figure and photo, of $100\ \Omega$ which is connected in series with a capacitor, marked C in the figure and photo, of $2200\ \mu\text{F}$.

The component is a printed circuit card with a resistor (R) and a capacitor (C), i.e., an RC circuit. The voltage over the RC circuit is 27 V. A rectifier delivers the power from AC 220 V to DC 27 V to the subsystem where the RC circuit is.

There is a fuse between the rectifier and the printed circuit card. The printed circuit card is connected in both sides with a $0.75\ \text{mm}^2$ wire, see Figure 10. This wire goes to a negative wire loop joining other printed circuit



Figure 10: The placement of the RC circuit in the rack, by the arrow, and the wire from RC components that have the same size both on the negative bound and on the positive side. The area of the wire is 0.75 mm^2 .

cards that have the same function (but in different subsystems) which at the end is grounded.

3.3 The Oxmyran station

Assume a train enters the Oxmyran station from the south, i.e., from Brattsbacka, see Figure 11. Then the relay on section S1a is grounded and the potential drops down to zero. The train positioning system indicates that there is a train in this section.

When the train enters the first section in the Oxmyran station, from the south, the S1a relay drops the potential and affects the RS1 relay, see Figure 12. Therefore, the RC circuit gets power while the train is in the station.

When the train moves over to the next section and the relay S1b or S1c, depending on the direction of travel, it drops the potential to zero and thereby indicates that the train is in this section. The RS1 relay is still holding and makes sure that there is power to the RC circuit, see Figure 13.

When the last wagon of the train leaves the station, in this case, when the last wagon leaves the station section as indicated by S1a, S1b, or S1c relay (in this example S1b or S1c), the RS1 relay drops the power immediately. The next section S1b or S1c, depending on which route the train takes through the station, indicates that the train is in this section now, see Figure 14. The relays after the RS1 relay need power to pull back. This power is supplied by the RC circuit, which can hold the power a few milliseconds.

It is the same procedure if the train comes from the north, but the first relay in the station is S1b or S1c, depending on whether the train comes from

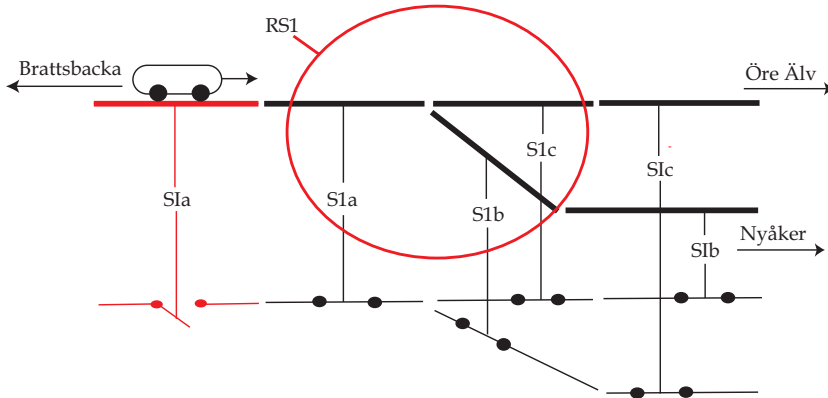


Figure 11: The state of the train positioning system when a train is entering the Oxmyran station from the south via Brattsbacka, the section (relay) before the station has dropped the potential down to zero.

Öre Älv or Nyåker.

4 Failure reporting system

Ofelia (0; zero; felia; faults) is a failure reporting system used by Banverket. Every time a failure occurs on the railway infrastructure the fault is registered in Ofelia. When a fault appears, it is indicated in the CTTC office from alarms via the signal and detector systems out by the track. Maintenance personnel receives a work order to repair or check the alarm. There are several fields in Ofelia that must be filled in; e.g., real fault, definition of the fault, cause of the fault, description of corrective action, remedial action, etc. [10].

4.1 Fault Detection

When the RC component fails, the CTTC office personnel notice on their computer system that the train route on the track still indicates red after the train has left the station, i.e., the signalling system still indicates that there is a train there. Then, it is not possible to set a new train route with the signalling lights for the locomotive driver. When this failure occurs, the CTTC office staff report to Ofelia and calls the contractor's contact person, in turn instructs personnel to remedy the fault. The maintenance workers must report the repair to the CTTC office, which finalises the report in Ofelia [10]. The maintenance worker has to analyze the situation in the signal box

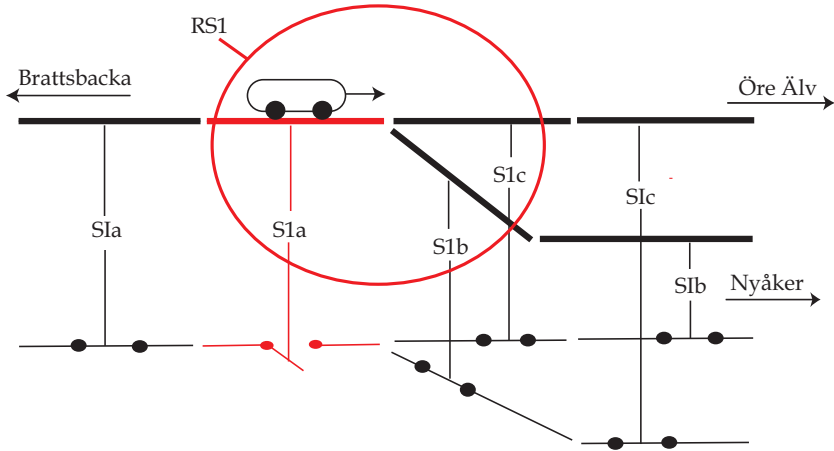


Figure 12: The train is in the Oxmyran station. The relay S1a has dropped potential and therefore influences the RS1 relay.

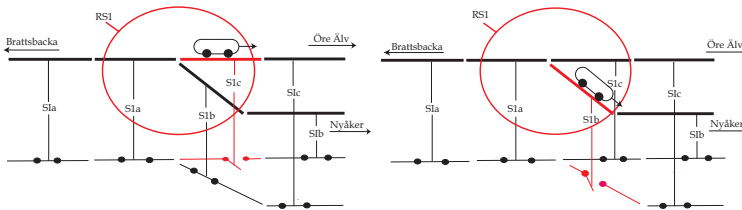


Figure 13: A train is in the Oxmyran station and the S1b or S1c relay has dropped the potential, depending on which way the train is going. The RS1 relay is still activated by the presence of train in the station.

and report the source of the fault and cause of the failure.

After many similar faults, the first check made by maintenance personnel upon arrival at Oxmyran, is the RC circuit. The RC circuit may appear to be unbroken or as in the case in Figure 15, the resistor may be broken.

The maintenance personnel have difficulty getting a new RC circuit because they have only one in stock locally. Getting a new one might take several days. In this case, anticipating that the RC component will fail again, the maintenance staff order more than one RC component and place the spare parts in the Oxmyran signal box.

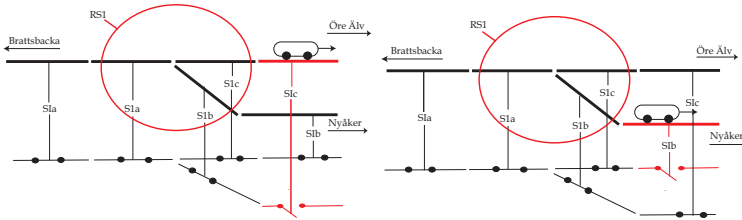


Figure 14: The train is leaving the Oxmyran station and the S1b or S1c relay has dropped the potential, depending on which route the train is going. The RS1 relay is pulled back and drops the power when the last wagon is out of the station.

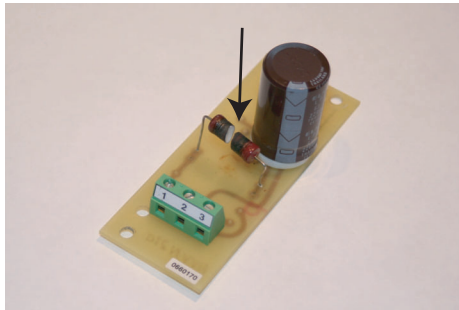


Figure 15: A faulty RC component. The resistor has been overloaded, causing a material breakage, see the arrow.

5 Possible causes

There are many possible causes of this fault. In this article, some of the possible causes are studied in depth to find a cause of the fault. The hypotheses of possible causes are as follows:

- *Lightning discharge*; in Ofelia the reported cause is “thunder”. Lightning is thought to have discharged close to railway infrastructure when the fault occurred in the signal box.
- *Trains*; one type or one specific train sends out disturbance (interference) and is the cause of EM interference.
- *Construction*; big differences between signal boxes and therefore the sensitivity of interference differs considerably.
- *Installation*; different companies have installed the two signal boxes. Therefore, the EM protection is worse in one of the signal boxes.

- *Grounding*; different ways of grounding result in different pathways for the current to travel in the system.
- *Surrounding system*; transformers or a contact wire that is so close to the signal box that it can affect the items in the signal box.
- *Components*; the components in the system do not follow the specifications and are more sensitive than the specification.

5.1 Lightning discharge

Of the 35 reported faults of the RC component in Ofelia reporting system, during the period 1 January 2001 to 1 May 2007, 13 reports have been filed stating the cause as “thunder”, see Table 1. For this to be the cause, lightning has to hit or discharge close to the railway infrastructure to cause interference in the railway infrastructure and its system, especially the RC components system. After the discharge, a train has to pass the location, Oxmyran, in order to CTTC to detect the failure.

The dates and times of the discharges, from clouds to ground, that have occurred [11] should agree with the dates and times that faults have been reported into Ofelia. In this comparison there are six occasions on which the location of the lightning strike has been within a 25 km radius of Oxmyran. The shortest time a fault occurs after a discharge was reported was 40 minutes before the alarm, and the longest is 3 hours and 36 minutes. Five of the faults were reported into Ofelia as “thunder” and one as “uncertain electric disturbance”, see Table 1.

According to the train traffic roll of passing trains in Oxmyran, several trains pass between the time the strike is registered and the alarm goes. On two occasions the alarm is triggered by the first train that passes after the discharge. On these two occasions it could be a discharge, “thunder” or lightning that has caused the alarm; but if it had been a discharge or a strike, the equipment should have been damaged in several places. The energy from a strike should have broken more than one item in the station; therefore, many more components should have been broken in the station.

Lightning occurs mostly during the summer. Oxmyran is not in an area where there is much lightning. According Ofelia, most faults occur during the summer period, but occasional faults occur during the winter.

In this case, it is only slightly probable that “thunder” or lightning has caused the RC component to fail immediately. The damage to the rest of the equipment should have been greater if lightning had caused the occurred failures.

Table 1: The dates of registered, in Ofelia of fault occurrences of the RC components in Oxmyran; registered the actual fault in Ofelia under the column the real fault and registered cause in Ofelia under the cause column.

Date and time	The real fault	Cause
2001-06-18 18:15	Material breakage	Broken component
2001-06-19 00:30	Not able to detect	Broken component
2001-07-04 22:28	Interrupt	Thunder
2001-07-05 09:29	Interrupt	Thunder
2001-08-11 16:13	Material breakage	Broken component
2002-06-07 17:45	Interrupt	Thunder
2002-07-04 08:07	Interrupt	Thunder
2002-07-12 08:42	Interrupt	Thunder
2002-07-16 00:10	Interrupt	Thunder
2002-07-24 09:12	Interrupt	Thunder
2002-07-26 19:37	Interrupt	Thunder
2003-06-25 17:12	Interrupt	Uncertain electric disturbance
2003-07-21 16:24	Interrupt	Thunder
2003-08-15 08:56	Material breakage	Thunder
2003-12-03 06:06	Interrupt	Broken component
2004-05-10 06:01	Material breakage	Thunder
2004-07-17 05:40	Interrupt	Broken component
2004-07-19 09:17	Not able to detect	Broken component
2004-08-20 11:18	Interrupt	Broken component
2004-11-19 06:07	Interrupt	Loosen detail
2004-12-22 16:34	Interrupt	Uncertain electric disturbance
2005-04-09 22:29	Interrupt	Broken component
2005-04-15 05:33	Not able to detect	Broken component
2005-05-25 09:40	Interrupt	Broken component
2005-05-26 07:35	Interrupt	Uncertain electric disturbance
2005-09-04 17:31	Interrupt	Broken component
2005-09-09 22:31	Interrupt	Thunder
2006-06-01 09:25	No fault	Not fault found
2006-06-02 06:01	No fault	Not fault found
2006-06-08 14:54	Not able to detect	Broken component
2006-06-21 07:11	Interrupt	Thunder
2006-07-11 05:45	Not able to detect	Investigated the electronics
2006-09-07 13:33	Interrupt	Uncertain electric disturbance
2006-09-14 13:05	Interrupt	Broken component
2006-10-22 08:04	Interrupt	Broken component

5.2 Trains

The steel train from SSAB (Swedish Steel Inc), in Luleå, uses this route to transport the steel slabs to SSAB in Borlänge, from which sheet metal is made. The steel trains are very heavy and therefore require high power. More power gives higher electromagnetic fields [12]. Could this higher power be a cause of interference in the system and failure of the component? It could also be a specific locomotive that disrupts the equipment at Oxmyran. This depends on the direction of travel of the train, since there is a slope. The steel trains always travel south with a heavy load from Öre Älv to Oxmyran or coming from Nyåker to Oxmyran, but the latter route is seldom used. When a train is starting after having waited for another train, the train needs high power to accelerate and the disturbance and EM interference is larger.

The correlation between the type of train passing the Oxmyran station and the faults that occur there is not that clear. Different types of locomotive, type of train, direction, or weight of the train do not explain the faults. Similar faults do not occur in Öre Älv.

5.3 Construction

Both signal boxes were built by the same company, Banverket's Industry Division, in Nässjö. The signal boxes were transported by truck to the locations in Öre Älv and Oxmyran. They were built according to the same instructions and drawings, and appear upon visual inspection to be alike. The only apparent difference, upon visual inspection, is the location of a wire that is drawn to the ground, see Figure 10. It whether has not been fully investigated to verify every wire or cable in the signal box are in same place.

The construction in Oxmyran and in Öre Älv is similar, as far as has been seen in this investigation. The construction therefore has no impact on the RC component.

5.4 Installation

The same contractor, Banverket Production, installed the signal boxes in both Öre Älv and Oxmyran. There is no information as to whether there were different instructions during the installation of the two signal boxes, according to the installation supervisor. Therefore, all the information in this case indicates that both signal boxes were installed in the same manner.

5.5 Grounding

The grounding system should be similar in both signal boxes, because the same contractor has installed both boxes according to the same drawings and instructions. Since it is difficult to know the location of ground cables from the signal boxes to the S-rail, i.e., the ground in the railway system, there is no guarantee that both the grounding systems are placed or drawn grounding cable in the same way to the S-rail. This is because the ground cable is under the ground surface. The grounding is a complex area and is influenced by many parameters, it is not easy to discount it as a fault source. However, there is no visual indication of a difference in the grounding, according to construction drawing and interviews with project manager for those signal boxes, and ground cable placement and therefore the grounding should not be an obvious possible cause in this case.

The Oxmyran box has been complemented afterwards with a ground loop in the ground around the signal box and iron bar in the ground and it is connected to the S-rail to get a better grounding in Oxmyran, especially for lightning strike.

5.6 Surrounding system

The distances from the overhead contact wire to the signal box at Oxmyran are about the same distance as to the signal box in Öre Älv. The reserve power to Banverket power, if Banverket's power is interrupted, have a transformer about 10 m away from the signal box both in Oxmyran and Öre Älv.

There was no significant correlation between times of reserve power switching on or off and the faults of the RC component. The old database for times when the reserve power has been turned on or off, has been erased, and the new data in the new database does not indicate a relation between time of failure and switching.

5.7 Components

There are six rectifiers in every signal box. On the primary side, each has 220 V and on the secondary side the variation is from 6 V to 220 V. The same company made the rectifiers that were installed in Oxmyran and Öre Älv and the specifications are the same in the subsystems.

Upon visual inspection, the RC circuit is different but the specifications are the same. As mentioned earlier, the placement of wire are not the same in Oxmyran and Öre Älv; see Figure 10, which shows the placement of a wire

from the RC circuit, negative side, in Öre Älv. In Oxmyran it is placed on the opposite side in the cable channel above.

Here, there is nothing to indicate that the new components differ or that differences in the rectifiers could be the source for the fault.

5.7.1 Quality assurance

Banverket conducts no quality inspection of components before they are taken into operation on the system. They rely on suppliers to do a good job and build RC circuits according to the correct specifications.

Every rectifier is tested and inspected by the manufacturer.

6 Measurement on site

Measurements have been taken to determine if there are any differences between the systems in Oxmyran and Öre Älv. The measurements were taken in the signal box in Oxmyran, and the signal box Öre Älv was measured as a reference. A laptop computer and an NI USB-6251 DAQ (data acquisition) unit from National Instruments was used. The software that was used to control the measurements and collect the data from the DAQ was LabView from National Instruments. The DAQ has an input span of -10 V to +10 V and could not measure the 27 V DC that is supplied by the rectifier to power the RC. Therefore, a probe that damped the signal 10:1 was used and the measurements could therefore manage voltages up to 100 V. The trigger was adjusted at 3 V, i.e., started the measuring at 30 V. The probe measured over the RC circuit.

Levels below 30 V are not sufficient to cause the RC to fail, because that is close to the operational voltages for the system. The specification for the capacitor and the resistor is 64 V. The sample rate was set to 10 kHz.

Two DAQ units and computers were used, one in Oxmyran and the other one in Öre Älv. Both the units were tested in labs to verify that they show the same values from the same source and behave in a similar way. It was checked with a Keithley 2400 SourceMeter. The reference measurement on site was done in Öre Älv, which has no problems with the RC circuit, and it was possible to measure the same train on the same passage. The function of the measurement program was periodically verified by a test measure when the computer clock changed over to a new date. There is no train passing at this time.

The two measurements that can be compared are the curves from the measurements that were done in the middle of the night. These clearly show

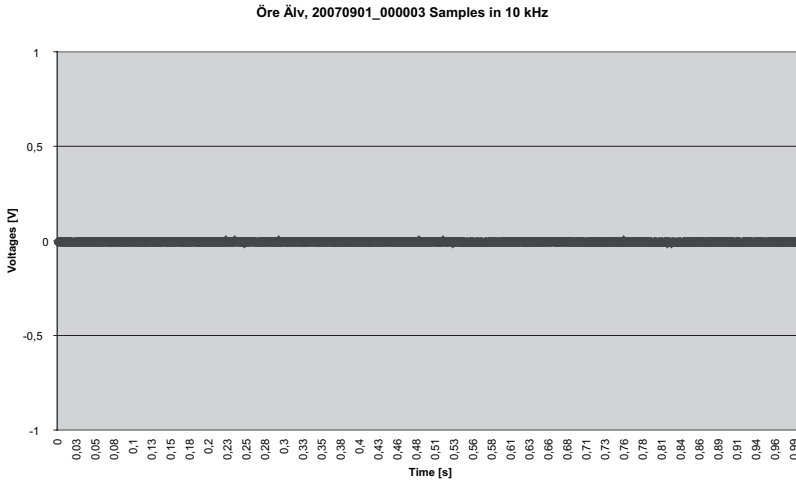


Figure 16: Measurement during the night when there are no trains in the Öre Älv station.

that there is a difference between the two stations, Öre Älv, see Figure 16 and Oxmyran, see Figure 17. The curve in Oxmyran has a ripple and is not as stable as in Öre Älv.

7 Results

As a fault source, “thunder” or lightning is a likely candidate as a cause of EM interference. In this case lightning was a probable cause of failure in only in 2 of 13 cases, since there were only two possible occasions of lightning discharge just before a failure, according to the data from the lightning registering system and Ofelia.

The failures occurs regardless of type of train, the direction of travel, and weights of the trains.

In terms of construction, installation, grounding, surrounding systems, and components, there are no differences between the stations that would explain why failure should occur more often in Oxmyran than in Öre Älv.

The measurements indicate that there are differences between the stations in Öre Älv and Oxmyran. The diagram from a measurement of the RC circuit from Öre Älv, see Figure 16, shows a good level and indicates the zero level. The diagram from measurement at the same point in Oxmyran shows a different voltage level, 80 times higher than in Öre Älv, see Figure

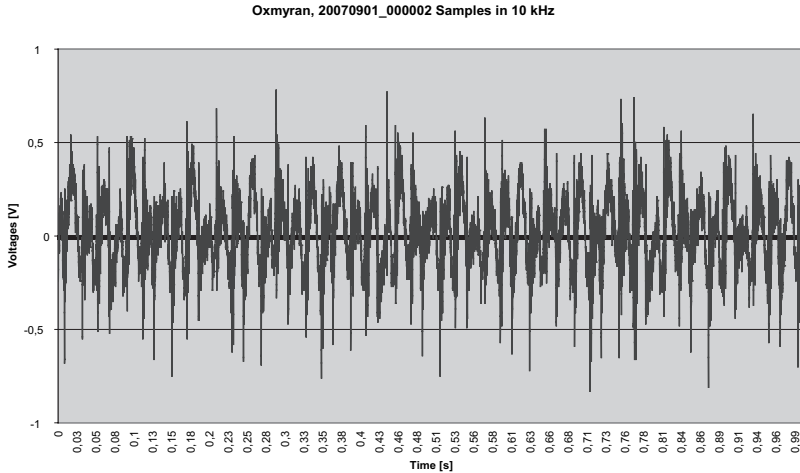


Figure 17: Measurement during the night when there are no trains in the Oxmyran station.

17. The diagrams in both figures are diagrams that were produced during the night and no trains were passing. The voltage level should be zero, but it clearly shows that there are differences in the system that probably accelerate during the time before the RC circuit failure. The difference in this case causes problems in the subsystem that contains the RC component.

Three measurement runs were triggered in Oxmyran and none triggered in Öre Älv. The trigger level is set over the normal level and differs from the specification in the system and the specific RC circuit. The levels of the transients are 50 V or more.

8 Conclusion

The maintenance personnel have found a temporary solution at Oxmyran as they have left several spare components in a plastic bag in the signal box. This case shows the complexity of finding the source of a simple yet frequent RC circuit failure.

The conclusions from this study are summarized as follows:

- The failure reporting system to some extent conceals that the component recurrently fails in the same failure mode. Inaccurate information in the reporting system makes it more difficult to study and under-

stand the EM interference problem and thus to improve EM design and maintenance.

- Regarding the Ofelia failure reporting system, it is clear that “thunder” is often given as a cause of failure, which is probably erroneous. It is suggested that Ofelia records which are categorized as “thunder” should be checked and linked with weather data.
- The high number of alternatives in several fields make it difficult to use and to select appropriate alternatives in the Ofelia reporting system today. Fewer alternatives are recommended to have under every category instead of speculations.
- The variation in the measurements indicates that there are differences in the EM environment in the two signal boxes.

9 Further Work

The issue to the problem in this paper and case was an investigation of the probable source of the fault that breaks the RC circuit in Oxmyran. This task remains to be completed and requires meticulous investigation of the system. There are no measurements during the period immediately before and during a RC circuit failure. Destructive tests may be performed on the RC circuit to measure the possible voltages that could break the RC circuit.

One possibility is to exchange the rectifiers in Öre Älv and Oxmyran and see if the change gives the same measurements.

The ground dampness may affect the grounding system which needs to be investigated and if the stations differ in this respect, this might explain the failure in the RC circuit.

References

- [1] M. Kans. A holistic perspective of computerised maintenance management support. In U. Kumar, A. Parida, R. B. K. N. Rao, editors, *Condition Monitoring and Diagnostic Engineering Management*, pages 231–241, Luleå, Sweden, June 2006. Division of Operation and Maintenance Engineering, Luleå University of Technology.
- [2] P. Söderholm. A system view of the no fault found (NFF) phenomenon. *Reliability Engineering & System Safety*, 92:1–14, Jan 2006.

- [3] S. Mizuno, editor. *Management for Quality Improvement, The Seven New QC Tools*. Productivity Press, Cambridge, 1988.
- [4] L. Pintelon, N. D. Preez, F. Van Puyvelde. Information technology: opportunities for maintenance management. *Journal of Quality in Maintenance Engineering*, 5(1):9–24, 1999.
- [5] T. Williams. *EMC for product designers*. B.H. Newnes, Oxford, second edition, 1996.
- [6] F.M. Tesche, M.V. Ianoz, T. Karlsson. *EMC Analysis Methods and Computational Models*. John Wiley & sons, INC, New York, 1997.
- [7] IEC 50. Glossary of electrotechnical, power, telecommunication, electronics, lightning and colour terms: Electromagnetic compatibility. Standard, IEC, International Electrotechnical Commission, Geneva, Switzerland, 1999.
- [8] European Standard. En 13306, maintenance terminology. Standard, CEN, European Committee for Standardization, Brussels, 2001.
- [9] S. Mizuno. *Company-Wide Total Quality Control*. Asian Productivity Organization, Japan, fourth edition, 1990.
- [10] B. Nyström, U. Kumar. Analysis of train delay information. In *Proceedings of the World Congress on Railway Research*, pages 253–261, Edinburgh, September-October 2003.
- [11] Uppsala University. Lightning Research, Faculty of Science and technology. 2007.
- [12] S. Benda. *Interference-free electronics, Electromagnetic Compatibility*. Studentlitteratur, third edition, 2001.

Paper VIII

Selection of maintenance actions using the Analytic Hierarchy Process (AHP): decision-making in railway infrastructure

B. NYSTRÖM and P. SÖDERHOLM*

Division of Operation and Maintenance Engineering & JVTC Luleå Railway Research Centre, Luleå University of Technology, SE-971 87 Luleå, Sweden

A methodology for prioritising between different maintenance actions in the railway infrastructure is presented. The consistency of the prioritisation and the feasibility of the applied methodology are investigated. Criteria describing the diverse effects of maintenance are developed and presented to track managers, together with a set of maintenance actions, specific for each track manager. Then, the Analytical Hierarchy Process (AHP) is used to obtain preferences for the criteria and for the different actions. The track managers roughly agree on the prioritisation of criteria. However, the discrepancies between the results of the two ways employed to elicit the preferences for the actions are rather large. The track managers consider it easy to understand the rationale of AHP and to enter their preferences. It is proposed that preferences are recorded as they are in this paper, in order to document the rationale of the decisions and to facilitate mutual learning among decision-makers and over time.

Keywords: Railway infrastructure management; Multi-criteria decision-making; Analytic Hierarchy Process (AHP); Maintenance allocation

1. Introduction

Decisions regarding investments can be performed in four steps (Butler *et al.* 1993). The first step is *Identification of opportunities*, the second *Project development*, the third *Selection of alternatives* and the fourth *Control*. The steps involved in maintenance decisions might be considered to be the same. All four steps are necessary in order to give feedback and evaluate whether decisions made are good or not, and hence achieve continuous improvement, see Shewhart (1939) and Deming (1993). However, a decision might be correct even though it turns out to be bad. A correct decision is based on relevant data that is identified, collected and analysed in a rational way (NUREG 1981). On the other hand, a good decision leads to good results (i.e. high utility). Hence, whether a decision was good or bad is only possible to tell in retrospect.

Economic theory prescribes that the expected value (or expected utility) should be maximised. However, expected utility theory runs into problems when describing how

*Corresponding author. Email: Peter.Soderholm@ltu.se

decisions are actually made. The experimental psychologists Kahneman and Tversky (1979) developed the prospect theory to describe how people are affected by risk when they choose between alternatives, e.g. different investments. The choice is also highly affected by whether the question is stated as a loss or a gain (Tversky and Kahneman 1981, Tversky and Thaler 1990). When the decisions are multi-criteria, i.e. comparing alternatives with different kinds of utilities, the difficulties with the expected utility theory become even more pronounced.

There are many different requirements placed on a railway Infrastructure Manager (IM), creating a decision environment characterised by multiple criteria. In Sweden, the government has formulated aims on connectivity, quality of transport, safety, environment, regional considerations and gender equality for Banverket (the Swedish Rail Administration). For instance, measures of quality of transport include the number of train delays (the Swedish Ministry of Enterprise, Energy and Communications 2005). At the same time, the amount of train traffic increases (EU 2004). Hence, as the IM has a limited budget, it is important to carry out the maintenance actions that are most beneficial. The requirements and goals for the IM are related to the whole railway network. However, the decisions are based on local assessments and circumstances. The aforementioned factors highlight the importance of being able to make both correct and good decisions regarding maintenance of the railway infrastructure.

In a study performed by Wiklund (2006), experts in a local group setting were requested to estimate the necessary increase in maintenance efforts to reduce large train disturbances by 20% respectively 50%. It was found that the judgments differed considerably between the experts. The study presented in this paper goes in the opposite direction, i.e. from specific maintenance actions to their effects.

A track manager is responsible for the railway infrastructure in a certain area, containing one or several track sections, typically totalling 100-200 km of railway line. The track manager works in cooperation with adjacent track managers, technical specialists and the relevant Maintenance Contractor (MC). The decisions regarding which railway infrastructure maintenance actions to perform are multifaceted. They include decisions on rules and regulations governing railway infrastructure maintenance, decisions on train traffic and funding. The track managers make explicit decisions by selecting among alternative maintenance actions. As mentioned earlier, there are also many different aims for the track manager to consider. Furthermore, the effects of carrying out a certain maintenance action might be uncertain. The cost might be uncertain due to varying prices of material or labour. Neither are the future changes in traffic known, which makes the benefits (i.e. utility) of some actions even more uncertain.

The research questions considered in this paper are:

- a. How important do decision-makers consider different criteria affected by infrastructure maintenance to be?
- b. How consequent is the selection of maintenance actions?

The questions are answered by comparing the results from different ways of ranking the same alternative maintenance actions; i.e. ranking by criteria, ranking by alternatives and factual ranking. For the first two ways, the rankings are obtained by using the Analytic Hierarchy Process (AHP), which is described by Saaty (1980).

This paper first describes AHP, and then provides a description of the applied methodology. In Sections 4 and 5, the empirical work is presented. Finally, some conclusions and discussion are given.

2. Multi-criteria decision-making and the Analytic Hierarchy Process (AHP)

Different maintenance actions often have different kinds of effects. The problem of selecting the alternatives that best help to reach different objectives (criteria) is studied in the field of Multi-Criteria Decision Making (MCDM). Triantaphyllou (2000) gives an overview of different MCDM methodologies. One such methodology is the Analytic Hierarchy Process (AHP), which was presented by Saaty (1980). When applying the AHP methodology, the decision-maker is asked to state the preferences when comparing every possible pair of alternatives. From these pair-wise comparisons, the AHP allocates priority (weight) to each alternative. As an illustration, say that the decision-maker regards alternative A to be three times better than alternative B, i.e. $A = 3B$. The decision-maker also considers B to be twice as good as C; i.e. $B = 2C$. Hence, if consistent, the decision maker should consider A to be six times better than C; i.e. $A = 6C$. However, human decision-makers are not entirely consistent, as the decision-maker might consider C to be better than A (i.e. $C > A$), or consider A to be just five times better than C (i.e. $A = 5C$). In both cases, there is an inconsistency, which can be quantified by AHP. The inconsistency ratio is an indicator of the reliability of the resulting priorities. An inconsistency ratio of about 0.10 or less is considered acceptable. Inconsistency might be thought of as the required adjustment to improve the consistency of the pair-wise comparisons, according to Saaty (1994). The inconsistency indicates how reliable the resulting priorities are.

In order to sort n alternatives, one needs $n-1$ comparisons, when the alternatives are compared pair-wise both regarding which is preferred and the degree of preference. In contrast, the AHP requires each single alternative to be compared to every other alternative, necessitating $n(n+1)/2$ comparisons. The 'extra' comparisons introduce redundancy, which makes the resulting priorities more trustworthy. The strength of AHP lies in that alternatives are pair-wise compared to each other. Thus, the decision-maker does not have to quantify the alternatives using absolute numbers. AHP is also considered to be easy to use for the layman. A drawback with the AHP methodology is the time required to make the comparisons, which increases rapidly as the number of alternatives, n , increases (Saaty 1980, 1994).

An overview of AHP applications is given by Vaidya and Kumar (2006). Davidson and Labib (2003) combined AHP and FMEA, and investigated possible solutions to a Concorde malfunction. This combination was also employed by Bertolini and Bevilacqua (2006), who investigated maintenance strategies for pumps. When prioritising building renewal projects, Karydas and Gifun (2006) compared the priorities obtained by AHP to the ones obtained by discussion. They concluded that one advantage of using AHP is that the decision-maker sees the risk of not performing a certain action. AHP was used by Hagquist (1994) to establish priorities (weights) for the criteria used for prioritising road improvements and maintenance, which were to be used as decision support. Ramadhan *et al.* (1999) also used AHP to prioritise road maintenance. However, no application of AHP in the prioritisation of railway infrastructure maintenance actions has been found.

3. Methodology

Although the track manager might be a subordinate to the person who signs the final papers, the track manager is most often the real decision-maker. Furthermore, even though Banverket is currently reorganising, the track manager role has remained fairly stable. Therefore, this study investigates maintenance actions that are subject to an explicit selection by the track manager. Most of the considered actions has a budget of € 10 000 – 100 000. The applied methodology consists of three steps as illustrated in Figure 1, where the information in the study flows from left to right.

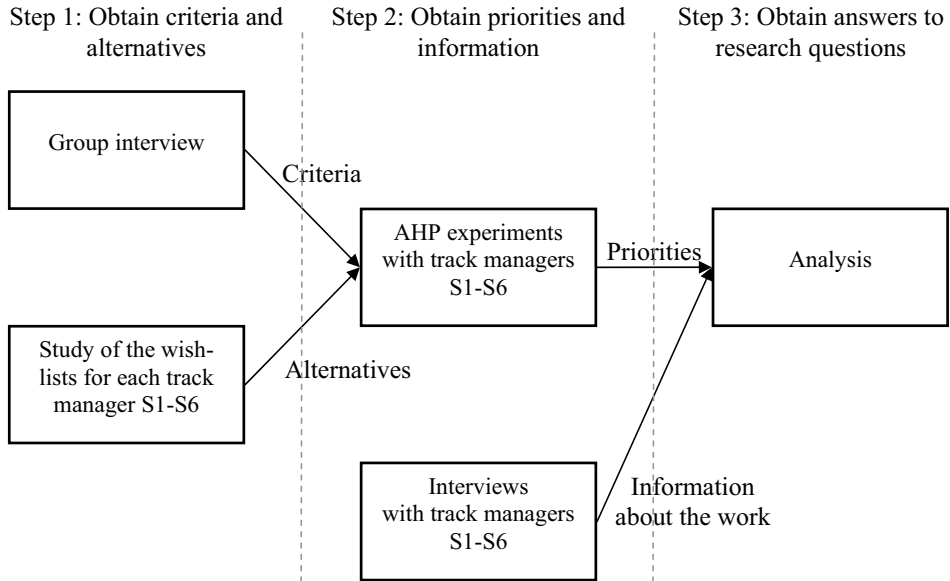


Figure 1. The information of the study follows the arrows from left to right. The applied methodology begins with a group interview and document studies, which provide the basis for the criteria needed for the AHP experiments carried out with the track managers (S1-S6). The resulting priorities are then analysed, taking into consideration the information given by respective track managers in individual interviews conducted after the experiments.

3.1 Steps of the methodology

In the first step of the applied methodology, a group interview and a document study were performed, in order to obtain criteria (effects of maintenance actions) and maintenance actions respectively. The group interview was performed with ten track managers, from one track region. First, they each filled in a questionnaire (Nyström 2007a), in which one of the questions was to name five important maintenance actions. Then, they were asked to write down individually, on sticky notes, the factors needed to describe any maintenance action, without divulging its name. Thereafter, the group was asked to cluster these factors and give names to the obtained clusters, i.e. to make affinity diagrams (Mizuno 1988). After this, the group discussed the maintenance actions most frequently named by them in the questionnaire. The aim was to formulate these maintenance actions in the developed hierarchy of the clusters. The top-level clusters are hereafter termed criteria. After the meeting, the information was compiled in a short report and distributed for clarifications. The rationale of having a fixed set of criteria is that it helps decision-makers to consider every aspect, including both pros and cons. According to Tversky (1972), people tend to focus on features that make the alternatives differ from each other. Hence, different divisions of the features of the alternatives sometimes give different preferences. In the study presented in this paper, the aim is to reduce this effect. This is done by describing the alternatives using several fixed criteria. The performed document study investigated the lists of desired maintenance actions for each track section (typically 50–100 km of railway line). This wish-list is recorded in a spreadsheet, designed differently in different track regions. For the track managers who were responsible for more than one track section, just one track section was selected for the

Selection of railway maintenance actions using AHP

experiment. To help the track managers remember the maintenance actions, the alternatives that had most recently been added to the list were used. In the experiments, five of the track managers compared 12 maintenance alternatives. Track manager S2 was presented with only eight alternatives, due to time constraints.

In the second step of the applied methodology, individual experiments and interviews with six track managers (denoted S1-S6), who did not participate in the group interview, were carried out. Two track managers each from three different track regions of Banverket were interviewed. The track managers were responsible for track sections carrying both freight and passenger traffic, outside large urban areas. Their experience of the railway sector ranged from three years (track manager S3) to more than twenty years (the other five track managers). The informants in this study were all male, as are the vast majority of track managers in Sweden. In the experiment, the experimenter asked the track manager to state his preferences by using the Expert Choice® software. The track manager chose one of the two alternatives based on which he thought was better. The preferences were entered using bars on the graphical user interface. This made input simple, although a drawback might be that it allows the subject to enter extreme preferences, outside the traditional 1,2,...,9 scale of AHP, as described in Saaty (1994). In the experiments, the track managers' preferences were elicited in the following order:

1. Each one of the eight criteria (effects of maintenance) was compared to all the others. This procedure gave the *criteria prioritisation*.
 2. Each maintenance action was compared to every other maintenance action, with respect to criterion 1, 'Cost'. Subsequently, comparisons of the maintenance alternatives regarding each of the remaining criteria, 2-8 were performed. The result of this procedure is called a *ranking by criteria*.
 3. Each maintenance action was compared pair-wise to every other maintenance action, without using any criterion. The result of this procedure is called a *ranking by alternatives*.
- After the experiment, the track manager was interviewed regarding the work in general.

The information from the AHP experiments and the interviews were jointly analysed in the third step of the applied methodology, in order to receive answers to the stated research questions.

3.2 The experiments

The track managers were asked how they would assess the alternatives presented to them. See Figure 2 for a screenshot from the experiment carried out with track manager S6. The track manager enters his preferences, using any of the two bars. Below the bars is the data grid for the comparisons carried out regarding the considered criterion. The coloured bars underneath the names of the alternatives show the resulting priorities, calculated by Expert Choice® using the distributive mode of the AHP.

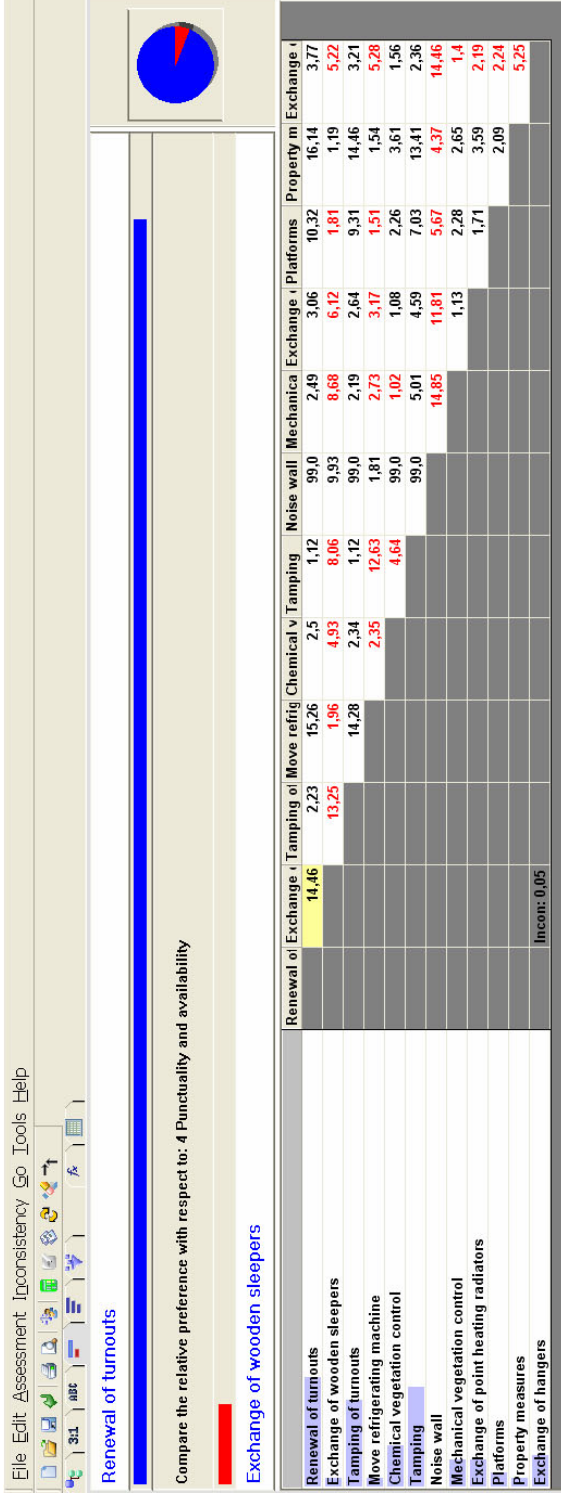


Figure 2. A screenshot of Expert Choice® as seen by the track managers in the experiments. The track manager enters his preference for one alternative to the other by dragging any of the two bars. Here, track manager S6 prefers the upper alternative to the lower alternative, with respect to the criterion '4 Punctuality and availability'. In the comparison matrix, the numerical representation of the preferences is displayed. The coloured bars underneath the names of the alternatives show the resulting priorities. The alternatives are translated from Swedish.

No commonly acknowledged set of criteria exists to describe the effect of maintenance actions. Actually, the group interview showed that track managers think in terms of specific maintenance actions, not the effects of the actions (i.e. the criteria).

Another finding of the group interview was that non-documented actions are a problem. Documentation as to the grounds for a certain decision and comparisons of the results obtained to the ones planned are also lacking. Hence, the scope for evaluating decisions made and identifying whether they were correct or good, and thereby achieving continuous improvement, is undermined.

The developed criteria are shown in Table 2. In the criteria document given to the track managers in the experiments (Nyström 2007b), definitions of each criterion and the factors under it are stated. Each criterion has a maximum of three levels under it.

Table 2. The developed criteria, as they were presented to the track managers in the experiments. Each criterion has a definition and a number of factors, in order to facilitate the track manager's understanding of what is included in the factor. Here, the definition and factors are shown only for criterion 2, Track work time.

1. Cost
2. Track work time
A track work hinders, entirely or partly, train traffic (and other track works). This might happen before, during, as well as after, the 'actual' work in question.
2.1 Impact on traffic
2.2 Difficulty to shift the impact on traffic over time
3. Safety
4. Punctuality and availability
5. Condition
6. Environmental impact
7. Own abilities and development
8. Collaboration with stakeholders

5. Analysis and results from the AHP experiments

Each of the six track managers (S1-S6) compared the eight criteria pair-wise. From these pair-wise comparisons, the priorities and rank orders of Table 3 were calculated. The criteria are listed from high priority to low.

Table 3. The different track managers' rankings of the criteria. The priority column shows the arithmetic mean of the six track managers' priorities for the respective criterion (not shown). This number gives the rank.

Criterion	S1	S2	S3	S4	S5	S6	Priority	Rank
Safety	1	1	2	1	1	1	0.406	1
Punctuality and availability	2	3	1	3	5	2	0.154	2
Track work time	3	2	3	5	4	5	0.093	3
Cost	4	4	4	7	7	3	0.085	4
Condition	5	8	5	4	2	4	0.084	5
Own abilities and development	8	7	7	8	3	7	0.063	6
Collaboration with stakeholders	7	5	6	6	6	8	0.058	7
Environmental impact	6	6	8	2	8	6	0.057	8

Selection of railway maintenance actions using AHP

In Table 3, the rankings of the eight criteria according to each track manager (S1-S6) are shown. In the two right-most columns are the mean of the different track managers' priorities and, from these weights, the resulting overall ranking. As can be seen from Table 3, 'Safety' is the top-ranked criterion, accounting for 40.6% of the total priority. 'Punctuality and availability' is second, while 'Track work time', 'Cost' and 'Condition' constitute the middle section. The lowest priorities, around 6%, were given to 'Own abilities and development', 'Collaboration with stakeholders' and 'Environmental impact', which comes last.

To aggregate the priorities of the eight individual track managers, the arithmetic mean was calculated, according to Forman and Peniwati (1998). Their alternative method, to calculate the geometric mean, would reverse the ranking between the criteria 'Cost' and 'Condition', and also between 'Collaboration with stakeholders' and 'Environmental impact'. This is reasonable, because the mean priorities differ so little within these pairs of criteria.

In Table 4, the inconsistency ratio of the criteria prioritisation for the different track managers is shown. One possible explanation for the high inconsistencies of some track managers (S2 and S3) is that the subjects were not allowed to reconsider the pair-wise comparisons that contributed the most to the inconsistency (but such changes were allowed for the pair-wise comparisons of the maintenance actions later in the experiment).

Table 4. The inconsistency of the criteria prioritisation for each of the track managers (S1-S6).

Track manager	S1	S2	S3	S4	S5	S6
Inconsistency of criteria prioritisation	0.01	0.24	0.28	0.03	0.11	0.11

Each track manager was presented with the criteria in the order of Table 2, i.e. beginning with a comparison of the maintenance actions with respect to 'Cost'. Table 5 shows the inconsistencies of the pair-wise comparisons of the alternative maintenance actions, with respect to each criterion (1-8) and track manager (S1-S6). The two highest inconsistencies of each track manager are starred (if 0.10 or greater). The highest inconsistency is 0.71 of S1. The last row shows the inconsistencies of the pair-wise comparisons with respect to the alternative maintenance actions. When S1 learned about this, he said that the definitions of the criteria 'Punctuality and availability' and 'Track work time' might have blurred together, as availability might be incorrectly construed as the availability of track maintenance activities. The exceptionally low inconsistency of S3 is at least partially explained by the fact that he entered mostly small and moderate preferences.

Table 5. The inconsistencies of the comparisons carried out by the track managers (S1-S6), with respect to each of the eight criteria and with respect to the alternatives. The two highest inconsistencies of each track manager are starred (if 0.10 or greater).

Criterion	S1	S2	S3	S4	S5	S6
1. Cost	0.11	0.34*	0.01	0.08	0.11	0.05
2. Track work time	0.26*	0.02	0.00	0.41*	0.19*	0.13*
3. Safety	0.22	0.16*	0.00	0.05	0.08	0.05
4. Punctuality and availability	0.71*	0.14	0.00	0.01	0.15*	0.05
5. Condition	0.08	0.04	0.00	0.04	0.04	0.01
6. Environmental impact	0.01	0.07	0.01	0.01	0.06	0.03
7. Own abilities and development	0.02	0.15	0.00	0.01	0.01	0.01
8. Collaboration with stakeholders	0.00	0.08	0.00	0.01	0.01	0.02
Alternatives	0.00	0.12	0.00	0.01	0.09	0.01

Table 6 shows the twelve maintenance actions that track manager S6 had to select among, as presented to him. The alternatives ‘Move refrigerating machine’ and ‘Property measures’ during the interview both turned out to be linked to the same power converter station, i.e. they should be decided and carried out as a single action. Such a synergistic characteristic was also described by track manager S1, who put it well by saying ‘A culvert is not just a culvert’. The drainability of a culvert is heavily dependent on other civil engineering activities. Track manager S5 had an alternative that is on the border between maintenance and investment; ‘Refrigerating machine signal box’. This is the installation of a refrigerating machine, motivated by the installation of new, heat-emitting, equipment. As it is a consequence of an investment, it might be considered to be a part of this investment (although not included in its budget). On the other hand, from the point of view of the previously installed items, their environment is maintained at the same temperature by the action, so it might be considered a maintenance action.

Table 6. The list of maintenance actions that the track manager S6 had to select among (also seen in Figure 2). ‘Tamping’ does not include tamping of turnouts.

Renewal of turnouts
Exchange of wooden sleepers
Tamping of turnouts
Move refrigerating machine
Chemical vegetation control
Tamping
Noise wall
Mechanical vegetation control
Exchange of point heating radiators
Platforms
Property measures
Exchange of hangers

The experiment with the track managers resulted in a ranking by criteria and a ranking by alternatives, as exemplified in Table 7 for track manager S6. The three arrows show parts of the relation between the two rankings. The topmost arrow highlights that the top-ranked criterion according to the ranking by criteria is also top-ranked according to the ranking by alternatives; i.e. there is a difference of 0 positions. The next arrow points to the alternative ranked six on the ranking by criteria, ‘Exchange of point heating radiators’, which is found in fifth place in the ranking by alternatives, i.e. there is a difference of -1. The lowermost arrow points to ‘Property measures’, which has a difference of +2 positions.

Selection of railway maintenance actions using AHP

Table 7. The rankings by criteria and by alternatives as given by track manager S6. The arrows exemplify the differences in positions between the two rankings.

Ranking by criteria								Ranking by alternatives	
1	2	3	4	5	6	7	8		
Renewal of turnouts								0	Renewal of turnouts
Tamping of turnouts									Tamping of turnouts
Tamping									Tamping
Exchange of hangers									Exchange of hangers
Move refrigerating machine								-1	Exchange of point heating radiators
Exchange of point heating radiators									Chemical vegetation control
Chemical vegetation control									Mechanical vegetation control
Mechanical vegetation control									Platforms
Property measures								+2	Move refrigerating machine
Exchange of wooden sleepers									Noise wall
Noise wall									Property measures
Platforms									Exchange of wooden sleepers

Table 8 shows the differences in rankings for track managers S1-S6, where the same sign convention as in Table 7 is employed.

Table 8. The differences between the rankings by criteria and by alternatives for track managers S1-S6. Track manager S2 was subjected to eight maintenance actions in his experiment, the others to 12.

	S1	S2	S3	S4	S5	S6
	0	0	3	6	3	0
	3	1	9	10	-1	0
	1	-1	-2	3	4	0
	-1	3	1	1	2	0
	7	-1	1	3	4	4
	-4	2	-4	4	-3	-1
	1	-1	3	2	-2	-1
	-2	-3	0	3	0	-1
	-2		-6	-8	-7	2
	0		-3	-8	1	2
	-2		-2	-8	1	-1
	-1		0	-8	-2	-4

The degree of correlation between two rankings (i.e. only the order is taken into account, not the priorities from which it was calculated) might be calculated using the Spearman rank correlation coefficient, r_s (Sachs 1982). The coefficient r_s attains values between and including -1 and 1, with 1 meaning that the two rankings are identical and -1 meaning that they are reversed. Table 9 shows r_s for the ranking by criteria and ranking by alternatives, for each of the track managers (S1-S6). The rankings of S6 have the highest correlation ($r_s = 0.85$). For the rankings with the lowest absolute correlation, the rankings of S3 ($r_s = 0.41$), the hypothesis that the rankings are independent is rejected at the 20% level. For the other track managers it is rejected at the 10% or lower levels. S4 has a negative correlation coefficient, which means that the rankings are, to some extent, the reverse of each other. The major differences in the rankings of S4 are illustrated in Figure 3.

Table 9. The Spearman rank correlation coefficient, r_s , between the ranking by criteria and the ranking by alternatives, for the track managers (S1-S6).

Track manager	S1	S2	S3	S4	S5	S6
r_s	0.69	0.69	0.41	-0.54	0.60	0.85

One of the track managers (S6) placed the four top-most alternatives in the same positions on both rankings, contributing to an r_s of 0.85. Two more track managers (S1 and S2) had the same alternative on top in both rankings. S4 is the track manager with the greatest difference between his rankings ($r_s = -0.54$). He has the alternative with the greatest difference between the two rankings, as his ranking by criteria places the alternative ‘Make an inventory of cables’ in second place, while in the ranking by alternatives, it ends up in twelfth place, i.e. last. For S4, the last four alternatives according to the ranking by criteria end up as the first four in the ranking by alternatives. This is schematically illustrated by Figure 3. For clarity, the names of the maintenance actions are omitted. Performing a sensitivity analysis, by changing the weights of the criteria contributing to the four lowermost criteria being at the bottom, does not bring the ranking by criteria order much closer to that of the ranking by alternatives. Track manager S4 did not object to any of the two rankings of the maintenance alternatives.

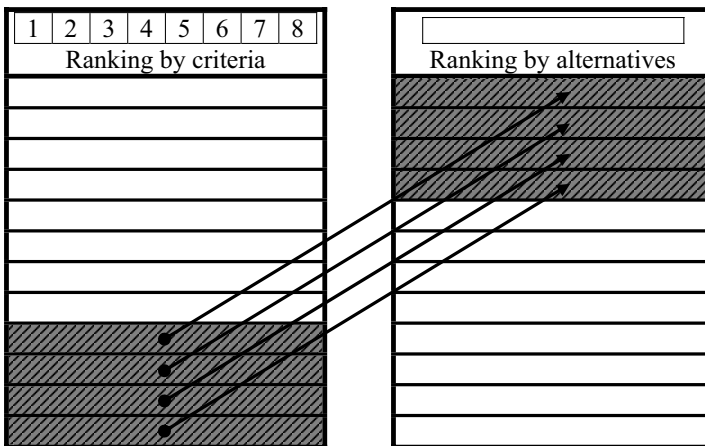


Figure 3. The track manager S4 had the greatest difference between the ranking by criteria and ranking by alternatives. The four lowest ranked alternatives according to the ranking by criteria have each moved up eight positions and constitute the top-four of the ranking by alternatives.

The track managers S3 and S6 placed alternatives involving maintenance of platforms higher when ranking by alternatives than when ranking by criteria. The high ranking by alternatives might have been caused by the phenomenon that, when it comes to such ‘popular’ items, the track manager may well be affected by many other people’s opinions, according to one track manager.

Much of a track manager’s job is about deciding which maintenance actions can be postponed, e.g. to the following year. It is also about how one should prepare, e.g. if and when to make an audit of a certain type of item.

6. Discussion

All track managers, except one, ranked the 'Safety' criterion first of the eight criteria. The exception (S3) regarded 'Safety' as a hygiene factor, assuming that it is already in place, and ranked it as the second criterion. It is also noteworthy that the pair-wise comparisons of S4 put 'Environmental impact' in second place, while the other five track managers ranked that criterion sixth or eight.

Three out of the six track managers had 'Track work time' as the criterion with the highest inconsistency. This indicates that the definition of the criterion (see Table 2) is unclear or that the track work time is unknown or hard to assess. Assessment might also be made difficult by the fact that the criterion is 'two-dimensional'. As some track managers said that factor 2.2 ('Difficulty to shift the impact on traffic over time') did not pose any problem at all, it is more likely that the factor 2.1 ('Impact on traffic') is partly unknown or hard to assess. The large variation between track managers regarding the inconsistencies of their preferences of the criteria indicates that the degree of understanding varies between the track managers in the experiments.

When one of the alternatives had a 'zero' value, some of the track managers entered extremely strong preferences for one of the alternatives. An example was when one of the alternatives did not need any (or a very brief) track work time (criterion 2). Entering such preferences increases inconsistency. The most extreme example is track manager S4, where the existence of alternatives that included only projecting, thereby not requiring any track work time, strongly contributed to increasing the inconsistency for the criterion 'Track work time' to 0.41.

It is desirable to obtain quite large differences between the priorities of the maintenance actions, resulting in a reliable ranking. The clarity of the priorities is about as good for the priorities according to criteria, as for the priorities according to alternatives. Track manager S1 might be considered an exception, as the priorities according to criteria show clearer steps between them than do the priorities according to alternatives. The rankings of track manager S4 are strongly unrelated to each other, and his rankings are equally unclear.

There are dissimilarities between the ranking by criteria and ranking by alternatives, although these are not very large. However, track manager S4 is a clear exception. No explanation for the large disparities between his rankings has been found. It might also be noteworthy that the track managers with the lowest rank correlations, S3 and S4, are from the same track region.

Track manager S3, with the least experience (three years), shows high inconsistency as regards criteria prioritisation (0.28 as of Table 4), a lack of strong preferences (as stated in connection with Table 5), as well as a fairly low correlation of his different rankings (Table 9). This might be caused by his relative lack of experience. The more experienced track managers might have been better able to internalise the set of criteria presented to them in the experiments, due to their longer exposure to discussions about the pros and cons of various maintenance actions. However, the fact that the more experienced track manager S2 shows an inconsistency (0.24 as of Table 4) almost as high as S3 does not support this hypothesis. Another possible explanation for the fact that the least experienced track manager stands out in Table 5 and Table 9 might be that he did not know as much about the maintenance actions as the other track managers did. To follow maintenance actions from their infancy in loose discussions, to maintenance execution and to experience the results requires several years of practice. This is especially true for areas like railway infrastructure, where the useful life of some items is as long as if not longer than that of an individual's normal working life. However, several of the more experienced track managers also said that they lacked first-hand knowledge of some of the maintenance action alternatives. But, in order to obtain a familiarity

with maintenance actions that is appropriate for decision-making, a good dialogue with relevant specialists might be more important than detailed technical knowledge.

The Spearman rank correlation coefficient might also be used to compare the correlation between the two rankings obtained in the experiments to another ranking of the maintenance actions e.g. derived from a factual rating. The correlation is possible to estimate for S5 and S6, although not every alternative has a factual ranking. It is clear that the correlation between each of the two rankings obtained in the experiments and the factual ranking is considerably lower (at most 0.05 for S5, at most 0.55 for S6) than the correlation between the two rankings obtained in the experiment (0.60 and 0.85, as shown in Table 9). Possible explanations include the fact that it is not the track manager who makes the factual ranking and that the difference in ranking procedures causes the dissimilarities. The existence of the latter phenomenon was illustrated in experiments by Lichtenstein and Slovic (1971), where subjects chose between alternatives and also gave bids for the same alternatives. Often, these two ways of eliciting preferences gave reverse preferences. In maintenance, one might see an alternative as a loss (maintenance is a necessary evil) or a gain (maintenance assures performance). Therefore, the preferences of maintenance actions might be affected by how they are described and perceived, as well as by the employed procedure for elicitation.

The ranking by one of the criteria might also be investigated in terms of correlation to a factual ranking. However, the only obtainable factual rankings stem from the budgets (of S2 and S5), which would make it possible to compare the ranking by the criterion 'Cost' to a ranking given by the cost figures of the budget. These do not correlate well ($r_s=0.40$ for S2). However, as the budget only gives a point estimate, with the risks untold, one cannot from this tell whether the track manager has a good grasp of the costs or not.

As each track manager has a unique set of alternative maintenance actions, it is generally not possible to compare the rankings of maintenance actions between different track managers. The only maintenance actions that are likely to mean the same thing to different track managers, thus being comparable, are 'Tamping' (excluding the turnouts), 'Tamping of turnouts' and maybe 'Renewal of turnout', all on the wish-lists of both S5 and S6. In Figure 4, the priorities according to the criteria (1-8, as listed in Table 2) are shown for track manager S5. It can be seen that the profile of criteria, according to S5, is fairly similar for 'Tamping' and 'Tamping of turnouts'.

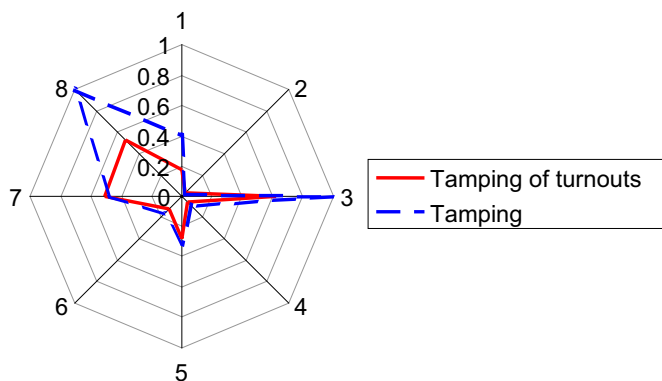


Figure 4. A spider diagram describing the maintenance actions 'Tamping of turnouts' and 'Tamping' (excluding the turnouts), according to S5. Each priority axis ranges from 0 to 1. The criteria 1-8 are listed in Table 2.

Selection of railway maintenance actions using AHP

Regarding S5 and S6, considering the three aforementioned maintenance actions, the similarities of the criteria profiles are greater between different actions for the same track manager than between the same action for different track managers. This might indicate that different track managers do not have the same preferences for identical maintenance actions. This may be due to the fact that the effects of identical maintenance actions vary between different parts of the railway network. Furthermore, the choice of other actions on the track manager's wish-list will affect the relative priority of a certain maintenance action.

If the priorities of the criteria for the alternative maintenance actions are correlated, it might imply different things. The criteria '7 Own abilities and development' and '8 Collaboration with stakeholders', are strongly correlated by four out of the six track managers. This is exemplified, using data from track manager S3, by Figure 5. In Figure 5, the priorities for criteria 7 and 8 are plotted for each of the twelve alternatives. The correlation might imply that these criteria do co-vary with the maintenance actions. It might also mean that these criteria are not formulated in a mutually exclusive way or reflect a transfer from criterion 7 to criterion 8. This is motivated by the fact that the track managers participating in the experiment might have got tired towards the end of the experiment, and this might have caused them not to relearn completely, i.e. to consider criterion 8 instead of 7.

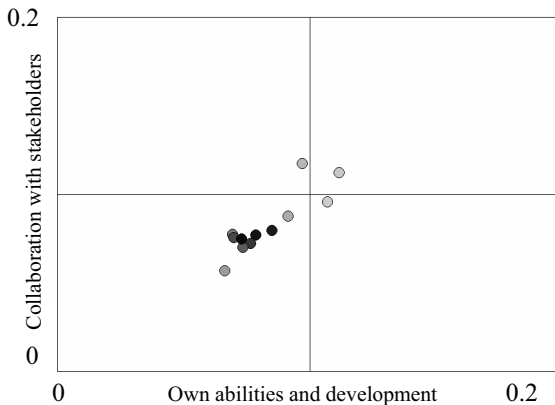


Figure 5. The priorities of criterion '7 Own abilities and development' are on the x-axis, the priorities of criterion '8 Collaboration with stakeholders' are on the y-axis. The twelve maintenance alternatives given to track manager S3 are plotted.

There are also a strong correlation between criteria '3 Safety' and '4 Punctuality and availability' and between '3 Safety' and '5 Condition', according to three of the track managers. This might be explained by the fact that improved condition leads to higher reliability, which results in better safety as well as increased availability and thus punctuality, to different extents. An example, representative of these correlations, is given in Figure 6, where the correlation between '3 Safety' and '4 Punctuality and availability' for track manager S1 is illustrated. For example, the maintenance action 'to exchange rail that has been classified as unsafe', has the highest priority regarding both '3 Safety' and '4 Punctuality and availability'. Figure 6 also contains an outlier, 'Installation of protection wall', for which the priority with respect to 'Safety' is about 0.27, but the priority with respect to 'Punctuality and availability' is close to 0.00.

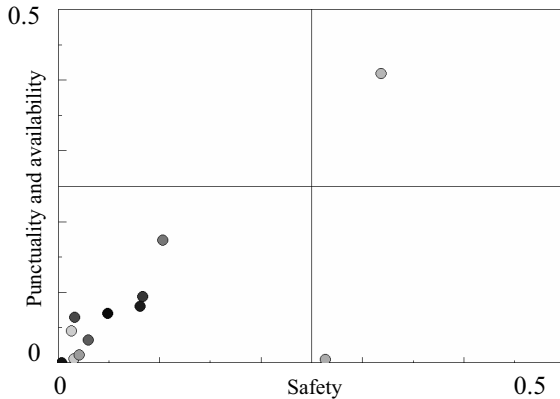


Figure 6. The priorities of criterion '3 Safety' are on the x-axis, the priorities of criterion '4 Punctuality and availability' are on the y-axis. The twelve maintenance alternatives given to track manager S1 are plotted.

The fact that the highest and second highest inconsistency both belong to the criteria 1-4 for each track manager (see Table 5) implies that the criteria definitions or the order in which they are presented affect the consistency.

The kinds of decisions made by the track managers vary among the track managers and over time. Significant characteristics include:

- Urgency: contracts include remediation of inspection remarks to varying extents
- Trade: e.g. electric power might be decided by others and not by the track manager
- Effect: e.g. actions to improve punctuality have been separately financed.

Access to experts in certain specialities, e.g. sleepers, will also impact decisions.

In this paper, the decisions of track managers have been investigated. By comparing different rankings; by criteria, by alternatives and factual, possible variations in the decisions are indicated. However, a low variation indicates only that correct decisions are made, not whether they are good or bad. None of the informants, on the other hand, had heard about the effects of the decisions being evaluated, i.e. planned effects being compared to attained effects. As the motivations for prioritising are not documented, one might, for example, know that the decision was bad, but not whether it was correct or not. The need for documentation is stressed by the long life of some railway infrastructure items; longer than a human's working life. Documentation would also enable continuous improvement, based on an evaluation of decisions made.

The different ways of describing the alternatives on the wish-lists might impact the decisions taken, although the possible impact is unknown. The wish-lists are, in most cases, not entirely appropriate to be used as support for decisions and follow-up. The track manager, for most of the alternative maintenance actions knows a lot more than is written on the list. Therefore, one cannot from this conclude that the correct decisions are not taken. However, as it is impossible to consider all this information simultaneously, a more systematic description of the maintenance actions might facilitate the decisions taken by the track manager. It will facilitate the communication between different specialists, as the focus will be on the effects (criteria), not the action itself. The authors advocate a list including at least; unique identification of the maintenance action, motivation, description, anticipated effects and possible consequences if the action is not taken.

7. Conclusion

The first research question was concerned with how important decision-makers consider different criteria affected by infrastructure maintenance to be. The answer to this question is summarised in Tables 3 and 4, showing the ranking of the criteria, and is also partly depicted in Figures 5 and 6.

The second research question dealt with how consequent the selection of maintenance alternatives is. Aspects of this question are summarised in Table 5, showing the inconsistencies in criteria assessment, as well as in Tables 7-9 and Figure 3, showing differences in the results from the different procedures employed to rank the maintenance actions.

The AHP methodology is applicable to selection among different maintenance actions related to railway infrastructure. The track managers easily accepted the rationale of the methodology and found it on the whole easy to work with the software. The major drawback was the long experiment time, 4-5 hours, needed to do all pair-wise comparisons. Regarding the maintenance alternatives, one track manager, who had several alternatives projecting maintenance on his list, found it difficult to compare the projecting alternatives to the other, 'ordinary' action alternatives. Hence, the constructed list of criteria (effects of maintenance) worked well, but the alternatives that only involve projecting might need to be treated separately.

The experiments posed questions about the alternatives, seen as isolated maintenance actions. However, some maintenance actions need to be considered together, due to synergy effects. Furthermore, in real life, the track manager has budget limitations, i.e. has to keep total expenditure within a certain limit. The latter circumstance highlights the prospective action that receives the lowest priority, i.e. which is on the border of being funded. The differences between the ranking by criteria, the ranking by alternatives and the factual rankings, indicate that prioritisation should be considered from a wider perspective.

This paper has investigated the track manager as the sole decision-maker. However, the information and knowledge that a track manager gets from other professionals is embedded in his/her selections. Two of the track managers who participated in the experiments, have, to some extent, a formal way of doing this for maintenance works of a more acute character than those investigated in this paper. They put frequently failing railway items on a list, where the IM and the relevant Centralised Train Traffic Control Office (CTTC) each give rankings regarding the importance they attach to remedying different items. The AHP might also be used for simultaneously considering the preferences of several decision-makers, e.g. IM and CTTC. It was not possible to determine whether the experience of track managers affects their consistency in decision-making. However, there were some indications that the decisions made by the track manager with the least experience were less correct than those made by the managers with more extensive experience.

This paper has considered the selection among maintenance alternatives that lead to different types of effects. Another application of the AHP is to choose among different alternatives that strive for the same aim, e.g. noise reduction.

Most of the track managers use a list of desired actions that provides too little description to make informed selections, i.e. correct decisions. Hence, a track manager can learn from other track regions' ways of working with maintenance prioritisation, i.e. by benchmarking.

The AHP is a methodology with high reliability in itself. Several ways of estimating reliability exists, although they were not applied in this study. These include using several researchers to perform the experiments with some of the track managers (with the researchers in alternating order). Also, the order in which the criteria are presented might be randomised.

The impact on the selections made by the presentation of the alternatives, e.g. when alternatives are split into several smaller maintenance actions, also remains to be investigated.

Acknowledgements

The authors would like to thank Banverket and the EU regional fund for financial support. The authors are grateful to the staff at Banverket, especially the track managers who were subjects in the experiments: Christer Ahlin, Stefan Blomgren, Jakob Fors, Benny Harryson, Sune Johansson and Per-Erik Kenttä. Thanks also to Professor Uday Kumar and the colleagues at the Division of Operation and Maintenance Engineering for useful suggestions on the paper.

References

- Banverket, Måttbilaga för strategiska mål i SSP 2007-2009 (Appendix of measures for strategic goals in the general strategic plan 2007-2009), 2006 (Banverket: Borlänge). (In Swedish)
- Bertolini, M. and Bevilacqua, M., A combined goal programming – AHP approach to maintenance selection problem, *Reliability Engineering and System Safety*, 2006, **91**(7), 839–848.
- Butler, R., Davies, L., Pike, R. and Sharp, J., *Strategic investment decisions*, 1993 (Routledge: London and New York).
- Davidson, G.G. and Labib, A.W., Learning from failures: design improvements using a multiple criteria decision-making process. *Journal of Aerospace Engineering*, 2003, **217**(4), 207–216.
- Deming, W.E., *The new economics for industry, government, education*, 1993 (Massachusetts Institute of Technology, Center for Advanced Engineering Study: Cambridge, Massachusetts).
- EU, *Energy, transport and environment indicators. Data 1991-2001*, 2004 (Office for Official Publications of the European Communities: Luxembourg).
- Forman, E. and Peniwati, K., Aggregating individual judgments and priorities with the Analytic Hierarchy Process, *European Journal of Operational Research*, 1998, **108**(1), 165–169.
- Hagquist, R.F., High-precision prioritization using Analytic Hierarchy Process: determining state HPMS component weighting factors, *Transportation research record*, 1994, 1429, 7–14.
- Kahneman, D. and Tversky, A., Prospect theory: an analysis of decisions under risk, *Econometrica*, 1979, XLVII, 263–291.
- Kaplan, R.S. and Norton, D.P., *The balanced scorecard: translating strategy into action*, 1996 (Harvard Business School Press: Boston, Massachusetts).

- Karydas, D.M. and Gifun, J.F., A method for the efficient prioritization of infrastructure, *Reliability Engineering and System Safety*, 2006, **91**(1), 84–99.
- Lichtenstein, S. and Slovic, P., Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 1971, **89**(1), 46–55.
- Mizuno, S., *Management for Quality Improvement: The Seven New QC Tools*, 1988 (Productivity Press: Cambridge, Massachusetts).
- NUREG, *NUREG-0492: Fault Tree Handbook*, 1981 (U.S. Nuclear Regulatory Commission: Washington D.C.).
- Nyström, B., *Enkät om underhåll (Questionnaire about maintenance)*. Diariernr 01916-2007, Luleå University of Technology, 2007a. (In Swedish)
- Nyström, B., *Kriterier på olika sätt (Criteria in different ways)*. Diariernr 01973-2007, Luleå University of Technology, 2007b. (In Swedish)
- Ramadhan, R.H., Wahhab, H.I.A. and Duffuaa, S.O., The use of an analytical hierarchy process in pavement maintenance priority ranking, *Journal of Quality in Maintenance Engineering*, 1999, **5**(1), 25–39.
- Saaty, T.L., *The analytic hierarchy process: planning, priority setting, resource allocation*, 1980 (McGraw-Hill: New York).
- Saaty, T.L., *Fundamentals of decision making and priority theory with the analytic hierarchy process*, 1994 (RWS Publications: Pittsburgh, PA).
- Sachs, L., *Applied statistics: a handbook of techniques*, 1982 (Springer: New York and Berlin).
- Shewhart, W.A., *Statistical method from the viewpoint of quality control*, 1939. Reprint 1986. (Dover: New York).
- The Swedish Ministry of Enterprise, Energy and Communications, *Regleringsbrev för budgetåret 2006 avseende Banverket m.m. inom utgiftsområde 22 Kommunikationer (Appropriation directions for the fiscal year 2006 to the Swedish Rail Administration, etc., within expenditure area 22 Communication)*, 2005. Available online at: <http://webapp.esv.se/statsliggaren> (accessed 28 May 2007). (In Swedish)
- Triantaphyllou, E., *Multi-criteria decision making methods: a comparative study*, 2000 (Kluwer Academic: Netherlands).
- Tversky, A., Elimination by aspects: a theory of choice, *Psychological review*, 1972, **79**(4), 281–299.
- Tversky, A. and Kahneman, D., The framing of decisions and the psychology of choice, *Science*, 1981, **211**(4481), 453–458.

Tversky, A. and Thaler, R.H., Anomalies: Preference reversals, *Journal of Economic Perspectives*, 1990, **4**(2), 201–211.

Vaidya, O.S. and Kumar, S., Analytic hierarchy process: An overview of applications, *European Journal of Operational Research*, 2006, **169**(1), 1–29.

Wiklund, M., *Förebyggande underhållsåtgärders effekt på järnvägstransportsystemets sårbarhet – Försök med delfimetoden (The Effect of Preventive Maintenance Actions on the Vulnerability of the Railway System – Trials Using the Delphi method)*, 2006 (VTI: Sweden). Available online at: <http://www.vti.se> (accessed 2 March 2007). (In Swedish)

