DOCTORAL THESIS

Maintenance Decision Support Models for Railway Infrastructure using RAMS & LCC Analyses

Ambika Prasad Patra

LULEÅ
UNIVERSITY
OF TECHNOLOGY

DOCTORAL THESIS

Maintenance Decision Support Models for Railway Infrastructure using RAMS & LCC Analyses

Ambika Prasad Patra

Luleå University of Technology
Division of Operation and Maintenance Engineering

Printed by Universitetstryckeriet, Luleå 2009

ISSN: 1402-1544
ISBN 978-91-7439-023-0

Luleå 2009

www.ltu.se

Preface

The research work presented in this thesis has been carried out at Luleå Railway Research Center (Järnvägstekniskt Centrum, JVTC) and has been sponsored by Swedish National Rail Administration (Banverket) and ALSTOM Transport, France. I would like to thank Banverket and ALSTOM Transport for providing the financial support during my research.

First of all, I would like to express my sincere gratitude to my supervisor, Professor Uday Kumar for his invaluable guidance, suggestions, encouragement and support during the tenure of my research.

I am particularly grateful to Dr. Pierre Dersin (ALSTOM Transport, France) for providing me guidance, resources and support during my stay in Paris. I would like to thank Dr. Per-Olof Larsson-Kråik (Banverket) for his valuable suggestions and fruitful discussions; Dr. Ulla Juntti and Mr. Arne Nissen (Banverket) for sharing their knowledge in railway and Dr. Peter Söderholm (Banverket) for his contribution in my research work. I would also like to thank Professor Per Anders Akersten for his valuable comments and discussion.

I am also grateful to all my colleagues at the Division of Operation and Maintenance Engineering for their continuous help during the research. Special thanks are acknowledged to Prof. Jan Lundberg, Dr. Aditya Parida, Dr. Rupesh Kumar, Mr. Rajiv Dandotiya and Mr. Yuan Fuqing for their discussions.

My special thanks to all the Indian friends and families for their help throughout my stay in Luleå.

Finally, I would like to express my gratitude to my parents for their support and belief in me.

Ambika Prasad Patra
October, 2009
Luleå, Sweden

Abstract

Today's railway sector is imposing high demands for service quality on railway infrastructure managers. Since railway infrastructure has a long asset life, it requires efficient maintenance planning to perform effectively throughout its life cycle to meet these high demands. Traditionally maintenance decisions for the railway infrastructure have been based on past experience and expert estimations. The application of RAMS (Reliability, Availability, Maintainability and Safety) analysis for railway infrastructure is limited. The focus of this thesis is to demonstrate the applicability of RAMS analysis in effective maintenance planning. Within the scope of this research, various case studies associated with Banverket (the Swedish National Rail Administration and ALSTOM Transport have been carried out. The research presents approaches and models for estimating RAMS targets based on the service quality requirements of the railway infrastructure. The availability target of the infrastructure has been estimated by considering the capacity and punctuality requirements of the infrastructure, whereas the safety goal of the track has been estimated by calculating the probability of derailment by means of undetected rail breaks and poor track quality. Effective estimation of the RAMS targets will help infrastructure managers to predict the maintenance investment in the railway infrastructure needed over a period of time in order to achieve the targets. Nevertheless, the availability target of the infrastructure can lead to train delay. A model has been developed to achieve the availability target in both the scheduled and the condition based maintenance regimes by choosing an effective maintenance interval and detection probability respectively. This has been illustrated by a case study on track circuits. Different maintenance strategies can help in achieving the RAMS targets. In order to determine the cost-effective solution, LCC (life cycle cost) should be used. The maintenance strategy with lowest LCC will be the cost effective maintenance strategy. This has been demonstrated by a case study on a signalling system. Sensitivity analyses have been performed to calculate the maximum cost effectiveness of the system for different maintenance parameters. LCC estimation for a maintenance strategy should always consider the risks associated with the strategy. A fair degree of uncertainty is also associated with LCC estimation due to the statistical characteristics of RAMS parameters. An approach has been developed in this thesis to calculate the uncertainties associated with LCC estimation. Petri-Net analyses, Monte Carlo simulations, Design of Experiment have been used to develop models to achieve the objectives of this thesis. This thesis discusses the applicability of RAMS and LCC analyses for railway infrastructure and demonstrates models for effective infrastructure maintenance planning.

Keywords: Railway infrastructure, Maintenance planning, Reliability, Availability, Maintainability, Safety, Life cycle cost

List of Appended Papers

Paper I: **Patra, A. P., Kumar, U. and Larsson-Kråik, P-O.** (2010). Availability target of the railway infrastructure: an analysis. *Accepted for publication in Proceedings of Reliability and Maintainability Symposium (RAMS)*, San Jose, USA, 25-28 January.

Paper II: **Patra, A. P., Kumar, U. and Larsson-Kråik, P-O.** (2009). Assessment and improvement of railway track safety. *Proceedings of 9th International Heavy Haul Conference (IHHA)*, Shanghai, China, 22-24 June.

Paper III: **Patra. A. P. and Kumar, U**. (2009). Availability analysis of railway track circuit. *Accepted for publication in Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*.

Paper IV: **Patra, A. P., Dersin, P. and Kumar, U.** (2009). Cost effective maintenance policy: a case study. *Accepted for publication in the International Journal of Performability Engineering*.

Paper V: **Patra, A. P., Söderholm, P. and Kumar, U.** (2009). Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 2009, 223 (F3), 285-293.

Contents

1 Introduction

The railway transportation system is one of the most commonly used modes of transport and its importance and utility are very substantial for the society. With the advancement of technology, the changing environment and increasing customer demands, railways are having to upgrade their various operational activities constantly. A safe and reliable network with sufficient capacity and availability is a prime requirement. Railway infrastructure plays its part in achieving this requirement in its system life cycle. To fulfil this requirement in an effective manner, one needs to examine the various phases of the life cycle, such as inception, design, manufacturing, installation, operation and maintenance, and disposal. Once the infrastructure is installed, it is very difficult to modify the initial design. Therefore, the performance of the infrastructure depends largely on the maintenance and renewal decisions taken during its life cycle. The design phase of the track needs to consider not only the cost, but also aspects like Reliability, Availability, Maintainability, and Safety (RAMS) and Life Cycle Cost (LCC), with respect to technological advancements and changes. After the installation, during the operation and maintenance phase, LCC and RAMS are considered when making effective maintenance decisions.

Each of the infrastructure components, with its varying life and degrading conditions, will influence the quality and operability of the infrastructure. In order to maintain the quality of the infrastructure at an accepted level, two aspects of the infrastructure quality need to be considered, i.e. measurement of the infrastructure quality on a continuous basis and means to achieve the required infrastructure quality when the quality falls below the accepted level. The infrastructure quality is measured using various parameters, e.g. the service reliability, the track utilisation and accessibility, the infrastructure safety and the infrastructure system and cost effectiveness. High operation and maintenance costs act as a barrier to achieving a favourable financial performance of railway operations. The infrastructure quality is vulnerable to infrastructure system failures. With an increase in infrastructure requirements in terms of axle load, gross tonnage, speed, etc., the infrastructure experiences more failures, which require more maintenance. At the same time, the availability of the infrastructure to perform the necessary maintenance decreases, due to the increased traffic. This requires more budgetary and other resources. To optimise the maintenance activities in terms of cost-effectiveness and RAMS, a systematic analysis approach is required.

In order to minimise failures for the railway systems, the effects of decisions should be systematically evaluated. The infrastructure manager, which is responsible for the design, construction, maintenance, renewal and upgrading of the infrastructure, has a clearly defined role and is confronted with the increasing performance of its collaborative partners. Due to increases in operation and maintenance costs, infrastructure managers are compelled to optimise their budget, while reliability and availability have to be increased without endangering the traffic safety. A systematic approach is needed for communication with the infrastructure manager in order to guarantee the defined levels of performance. Since, in the current scenario, most of the maintenance and renewal decisions are based on past experience and expert estimations, a need for a systematic LCC approach arises. A life cycle costing approach in combination with RAMS analysis will provide a way to optimise the maintenance strategy, considering the short term budget requirements as well as the long term costs of ownership. Cost-effective decision making based on LCC usually does not consider the risk aspects. Therefore, when performing cost effective decision making based on LCC for the track system, one needs to consider the uncertainties associated with LCC. The associated uncertainties are the risk factors related to costs of the traffic disruptions/derailment and the variable costs due to RAMS parameters. Although some studies have been undertaken in the areas of RAMS and LCC separately (see e.g. Vatn, 2002; Swier, 2004; Zoeteman, 2006), there is a need for an integrated study of RAMS and LCC for the railway sector for enhancing the cost effectiveness of the railway system. When taking effective maintenance decisions based on LCC analysis, it is important to identify the uncertainties associated with LCC in order to support the decision taking process. The uncertainties associated with LCC can broadly be attributed to uncertainties in estimating RAMS parameters and uncertainties concerning the economic conditions of cost parameters over a long time horizon. In the railway sector, most of the efforts to implement RAMS and LCC approaches have been stand alone projects, and have not been integrated with the decision making process.

Under the increasing pressures to improve performance quickly, infrastructure managers are being forced to focus on supplying short-term cost and/or performance improvements only. Despite a substantial amount of research in recent years, many rail deterioration processes are not well enough understood for infrastructure managers to be able to translate them into unambiguous quantitative relationships between investment and maintenance decisions and long term quality effects (Ferreira, 1997; Veit, 2003); and uncertainty in these relationships might result in these effects not being sufficiently appreciated. Governments and shareholders have a preference for short payback periods for investments and quick performance improvements, which can seriously conflict with the nature of railways and optimal spending patterns. The long life spans of components and their high installation costs mean that decisions have a high degree of irreversibility. In addition, the consequences of low initial quality and insufficient preventive maintenance, i.e. high cost levels and low system reliability, often only come to light several years later. After reaching certain degradation levels, backlogs in maintenance lead to progressive degradation and, hence, capital destruction. Although the infrastructure manager should be the party capable of incorporating such effects into the decision making, either implicitly or explicitly, there are many impeding factors. The long term view of designing and maintaining usually conflicts with organisational and institutional boundaries, such as allocated budgets, standard operating procedures, established relations with other actors, and external regulations. Most of these boundaries have a long history, and decision makers usually consider only incremental changes. The present research focuses on the application of RAMS and LCC methodologies to develop a decision support system for a cost effective maintenance policy.

1.1 Research Problem

Infrastructure managers are facing increasing demands from traffic operators as well as passengers to ensure a safe, reliable and comfortable railway service. To achieve these objectives the quality of the infrastructure needs to be improved and maintained. The maintenance activities of the railway infrastructure have certain maintenance goals which are linked to the organisational goals and objectives and which help in achieving the overall objectives of the infrastructure. Karlsson (2005) presented Banverket's (the Swedish National Rail Administration) vision for maintenance activities based on overall goals for securing safety, reliability, comfort and cost-effectiveness. Usually, the overall maintenance strategy consists of various critical success factors that are necessary to achieve the overall goals for maintenance. The critical success factors include the guidelines for the functions (reliability, safety, and comfort) to be achieved; methods for establishing and measuring the relationship between the operational reliability, the condition of the infrastructure and the maintenance work carried out, and methods for measuring the cost effectiveness of maintenance operations, etc. Therefore, there is a need to develop effective maintenance decision support models which can achieve the overall goals of the infrastructure in a cost effective way.

The objectives of maintenance decision support models include determining when the infrastructure needs to be maintained, what maintenance actions need to be carried out, how the maintenance actions will be performed, which maintenance action will meet the infrastructure objective, what maintenance action will secure the safety of the system, etc. RAMS analysis will help in identifying different maintenance alternatives to be carried out on the infrastructure. RAMS provides a performance assurance with which the system can guarantee the achievement of the goals of the infrastructure. LCC analysis will help in optimising the cost effectiveness of the maintenance actions derived from RAMS analysis. Cost estimations through LCC help in foreseeing the cost implications of maintenance actions over the whole service life of the infrastructure, not just in the short term. Literature studies show that the maintenance decision support models in use for railway infrastructure do not consider RAMS and LCC methodologies explicitly. The use of RAMS and LCC methodologies in the maintenance decision models not only helps in enhancing the system effectiveness of the infrastructure, but also makes the maintenance strategies cost effective.

1.2 Research Objectives

The purpose of the present research is to illustrate and demonstrate the applicability of RAMS and LCC analysis in the decision making process governing the cost effective maintenance of the railway infrastructure, taking the associated risks and uncertainties into consideration.

The research has the following objectives:

- To study RAMS methodologies for the railway infrastructure and develop a framework for estimating the RAMS targets.
- To study the applicability of RAMS tools in track maintenance planning and to develop models.
- To develop cost effective maintenance models using RAMS and LCC and to discuss the variation in cost.

1.3 Research Questions

A literature review, in addition to our own industrial experience, and discussions with personnel within the railway infrastructure managers and the manufacturers gave rise to many interesting research areas in the field of RAMS and LCC. On the basis of the stated interests of the companies participating in this project, the following research questions were formulated:

1. How can RAMS targets be estimated for the railway infrastructure with the specified infrastructure capacity and safety requirements?
2. How can RAMS analysis be applied in railway infrastructure maintenance planning to achieve the RAMS targets?
3. How is LCC analysis used in combination with RAMS analysis to realize a cost effective maintenance policy?
4. How are the uncertainties associated with LCC analysis estimated?

1.4 Scope and Delimitations of the Study

Rail infrastructure consists of various sub-systems, such as the track system, the signalling and telecommunication system, and the power system. Each of these sub-systems contributes to infrastructure the RAMS and LCC of the infrastructure. In this research we considered only the track system and part of the signalling systems. Although RAMS and LCC activities are applicable in all the phases of the system life cycle, this thesis considers the RAMS and LCC methodologies only in the operation and maintenance phase.

The thesis presents different maintenance models for the railway infrastructure. However, the models do not consider how the interactive effect of the different systems on each other influences the maintenance policy. The models provide decision support for the infrastructure manager when making decisions on the inspection and maintenance so as to achieve the RAMS targets for the infrastructure.

2 Basic Concepts and Definitions

The optimisation of infrastructure constructions or infrastructure components regarding technical and economic requirements is essential for railway companies to fit into the market and to compete against other means of transport. Due to the long lifetime of the track and track components, pre-installation technical and economic assessments are necessary to optimise the track construction and obtain the return on investment (ROI) in a manageable timeframe. LCC and RAMS techniques are two acknowledged methods for assisting the optimisation process. In the past decade, RAMS and LCC analyses in the railway sector have attracted much more attention than before which has been demonstrated by many research reports and has led to the development of commercial applications. This increased interest in RAMS and LCC resulted in a number of projects being initiated at the European level in the field of railway engineering, but not specifically dealing with the track e.g. Cost, Reliability, Maintenance, and Availability (CRMA, 1998); Maintainability Management in European Rail Transport (REMAIN, 1998); IMPROVEd tools for RAILway capacity and access management (IMPROVERAIL, 2003); Progress in Maintenance and Management of Infrastructure (ProM@in, 2003); Light Rail Thematic Network (LibeRTiN, 2005); and Innovative Modular Vehicle Concepts for an Integrated European Railway System (MODTRAIN, 2007). The objective of CRMA (1998) was to develop LCC methodologies for rolling stock and to identify the parameters required to calculate the LCC, whereas REMAIN (1998) focused on condition monitoring and RAMS management for switches. The objective of IMPROVERAIL (2003) was to improve the existing LCC calculating methods by including costs due to vehicle infrastructure interaction and external costs, e.g. delay costs, accident costs, environmental costs, etc. ProM@in (2003) provided a comprehensive overview of RAMS and LCC analysis applied to railway infrastructure. The project provided an overview of LCC based maintenance planning, RAMS based track inspection and RAMS databases. LiberTiN (2005) discussed the use of LCC and RAMS principles in contracts. An LCC working group in UNIFE (2001) provided guidelines for LCC for total railway systems. It also developed the "UNILIFE-UNIDATA" LCC model for rolling stock. There are a few ongoing projects specific to track infrastructure; e.g. Lasting Infrastructure Cost Benchmarking (LICB, 2007), Urban Track (2007) and INNOTRACK (2009) deal with LCC. There is also some related literature, e.g. Burstrom *et al*. (1994), Stalder (2001), Zoeteman (2001), Esveld (2001), Zoeteman (2006) and Zhao (2006). The main focus of these publications (project descriptions and other literature) is on developing LCC calculation methodologies, as well as the use of LCC in maintenance planning.

However, these publications do not discuss the issues of LCC reduction and the integration of LCC concepts with RAMS. The application of RAMS has not yet been explored fully from the railway infrastructure perspective. In this chapter the basic concepts of RAMS and LCC are described and the application of RAMS and LCC concepts to railway infrastructure maintenance planning is addressed.

2.1 Reliability, Availability, Maintainability & Safety (RAMS)

Reliability and maintainability management is attracting new interest in today's corporate world. The quest to remain competitive and provide timely and accurate services is partly responsible for this interest. A company cannot adopt a rapid response strategy if its system is unavailable and unreliable (Madu, 2005). As engineering disciplines, reliability and maintainability are relatively new. Reliability and maintainability are not only important parts of the engineering design process but also necessary functions in life-cycle costing, cost benefit analysis, operational capability studies, repair and facility resourcing, the determination of inventory and spare parts requirements, replacement decisions, and the establishment of preventive maintenance programmes.

The first European standard (EN 50126) for the railway system in this context was published in 1999 by CENELEC and defines Reliability, Availability, Maintainability and Safety (RAMS) as a characteristic of a system's long term operation. The standard states that RAMS is achieved by the application of established engineering concepts, methods, tools and techniques throughout the life cycle of the system.

EN 50126 (1999) defines the basic RAMS elements as:

- *Reliability*: the probability that an item can perform a required function under given conditions for a given time interval.
- *Availability*: the ability of a product 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.
- *Maintainability*: the probability that a given active maintenance action, for an item under given conditions of use, can be carried out within a stated time interval when the maintenance is performed under stated conditions and using stated procedures and resources.
- *Safety*: the state of technical system freedom from unacceptable risk of harm.

2.1.1 RAMS Parameters

A thorough understanding of the technical description of the system is necessary to perform RAMS analysis of the system. In the case of railway infrastructure, there are three different systems, namely the track system, the signalling and communication system, and the power system. These systems have a combined effect on the degradation of the infrastructure. Each system is subjected to degradation due to various internal and external factors. All these aspects need to be considered to estimate the RAMS of the infrastructure, which makes the calculation more complex. The following sections present some of these factors affecting RAMS. To estimate the RAMS figures at the infrastructure level, one must evaluate the RAMS characteristics at the sub-system and component level. In general, the reliability and maintainability parameters are estimated both on the component level and on the system level, whereas the availability and safety parameters are estimated only on the system level. In order to achieve the required performance of the infrastructure, the failure modes should be identified and classified into the failure categories illustrated in the Table 2.1. A higher RAMS target is set for significant failure, whereas a not-so-high target is set for the minor failure category. Thus, the infrastructure managers should know which failure modes of the track should be given more attention in order to achieve reliability at the system level and incur less cost due to failure.

Table 2.1: RAM failure categories (EN 50126, 1999)

Failure Category	**Definition**
Significant (immobilizing failure)	A failure that - prevents train movement or causes a delay to service greater than specified time and/or generates a cost greater than a specified level
Major (service failure)	A failure that - must be rectified for the system to achieve its specified performance and - does not cause a delay or cost greater than the minimum threshold specified for a significant failure
Minor	A failure that - does not prevent a system achieving its specified performance and - does not meet criteria for significant or major failures

2.1.2 Factors Affecting RAMS

To achieve a dependable system, the factors which could influence the RAMS of the system need to be identified, their effects need to be assessed, and the causes of these effects need to be managed throughout the life cycle of the system. The RAMS of a railway system is influenced in three ways:

- *System conditions*: the sources of failures are introduced internally within the system at any phase of the railway system life cycle. These failures are incurred by the design and manufacturing of the components or the system.
- *Operating conditions*: the sources of failures result from the operating system methodology. These failures are also incurred by environmental conditions.

- *Maintenance conditions*: the sources of failures are caused by maintenance actions. Failures in the railway infrastructure can be caused not only by the maintenance actions on the infrastructure but also by the maintenance of rolling stocks.

These sources of failure can interact with each other and the relationship is shown in Fig. 2.1. In the figure, it can be seen that reliability is not explicitly shown, but is given through the group of internal and external failures in the system. The factors that influence RAMS, as shown in the figure, are generic and can be applied across all industrial applications with some applications, in transport systems. In order to achieve a dependable track system, the factors specifically affecting the track RAMS need to be identified. Table 2.2 identifies the specific factors that affect the track RAMS.

The factors mentioned above affect the characteristics of RAMS. Similarly, the quality of RAMS data affects the correctness of the RAMS estimation. Many types of data are relevant to the estimation and prediction of reliability, availability, and maintainability. Not all are collected in many instances, and the lack of information is sometimes a serious problem in RAMS analysis (Blischke and Murthy, 2003). Markeset and Kumar (2003) described some of the factors influencing the management of RAMS data. These factors concerned user skills and capabilities, and locations, etc., apart from the data type, data format and detail level. The different physical parameters that affect the track RAMS are listed in Table 2.2. In order to assess the effect of these parameters on the track RAMS, it is important to know the technical characteristics of these parameters.

Table 2.2: Factors affecting track RAMS

	Physical parameters	Technical parameters
System conditions	Track curvature (transient curve in, transient curve out, radius)	Quasi-static stress
	Track gradients (start, end, value)	Quasi-static stress
	Rail (rail type, jointed or welded)	Yield strength (Young's modulus)
	Ballast (ballast type, ballast size)	Stiffness, Damping
	Sleeper (sleeper type, sleeper spacing)	Stiffness, Damping, Bending stress
	Fastener (fastener type)	Damping
	Subgrade (geological condition)	Stiffness, Damping
Operating conditions	Track operating conditions:	
	Loads (annual MGT, maximum axle load)	Bending stress, Shear stress, Contact stress
	Environment (temperature)	Thermal stress
	Vehicle operating conditions:	
	Speed of trains	Vertical stress, Lateral stress
	Vehicle condition (hollow wheels)	Dynamic stress
Maintenance conditions	Grinding	Wear rate
	Tamping	Change in track stiffness
	Lubrication	Change in friction co-efficient
	Renewal of track components	Interval of renewal
	Corrective replacements of track components	Failure rate of track components

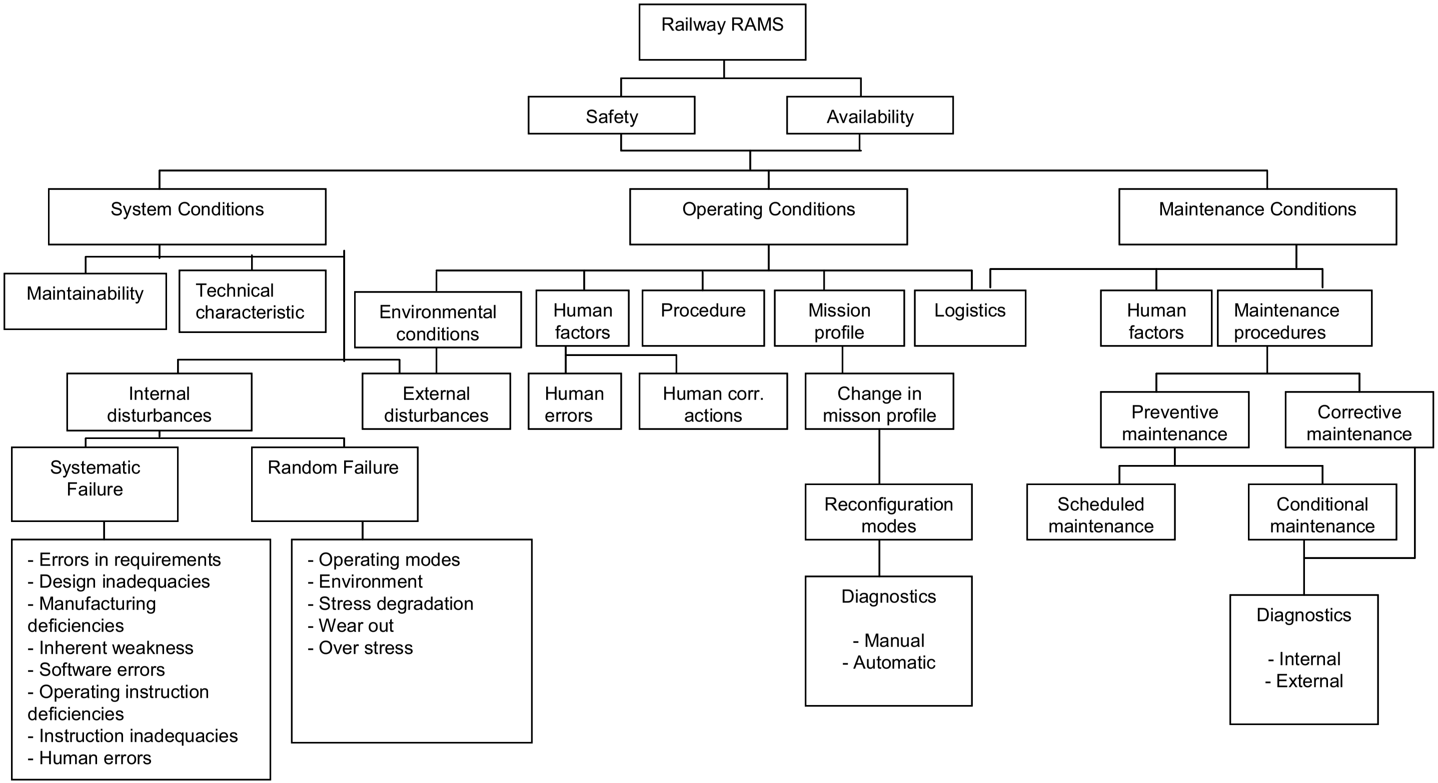


Figure 2.1: Factors influencing railway RAMS (EN 50126, 1999)

For example, in order to estimate the effect of the trainload on the RAMS characteristics of the track, one must know the bending stress, the shear stress and the contact stress imparted by the track load on the track. Similarly, the sleeper types and spacing determine the bending stress, stiffness and damping of the track. We can state that the technical parameters are the causes of the physical parameters which directly affect the track RAMS. The system conditions are mostly related to the design and manufacturing of the track components, whereas the operating conditions are connected to the rolling stock operations. In most of the cases it is difficult to change the system conditions and operating conditions of the track system in the operation and maintenance phase of the track, although sometimes the operating conditions (e.g. changes in the axle load) can change because of changes in railway regulations.

In order to identify relevant maintenance actions for the track, it is crucial to have a good understanding of the failure modes and their causes on the track. Failure Mode, Effects and Criticality Analysis (FMECA) acts as a tool to reveal these failure mechanisms. To support this task, four failure progressions are defined, namely (Jovanovic, 2006):

1. The component is subjected to gradual degradation which may be observed by suitable equipment.
2. The component is subjected to gradual degradation which cannot be observed.
3. The component is subjected to a sudden degradation which can be observed by suitable equipment.
4. The component is subjected to shock degradation, immediately leading to failure.

This classification is particularly useful when reliability parameters are being assessed, as these parameters have different interpretations for the four categories of failure progressions. The track failure modes and their corresponding limits can be categorised as stated below (Esveld, 2001). The limits can be set per unit length of the track.

- Track geometry condition (standard deviation in the longitudinal level, standard deviation in the cross level, and standard deviation in the gauge).
- Ballast condition (percentage of weedy ballast and percentage of surface soiling).
- Fastening condition (percentage of loose fastenings).
- Sleeper condition (percentage of bad sleepers, percentage of medium sleepers, and percentage of good sleepers).
- Rail defects (number of RCF defects, number of rail breaks, number of weld defects, and amplitude of corrugation).
- Rail wear (vertical wear of the rail head and lateral wear of the rail head).

Measuring the infrastructure condition is prerequisite for track maintenance planning. Banverket has a number of condition indices to describe the condition of their infrastructure facilities (Andersson, 2002). The main condition indices are known as the K-value and the Q-value. These are calculated from detailed inspection car measurements of the track. The inspection car measures the relative rail position (the lateral and vertical position), the rail profile and the rail gauge. The Q-value is a weighted index of the standard deviation of two inspection car measures calculated the deviation from the geometric comfort limits set for the specific track class. The Q-value is calculated per kilometre of track as:

$$Q = 150 - 100\left[\frac{\sigma_H}{\sigma_{H\lim}} + 2.\frac{\sigma_S}{\sigma_{S\lim}}\right]/3 \qquad 2.1$$

where σ_H and σ_S are the average standard deviation of the height and interaction on the section measured. The standard deviation for the interaction is calculated as a combined effect of the cant and the side position of the rail. σ_{Hlim} and σ_{Slim} are the comfort limits for a given track class. Track class classifications are based on the speed of the train

The K-value is calculated for a longer section of the track and is expressed as:

$$K = \frac{\sum l}{L}.100\% \qquad 2.2$$

where $\sum$l is the sum of the track length where all the σ values are below the comfort limits for a given track class and L is the total length of the track considered. The K-value is not suitable for shorter track sections.

The failure modes mentioned above must be categorised as per the failure categories given in Table 2.1 so as to proceed with RAMS analysis and define the RAMS targets for different failure categories. The goal of the railway system is to achieve a defined level of rail traffic in a given time, and safely. RAMS has a clear influence on the quality with which the service is delivered to the customer. Moreover, in-service safety and availability can only be achieved by meeting all the reliability and maintainability requirements and controlling the ongoing and long-term maintenance and operational activities and the system environment.

2.2 Infrastructure Maintenance Planning

Maintenance is defined as the combination of all the technical and administrative actions, including supervisory actions, intended to retain an item in, or restore it to, a state where it can perform a required function (IEV 191-01-07, 2007). Maintenance has long been considered as a reactive, “fire-fighting” approach. However, as the dependability targets for assets have become increasingly important, several proactive maintenance approaches and methods are being developed. All the decisions related to rail infrastructure maintenance are taken in order to keep a balance between economic and safety aspects. The goal is to find the effective maintenance procedure to optimise the track possession period and the train speed restriction regime and ultimately increase the track availability.

The different components of the railway asset are structurally and economically interdependent. Scale effects are involved in their maintenance and renewal, while their degradation is often structurally related. As operations have to be continued on the rail network and budgets are often restricted, all kinds of constraints have to be considered in the planning of infrastructure maintenance. The concepts of the maintenance planning process are developed in the following steps (Zoeteman, 2006):

- Generation of maintenance strategies for individual assets (e.g. corrective or preventive, time based or condition based, strategies are distinguished based on the criticality of the individual asset for the entire production system)
- Definition of clustering rules, which optimise the frequencies of activities on the basis of scale or scope effects
- Definition of rules for assigning time windows to maintain packages on the basis of opportunities that occur in the middle or short term.

The initial analytical work for track maintenance was carried out in the early 1980s. Fazio and Prybella (1980) pointed out a number of prerequisites for planning track maintenance. Track quality measures and track deterioration models are highlighted as key areas for a structured planning process to be established. Zarembski (1998) described three tools which railway organisations could use to improve the efficiency of maintenance operations (see Fig. 2.2) automated inspection systems, databases and maintenance planning systems. The lack of integration between these tools has prevented railway organisations from taking full advantage of their potential.

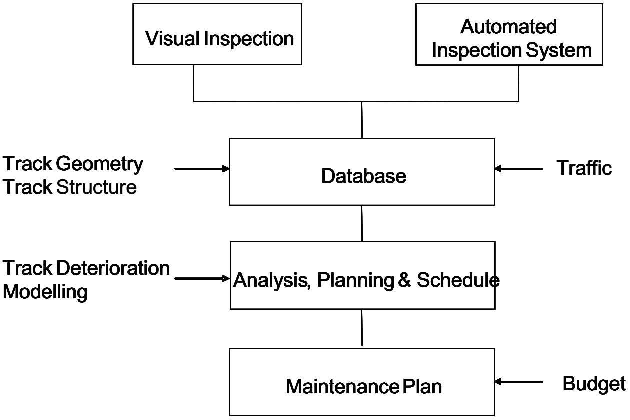


Figure 2.2: Maintenance planning overview (Zarembski, 1998)

These different data sources need to be linked in a general database for planning purposes. By adding models for track deterioration relationships, the state of the infrastructure can be assessed over time. The planning of specific maintenance activities will be affected by the conditions of the track. This requires a detailed knowledge of each component of the track and its relationship with other components of the track, as well as the degradation pattern of each component.

Therefore, the objectives of infrastructure maintenance planning can be described as finding answers to the following questions:

- What are the current conditions of the infrastructure? (Track quality indices)
- What will be needed in the short term as well as the long term as far as maintenance is concerned? (Forecasting of maintenance actions)
- What should be done first? (Prioritization of maintenance activities)

Table 2.3 illustrates the effect of the grinding strategy of the Canadian Pacific Railway on the reliability of the rail. It can be seen in the table that, as the grinding strategy moves from corrective to preventive grinding, the rail life increases considerably. Corrective grinding requires deep and infrequent cuts, whereas preventive grinding requires thin but more frequent cuts (Kalousek *et al.*, 1989). Generally for heavy haul railways, the minimum interval for rail grinding is in the range of 10-15 million gross tonnes (Canon *et al.*, 2003).

The wear rate in grinding is the parameter that controls the rail life, because as the wear reaches the maintenance/safety limit of the rail, the rail needs replacement. No grinding is a scenario where the life of the rail is determined mostly by RCF. Table 2.3 also gives a comparison of the fatigue lives in three grinding scenarios. The fatigue life of the rail is reached when the number of RCF defects in a specific track section reaches its limit.

Table 2.3: Grinding strategy vs. rail life for Canadian Pacific Railway (Magel and Sroba, 2007)

Wear Criteria	**No grinding**	**Corrective grinding**	**Preventive grinding**
Rail wear rate in mm/MGT	0.04	0.06	0.03
Rail life in MGT	469	367	844
Rail fatigue life in MGT	331	496	1322

The difference between the rail wear life and the rail fatigue life is illustrated in Fig. 2.3. As the material removal rate of wear and grinding increases, the rail wear life decreases as the wear approaches the maintenance/ safety limit of the rail. However, an increase in the material removal rate of wear and grinding increases the rail RCF life, because grinding and wear take away the RCF generated cracks before they become critical to the rail. Therefore, the grinding strategy (the wear rate) is seen to be an important parameter that affects the reliability of the rail. The figure also illustrates the “magic wear rate” phenomenon. The magic wear rate is the wear rate that the preventive grinding strategy should attain in order to achieve the highest reliability for the rail. As shown in the figure, when the wear rate is below the magic wear rate, the rail life is determined by the rail’s RCF life, whereas, when the wear rate is higher than the magic wear rate, the rail life is determined by the rail wear life.

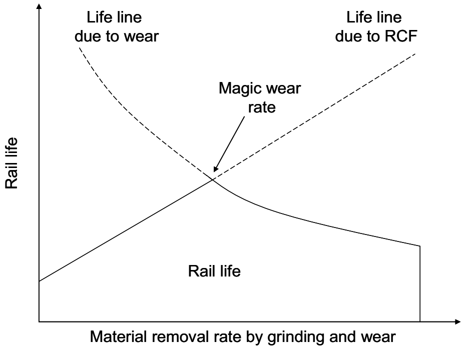


Figure 2.3: RCF, wear and rail life relationship (Magel and Sroba, 2007)

In order to assess the effects of the maintenance conditions on the reliability of the track system, it is necessary to consider their combined effect on the system. As described above, grinding affects the reliability of the rail. However, to perform an effective reliability analysis of the rail, the combined effects of other maintenance conditions, e.g. lubrication, rail replacements, etc., should be taken into account. For example, lubrication reduces the rail wear, especially in the track curves (Diamond and Wolf, 2002), and thereby increases the reliability of the rail. However, at the same time, lubrication is a factor for RCF defects, which are removed by grinding (Rinsberg, 2001). In order to fulfil the above mentioned objectives, RAMS analysis will play a major role in maintenance planning. The details are presented in the following sections of the chapter.

2.2.1 RAMS Analysis for Maintenance

RAMS analysis is a process which utilises the failure information from a system in order to develop probability distributions for the system’s ability to perform its intended functions. RAMS analysis for the track is based on the following elements:

- RAMS database
- Failure modes

- Methods and tools for RAMS analysis

The utilisation of failure and maintenance data is an important factor in RAMS analysis and the management of the system. There are several dimensions with respect to the collection of RAMS data. One should ascertain that the data being collected support all the types of RAMS analysis required for the system. Another important aspect is that the data should support the life cycle perspective of the system and, more importantly, the maintenance phase in this case. Fig. 2.4 illustrates the use of a RAMS database as a feedback to the RAMS analysis, as well as to the operation and maintenance phase of the system life cycle.

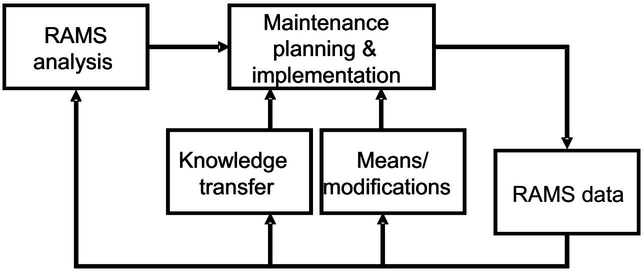


Figure 2.4: A process loop showing the use of a RAMS database in RAMS analysis (Prom@in, 2003)

For an effective analysis of RAMS, the traffic and track geometry databases should be considered along with the failure and maintenance databases as mentioned above. Therefore, the track must be divided into homogenous analysis segments with respect to the track curvature, grade, super elevation, traffic density, etc. The following data are also a part of the RAMS database, along with the failure data (Esveld, 2001).

I) *Layout and operating data*

- Curves (start and end km, transition curves, radius, etc.)
- Loads (annual load (MGT), maximum axle load (tonnes), date from which the data are valid, etc.)
- Speeds (speed of freight and passenger trains, date from which the speed is valid, etc.)
- Gradients (start, end, and value)

II) *Infrastructure*

- Subgrade (geological conditions, various monitored parameters, etc.)
- Ballast (ballast type, date of installation, ballast thickness, etc.)
- Sleepers (sleeper type, sleeper spacing, new/old sleepers when laid, type of fastenings, and date of installation)
- Rails (rail type, joined or welded track, weld type, date of installation, new/old rails when laid, date of installation, and cumulative tonnage on rails when installed)

III) *Work history*

- Renewals, grinding and tamping work history (start km, end km, and type)
- Speed restriction history (start and end date of temporary speed restriction, and value of reduced speed)
- Spot maintenance history (type, and date)

Banverket uses BIS (a track information system), BESSY (an inspection system), 0FELIA (a fault analysis system) and TFÖR (a train delay system) for maintenance planning. Recent developments are HANNES (a speed restriction system) and a track geometry data system. These existing systems are more or less stand alone modules using BIS as a reference system (Andersson, 2002). Therefore, there is a need to integrate the different databases for efficient analysis of RAMS and maintenance planning. It is also necessary to investigate the effect of track degradation/failure on the other systems of the railway infrastructure, e.g. the signalling systems. It is evident that, to achieve the availability target of the overall railway infrastructure, we need to assure the availability of all the systems that are part of the infrastructure.

2.2.2 RAMS in the Operation and Maintenance Phase

The system life cycle is a sequence of phases, each containing tasks, covering the total life of a system from the initial concept to decommissioning and disposal. The life cycle provides a structure for planning, managing, controlling and monitoring all the aspects of a system, including RAMS, as the system progresses through the phases, in order to deliver the right product at the right price within the agreed time scales. A system life cycle that is appropriate in the context of railway operation is shown in Fig. 2.5. The top-down branch (on the left-hand side) is generally called design and development and is a refining process ending with the manufacturing of system components. The bottom-up branch (on the right-hand side) is related to the assembly, the installation, the receipt and then the operation of the whole system. The "V" representation assumes that the activities of acceptance are intrinsically linked to the design and development activities, insofar as what is actually designed has to be finally checked in regard to the requirements. Therefore, the validation activities for acceptance at various stages of a system are based on the specification of the system and should be planned in the earlier stages, i.e. starting at the corresponding design and development phases of the life cycle.

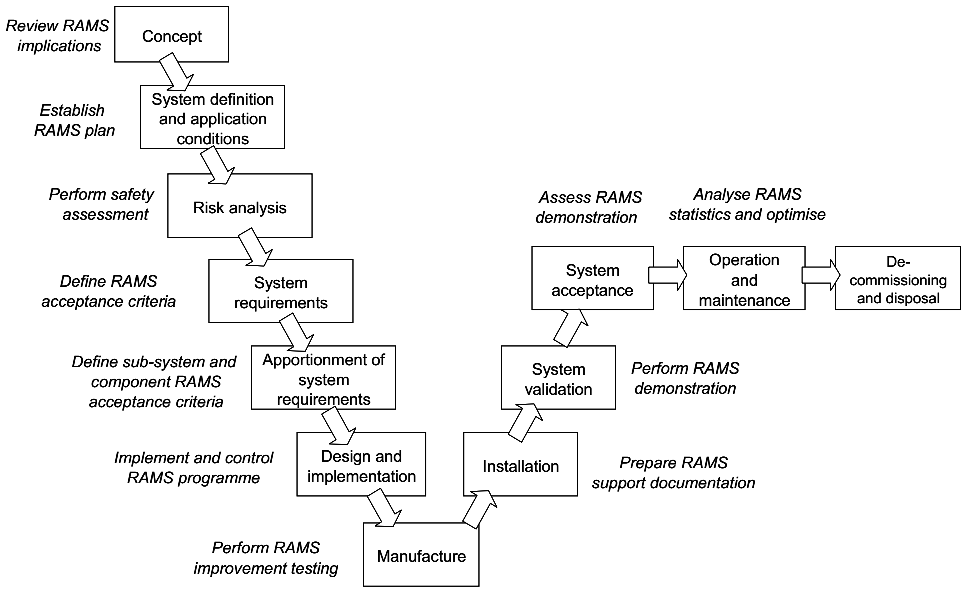


Figure 2.5: The "V" representation of the RAMS life cycle (IEC 62278, 2002)

The figure also describes the various RAMS activities being carried out at each phase of the system life cycle (IEC 62278 2002). To achieve the overall RAMS objectives of the system, it is important to follow systematic RAMS actions throughout the life cycle of the system. As far as RAMS activities are concerned, one of the important phases of the system life cycle is the operation and maintenance phase, where RAMS is optimised by the analysis of real life failure data.

The objective of this phase is to operate, maintain and support the total combination of components and subsystems in such a way that compliance with the system RAMS requirements is maintained. Fig. 2.6 illustrates the RAMS process for the railway infrastructure in the operation and maintenance phase of the system life cycle. It is a continuous improvement process throughout the operation and maintenance phase. The sources of failures are due to the system itself, train operation or maintenance activities carried out on the track. The failure data are collected by FRACAS (Failure Reporting And Corrective Action System). FRACAS is a closed-loop reporting system for identifying failure modes and their root causes and subsequently determining effective corrective actions for eliminating their re-occurrence.

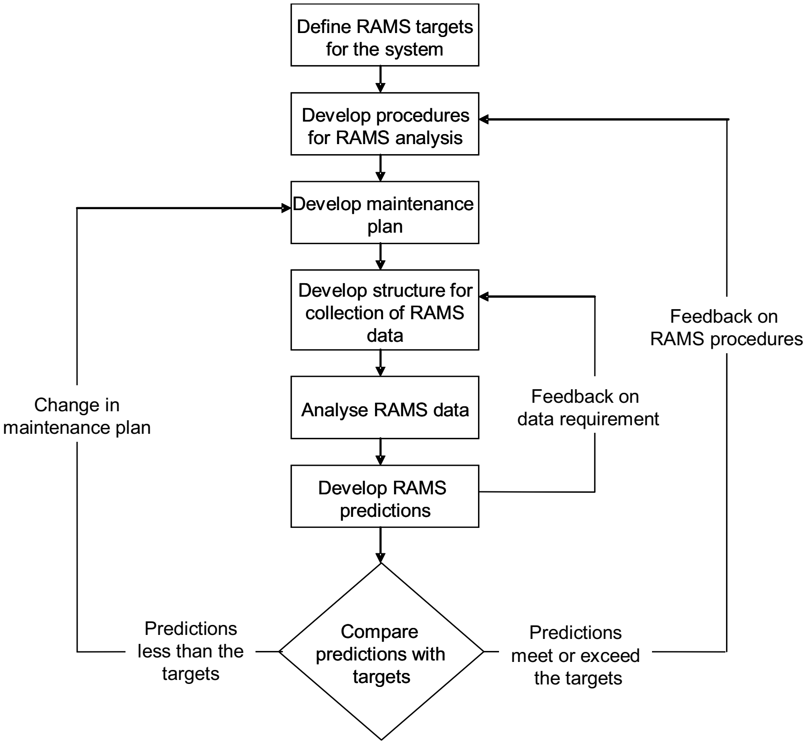


Figure 2.6: RAMS process in the operation and maintenance phase

Failure Mode, Effects and Criticalily Analysis (FMECA) is an analysis method involving two elements of risk namely failure frequency and consequence. FMECA analysis concentrates on the identification of the events and frequency resulting in failures and on analysing their effects on the components and systems. FMECA categorises the failures as non-safety critical failures and safety critical failures. Other tools that are being used for RAMS analysis are Reliability Centred Maintenance (RCM), Markov analysis, etc. The basic objective of the operation and maintenance phase is to monitor RAMS activities in order to meet the RAMS goals set for the infrastructure. The performance indicators for checking the goals are the RAMS parameters described earlier in this chapter. If the goals are not met at any point of time, then changes in the maintenance conditions are made in order to meet the goals. If the infrastructure operating conditions change during the operation and maintenance phase, changes in the maintenance conditions are required accordingly to meet the RAMS goals.

RAMS analysis of the infrastructure should not be performed without considering the operational characteristics of the rolling stock. As stated in EN 50126 (1999), the operational availability of the track hardly considers the train schedule. To have a realistic measure of the availability of the infrastructure, it is necessary to consider the demand availability when dealing with the operational availability. The demand availability is the probability that a system will be in a functioning state on demand (Kumar and Akersten, 2008). In the case of track, the demand availability defines that a unit length of track is available when trains pass over it. To achieve the demand availability of the track section, the following measures must be considered:

- The corrective maintenance on the track must be reduced. As failures on the track can occur at random, the lower the number of failures are, the better is the demand availability.
- All the preventive maintenance and renewal actions on the track must be carried out in the train free periods. The maintenance plans for the track need to utilise the train free periods to a maximum for all the maintenance actions.

In order to calculate the demand availability of a track section over a period of time, the reliability and maintainability of the track, along with the train timetable need to be considered.

2.3 Maintenance Optimisation

As rail infrastructure is an expensive asset with a long lifespan, the cost-effectiveness of long term design and maintenance decisions should be guaranteed. LCC analysis, an engineering economics technique, can be utilised to focus on maintenance strategies to minimise the life cycle cost, while meeting the dependability requirements. Fig. 2.7 depicts LCC calculations as being based on the business and technical requirements of the infrastrcuture, which are based on a specific operational scenario. The maintenance policy and budget constraints play a major role in selecting the alternative maintenance strategies. They act as a crucial input when deciding upon a particular maintenance strategy. Fig. 2.7 shows how RAMS affects the LCC calculation at various stages. The maintenance strategy (MS) with the lowest LCC is considered as the cost-effective solution to be implemented in the infrastructure operations. The maintenance strategy can be a single maintenance action (e.g. grinding) or a cluster of maintenance actions. For an effective decision on maintenance strategy, it is important to consider a cluster of maintenance actions for LCC calculations, as maintenance actions affect each other.

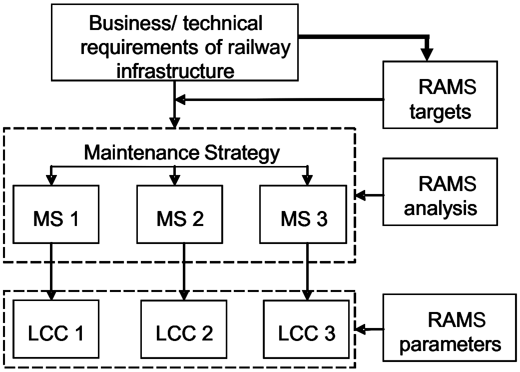


Figure 2.7: Maintenance decisions based on LCC

When considering the maintenance strategy with the lowest LCC as the cost-effective solution, it is important to consider the RAMS figures associated with that particular maintenance strategy. Therefore, when not considering the maintenance strategy with the lowest LCC as the best solution, a trade-off between the RAMS targets and the LCC value is necessary in order to achieve an effective maintenance strategy.

2.3.1 Life Cycle Cost Analysis

In many countries, the restructuring of railways and increasing efficiency and effectiveness requirements are causing a changing environment for infrastructure management. The responsibility for parts of the railway system is often handed over to different actors. In order to guarantee optimal long-term results for the railway systems, the effects of decisions should be systematically evaluated (Zoeteman, 1999). Putallaz (2003) states the three parameters (see Fig. 2.8) that influence the performance of the track infrastructure: capacity, substance and quality. The capacity of the infrastructure may be expressed in terms of usable train paths during a certain time span. The substance of the infrastructure refers to the average remaining useful lifetime of its components. Finally, the quality of the infrastructure represents the quality of track's geometry and components. Managing the infrastructure comes down to setting those three parameters at their most appropriate level, in order to maximize efficiency. Adjustments may be made to the capacity through the investment policy, to the infrastructure substance through the renewal policy, and to the quality through the maintenance policy. These three parameters cannot be adjusted independently. An old infrastructure (low substance) requires more maintenance (to increase the quality), whereas a bad geometry (low quality) increases the wear on the infrastructure (lower substance). Similarly, more engineering works (maintenance & renewal) require more track possessions (less capacity), while more traffic (high capacity) induces more wear of the infrastructure.

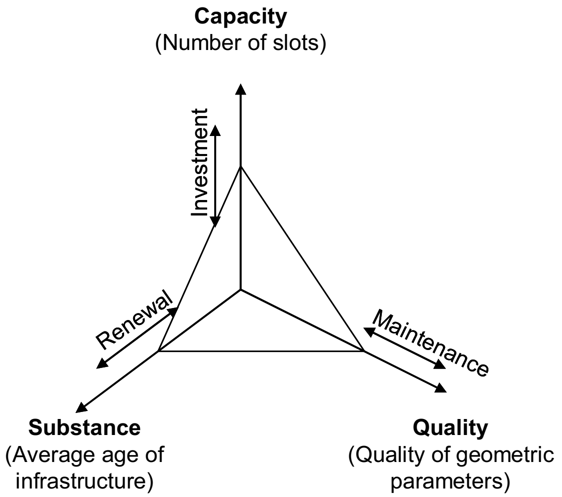


Figure 2.8: Three basic parameters of rail infrastructure influencing performance (Putallaz, 2003)

When adjusting these performance parameters simultaneously, the cost aspect of each activity should be considered. Since, in the current scenario, most of the maintenance and renewal decisions are based on cost models that rarely consider their effects on the whole life of the infrastructure, a need for a life cycle cost approach arises. Life cycle cost can be used as a tool to take cost-effective decisions on investment, renewal and maintenance, in order to adjust these three parameters to optimise the infrastructure performance.

The total costs can be observed from diverse points of view, i.e. from the viewpoint of the system's supplier or of the system's user or owner, or, even more broadly, from the point of view of society. A basic assumption providing motivation for the LCC approach is that it is usually possible to affect the future costs of a product beforehand, either by planning its use or by improving the product or asset itself (Markeset and Kumar, 2004). Asiedu and Gu (1998) stated that LCC analysis should be regarded not only as an approach for determining the cost of the system, but also as an aid for decision making in design, maintenance, etc. The use of life cycle cost analysis should therefore be restricted to the cost that we can control.

In order to be able to estimate the life cycle costs of the rail infrastructure, the factors influencing the performance of the railway infrastructure and their relationships need to be identified. The driving factor causing failures and maintenance is the degradation of the asset. Track degradation depends on many factors, such as the initial quality of the construction, the quality of the substructure and the loads on the track. Besides asset degradation, there are other factors that also influence the life cycle costs, such as the RAMS targets for the track, the amount of preventive maintenance, the market prices of labour, materials and machines, and the operational characteristics of the line (such as the axle loads, the traffic intensities and the duration of train free periods). The infrastructure manager can control some of these factors directly (e.g. the maintenance strategy) or with the cooperation of the transport operators (e.g. the quality of the rolling stock) and the government (e.g. negotiated grants) (Zoeteman, 2001). The performance of the railway infrastructure is influenced by factors such as the level of safety, riding comfort, noise, vibrations, reliability, availability, and the costs of ownership (see Fig. 2.9). Safety and noise standards indirectly influence the life cycle costs, since they determine the tolerances and thresholds for the design and maintenance parameters. The physical design influences the asset degradation, together with other conditions, such as the traffic intensities and axle loads, the quality of the substructure and the effectiveness of the performed maintenance. The quality of the geometric structure determines the required volume of maintenance and renewal (M&R). The chosen maintenance strategy also influences the amount of M&R. The realised M&R volume causes expenditures and planned possessions. The maintenance strategy also has a direct impact on the life cycle cost. The incident management organisation, the realised M&R volume and the transport concept determine the train delay minutes caused by the infrastructure and these train delay minutes can be converted into penalties for the infrastructure managers. The cost models used in the decision support systems or maintenance management systems should be able to provide means to evaluate and compare the costs and benefits of different maintenance strategies and options. In order to carry out an economic analysis, it is necessary to make adjustments to costs to ensure that they are all measured in the same units and represent real costs of resources (Larsson, 2002). According to Zoeteman (2001), the life cycle cost can be presented in three different ways, i) the total present value (TPV), ii) the internal rate of return (IRR), and iii) the annual equivalent or annuity (ANN).

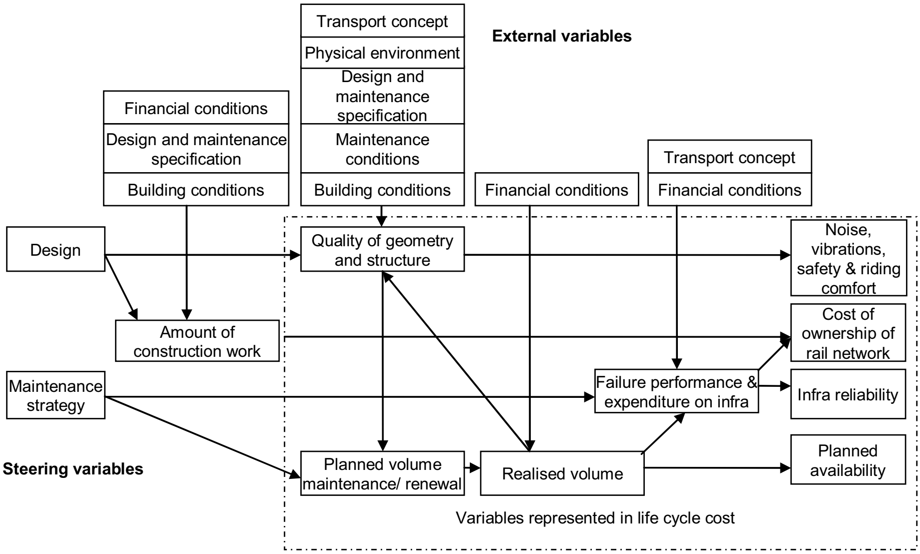


Figure 2.9: Factors influencing the performance of track infrastructure (Zoeteman, 2001)

To check the robustness of LCC models, two methods are used:

- Sensitivity analysis: The disadvantage of sensitivity analysis is that only one variable is tested at a time. Hence, possible interactions between factors are not revealed.

- Uncertainty analysis: In this approach the input parameters of the LCC model are considered to be random variables from which samples are drawn. Simulation techniques are used to determine the interaction of the input parameters with the outcomes.

The life cycle cost of the track infrastructure depends mainly on two aspects of the infrastructure, i.e. the network configuration and complexity and the network utilisation. Complexity is a predominant parameter for investment and the cost of maintenance. Some major indicators are (LICB, 2007):

- Density of switches
- Length of lines on bridges and tunnels
- Lengths of double track lines
- Degree of electrification

In addition configuration parameters like curvature, axle loads and speed level have their impact on the life cycle expenditure.

The utilisation of networks has a strong impact on the cost of maintenance and on the components' technical life until replacement. Some major indicators are:

- Average frequencies of trains per year
- Average gross tonnage per year (freight and passengers)

It is difficult to generalise the LCC per kilometre of track because of the track's variability in terms of complexity and utilisation. A harmonisation model is used to compare the cost data of different track configurations and utilities in the best possible way (Stalder, 2001). Various aspects of the harmonisation model are given by:

- *Single vs. multiple track*: The maintenance and renewal of single-track lines require more work per kilometre than that of double or multiple track lines (e.g. for work site logistics and preparatory work). Based on a detailed analysis of data from the French National Railway Company (SNCF) and surveys of other railways, it is concluded that the cost of maintenance per track kilometre on single track is typically 40% higher than that on double track. Therefore, this aspect should be taken into account when estimating the LCC per track kilometre.

- *Switch densities*: The switches on the main track have a major share in the cost of the track maintenance (with a high impact on the signalling and the power supply). With the switch densities varying between the main tracks, the need for harmonisation is evident.

- *Track utilisation*: The maintenance and renewal as well as the lifetimes of track elements depend heavily on the utilisation of networks. Data analysis has proven that maintenance expenditures can best be harmonised according to train frequencies, in particular because of the strong correlation between the track access times and the maintenance cost. Renewal expenditures are harmonised according to the gross tonnage, which has a great impact on the wear and tear of the track.

2.3.2 Maintenance Management

The asset strategy is the maintenance approach and plan developed for each item of the system. This strategy determines what planned and programmed maintenance work should be carried out, and considers what potential problems may require an unplanned, reactive response (Wilson, 1999). The approach to be taken in developing the maintenance activities and, on the basis of these, the asset strategies needs to be understood before the maintenance management strategy is completed. This is because, until it is known how much maintenance activity will be required, who will carry out, and what spare parts will be used, etc., then the approach to the organisation of maintenance activity cannot be finalised.

Infrastructure managers try to ensure the successful management of costs and quality, and the relation between the two. This is essential because the train operators as well as the passengers are imposing ever increasing quality requirements on the rail infrastructure. Therefore, the infrastructure managers require the best infrastructure quality at the lowest cost. The way to achieve this objective is through proper maintenance management.

Esveld (2001) gave examples of the type of data required for the Track Maintenance Management System (TMMS), as listed below:

- Measurements
- Planning
- Infrastructure
- Inspections
- Work carried out
- Costs

However, difficulties in the accurate anticipation of maintenance prevent extremely precise maintenance planning and management. Besides, the amount of funding allocated for maintenance work is often regarded as a compromise, as too much according to top management, and too little according to the operating and maintenance staff. Consequently, the selection of the optimal maintenance strategy can be challenging. A systematic approach for the determination of the deterioration of track components is necessary to gauge fully the status of the track system and components. This will require proper track condition assessments, the establishment of a standard condition rating system, and the development and regular updating of prediction models for various track components.

Esveld (2001) discussed the idea of rational rail management for infrastructure. Rational rail management aims at the objective evaluation of the qualitative and quantitative assessment of the rail infrastructure, after which, based on the system objectives, and on rules and standards, decisions may be taken regarding the maintenance and renewal of rail infrastructure. Rational rail management is summarised in the following objectives:

- To be less dependent on the individual know-how of co-workers
- To create optimal working conditions regarding business economy economics
- To bear responsibility and to report to the management

Therefore, an effective infrastructure maintenance management system requires RAMS management and life cycle cost management to be thoroughly integrated into the asset management of the system. The idea of system-effectiveness emerges so as to make the LCC analysis cost-effective. System-effectiveness concerns with RAMS characteristics of the system. Blanchard and Fabrycky (1998) define system-effectiveness as the probability that a system will successfully meet an overall operational demand within a given time and when operated under specified conditions. In short, system-effectiveness is the ability of a system to perform a job for which it is intended. It can be defined as a function of the system's operational availability, operational reliability and capability.

Operational availability of the infrastructure is defined as the probability that the infrastructure will be operationally available during the train traffic.

Operational reliability is the probability that during the train traffic operation, the infrastructure will not suffer from any failures.

Capability is the ability of the infrastructure to meet its required objectives.

System-effectiveness can be defined as:

System-effectiveness = Operational availability *Operational reliability *Capability

The higher the system-effectiveness is, the better the infrastructure is at achieving to achieve its objectives.

Cost-effectiveness analysis yields quantitative results to aid the decision maker with risk analysis and provides a useful decision tool.

Table 2.3 shows the calculation of cost-effectiveness from the LCC values of different alternatives. When taking a decision on maintenance alternatives it is necessary to calculate the cost-effectiveness of different maintenance alternatives. The higher the cost-effectiveness is, the better is the maintenance alternative.

Table 2.3: Cost-effectiveness of maintenance alternatives

Maintenance alternatives	Life Cycle Cost (LCC)	System effectiveness	Cost-effectiveness (SE/LCC)
M1	LCC1	SE1	SE1/LCC1
M2	LCC2	SE2	SE2/LCC2
M3	LCC3	SE3	SE3/LCC3

Fig. 2.10 illustrates the relationship between maintenance management, asset performance and asset maintenance. The asset management of the track concerns two important aspects of the asset, i.e. the asset performance and the asset maintenance. System-effectiveness and cost-effectiveness act as indicators for asset performance. Asset maintenance concerns activities ranging from small scale maintenance actions to the building of new infrastructure.

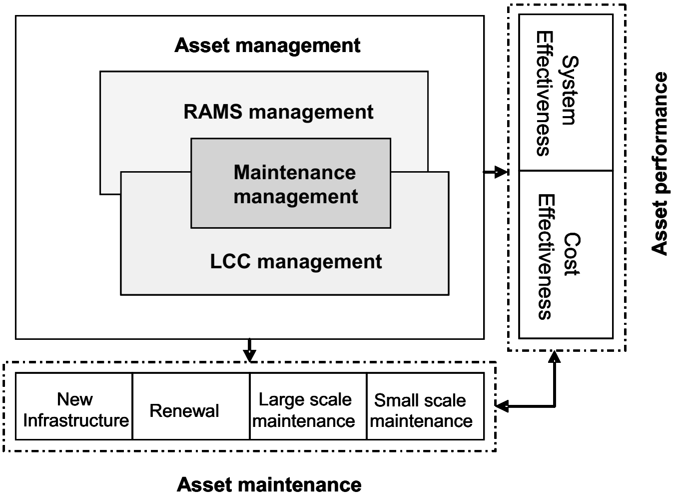


Figure 2.10: Factors influencing maintenance management (Adapted from Swier and Luiten, 2003)

As described in the previous sections, RAMS and LCC analyses act as tools for estimating the system and cost-effectiveness of the asset, as well as for taking effective decisions on the maintenance of the asset. There is a close relation between asset maintenance and asset performance, as effective asset maintenance increases the asset performance, while asset performance acts as a decision tool for asset maintenance.

3 Research Methodology

Research can be defined in many ways. Most generally defined, research is a process through which questions are asked and answered systematically. As a form of criticism, research can include the question of whether or not we are asking the right questions (Dane, 1990). In other words, research is a systematic examination of observed information, performed to find answers to problems. Research methodology is the link between thinking and evidence (Sumser, 2000). To conduct research, it is essential to choose a clear methodology. This provides a framework for integration of the different technical, commercial, and managerial aspects of study. The study of research methodologies provides the researcher with the knowledge and skills that are needed to solve the problems and meet the challenges of a fast-paced decision making environment (Cooper and Schindler, 2006).

There are many ways to carry out research, but the purpose of research can be classified into three main categories i.e. the exploratory purpose (to explore a new topic), the descriptive purpose (to describe a phenomenon) and the explanatory purpose (to explain why something occurs). The details of these are described in Table 3.1.

Table 3.1: Different kinds of research purposes (Neuman, 2003)

Exploratory	Descriptive	Explanatory
- Become familiar with the basic facts, setting, and concerns	- Provide a detailed, highly accurate picture	- Test a theory's predictions or principle
- Create a general mental picture of conditions	- Locate new data that contradict past data	- Elaborate and enrich a theory's explanation
- Formulate and focus questions for future research	- Create a set of categories or classify types	- Extend a theory to new issues or topics
- Generate new ideas, conjectures, or hypotheses	- Clarify a sequence of steps or stages	- Support or refute an explanation or prediction
- Determine the feasibility of conducting research	- Document a casual process of mechanism	- Link issues or topics with a general principle
- Develop techniques for measuring and locating failure data	- Report on the background or context of a situation	- Determine which of several explanations is best

The methodologies used in the present research are both descriptive and exploratory. The research purpose of this study is to describe the methodologies of RAMS and LCC analysis for the railway infrastructure, and to describe the methodologies for utilising both RAMS and LCC analysis in making track maintenance planning decisions.

3.1 Research Approach

Research may be fundamental or applied in nature, depending upon the kind of knowledge sought about a certain area and the solution intended. Fundamental research aims to widen the knowledge of a particular subject so that future research initiatives may be based on the extended knowledge. This research is designed to solve problems of a theoretical nature, with little direct impact on strategic decisions. Applied research addresses existing problems and opportunities (Cooper and Schindler, 2006).

This thesis concerns applied research with a purpose to apply RAMS and LCC methodologies in the railway infrastructure context and to develop models for maintenance decisions based on RAMS and LCC analysis. The knowledge gathered from an extensive literature study and from discussions and consultations with RAMS and LCC experts within Europe was applied to delineate the usefulness of RAMS and LCC analysis in railway infrastructure maintenance planning, so as to make the planning more effective and risk-based.

The research approach can be categorised as induction or deduction (Sullivan, 2001).

- The induction approach uses observations, a knowledge base and empirical data to explain and develop theories. The approach involves inferring something about a whole group or class of objects from our knowledge of one or a few members of the group or class.
- The deduction approach can be applied to generate hypotheses based on existing theories, the results of which are derived by logical conclusions.

The research approach can be quantitative or qualitative. In simple terms, quantitative research uses numbers, counts, and measures of things whereas qualitative research adopts questioning and verbal analysis (Sullivan, 2001).

In the present research, both deductive and inductive approaches have been applied. A deductive approach has been applied to develop a process of RAMS and LCC application in railway infrastructure maintenance, whereas an inductive approach has been applied to develop maintenance models. Both qualitative and quantitative research methodologies have been applied in this research.

3.2 Reliability and Validity

According to Neuman (2003), reliability means dependability or consistency. It suggests that the same things are repeated or reoccur under identical or very similar conditions. Reliability means that the implementation methods of a study, such as data collection procedures, can be applied by somebody else with the same result. Validity is concerned with whether or not the study actually elicits the intended information. Validity suggests fruitfulness and refers to the match between a construct, or the way in which a researcher conceptualises an idea in a conceptual definition, and a measure. It refers to how well an idea about reality fits in with actual reality (Neuman, 2003).

The data and information used in this research have been collected from reputed peer-reviewed journals, refereed conference proceedings and reports, or from company databases, which positively contributes to the research's reliability. Well-established RAMS analysis techniques have been applied in different case studies, which also contributes positively to the reliability of the research.

In this research, different maintenance models have been developed using RAMS and LCC methodologies. The obtained results are believed to support the validity of the research, as they matched the theoretical and logical expectations. These models can be implemented in different railway infrastructure systems in future to support the validity further.

4 Data Collection and Analysis

4.1 Data Collection

Data can be defined as the facts presented to the researchers from the studied environment. Data may be divided into primary and secondary types. Data collected by the researcher for the purpose of study through various experiments or onsite data recording are called primary data. Primary data are sought for their proximity to the truth and control over error. Data collected by other people/organisations and used by the researchers are called secondary data. They have at least one level of interpretation inserted between the event and its recording (Cooper and Schindler, 2006).

Qualitative data were collected through relevant scientific papers and articles from online databases. Relevant books were searched for from Lucia (Luleå University Library's online catalogue) and then perused, and relevant reports and licentiate and PhD theses from various universities were also studied. Different databases were searched to extract both qualitative and quantitative data. Relevant scientific papers and articles were extracted from online databases, such as Elsevier Science Direct, Blackwell Synergy, Emerald, IEEE Xplore, Proquest science journals, Inspec and Compendex, etc. Some of the articles were searched from the references of other relevant articles. Different keywords were used for searching these articles as mentioned in the abstract. Different combinations of these keywords were also used to narrow down the number of hits. Some of the known articles were searched directly from the journal databases. Quantitative data were collected from Banverket's BIS, BESSY and 0felia databases from the Iron Ore Line (Malmbanan), and from the internal databases of ALSTOM Transport, France. Cost-related data were collected through personal consultations with experts at Banverket and ALSTOM Transport. The details of Banverket's databases are given below:

BIS: This is Banverket's infrastructure register (computerized database) containing information about infrastructure or facilities, arranged geographically in accordance with Banverket's facility structure. In BIS, for example, information is collected prior to work on train timetables and work in connection with inspections (Karlsson, 2005). Apart from this, information about agreements, accident reports, the history of tamping and grinding, and curve-information can also be obtained.

BESSY: This is an inspection system in which comments are registered per facility on the completion of inspection. Data are also registered directly during the course of inspection with the aid of a palm computer.

0felia: This is a database containing information on all the faults in the infrastructure that have been registered for a particular railway facility. The faults are sorted on the basis of the structure used in BIS.

Raw data collected from these databases were treated to extract the information that is used in the models. Some of the rail break data was collected from Kumar (2008). Information was also collected through discussions and consultations with experts from Banverket. Failure and degradation data for the track and the track circuit were collected from these databases. The data were tested for trend and for dependency characteristics before proceeding with a specific reliability model. Fig. 4.1 illustrates the steps for failure data analysis before choosing the best fitting model.

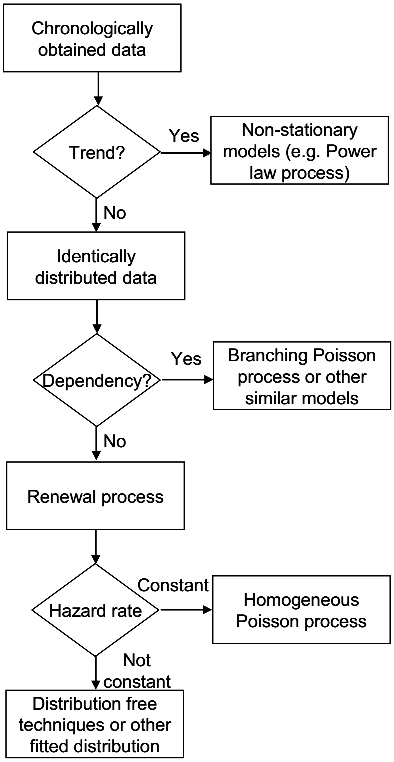


Figure 4.1: Possible exploratory steps in field failure data analysis before fitting distribution models (Asher and Feingold, 1984)

Failure, degradation and maintenance data for the signalling systems were collected from the internal databases of ALSTOM Transport. Details of the databases cannot be described due to confidentiality issues. The data acquired from the ALSTOM databases were altered before they were used in the models to keep their confidentiality. Different types of statistical distributions were examined and their parameters were estimated by using Reliasoft's Weibull ++ 6 software (Reliasoft, 2003).

Researchers generate information by analysing data after their collection. Data analysis is one step, and an important one, in the research process. Data analysis usually involves the reduction of accumulated data to a manageable size, developing summaries, looking for patterns, and applying statistical techniques. Further, the researcher must interpret these findings in the light of the client's research questions or determine if the results are consistent with the hypotheses and theories (Cooper and Schindler, 2006).

4.2 Data Analysis

The usual analysis of reliability and availability data is implicitly based on the assumption that the times between failures (TBFs) and times to repairs (TTRs) are independent and identically distributed in the time domain. Further, it is also assumed that they are independent of each other and form two independent series. This means that the TBFs and TTRs are free from trends and serial correlations. In such situation the recording of the TBFs and TTRs data by magnitude is valid for fitting the various distributions for representing the population of the TBFs and TTRs. The above discussion implies that before any reliability analysis is taken up, the test for trends and serial correlations must be done to check whether the usual assumption of independent and identically distributed (i.i.d) for the data sets are contradicted or not (Kumar, 1989). In this thesis, suspended failure data points have also been considered to estimate the time between degradations. Inspection data have been used to estimate the probability distributions of the degradation times. As the point of occurrence of the degradation between two inspections or between the functional failure and the last inspection is uncertain, the probability distribution of the degradation life can be estimated considering the interval data for degradation.

The failure rate of a repairable component depends on the type of corrective maintenance that is applied in a range stretching from perfect maintenance to minimal maintenance. Perfect maintenance repair brings the component age to zero (i.e. the component becomes as good as new (AGAN)) whereas minimal repair keeps the component's age un-modified (i.e. the component stays as bad as old (ABAO)). In real world cases the repairs are neither AGAN nor ABAO, but are something in between. In (Kijima, 1989), two models (type I and II) are proposed that estimate the virtual age of the component after a repair. Kijima models consider a parameter called the maintenance factor, which estimates the virtual age. If the maintenance factor is 1, the repair is ABAO and for the maintenance factor 0, the repair is AGAN. Model type I assumes that the repairs can only fix the damage incurred during the last period of operation. Thus, the nth repair can only remove the damage incurred during the time between the (n-1)th and the nth failures. Model type II assumes that the repairs fix all of the damage accumulated up to the current time. As a result, the nth repair not only removes the damage incurred during the time between the (n-1)th and the n^{th} failures, but can also fix the cumulative damage incurred during the time from the first failure to the (n-1)th failure.

If the times between the failures are denoted by $x_1, x_2, \ldots x_n$, t the virtual age of the component after the nth repair is given by

Vn = Vn-1 + (maintenance factor * x_n) →Kijima model type I
Vn = (maintenance factor*Vn-1) + (maintenance factor *x_n) →Kijima model type II

The maintenance factors have been calculated from the past failure times of the component by applying the Kijima models discussed above.

In this thesis, various system states (degradation and maintenance) of the system are modelled by Petri-Nets. Petri-Nets (see Fig. 4.2) are a graphical tool for the formal description of the flow of activities in complex systems. In comparison with other more popular techniques for graphical system representation (like block diagrams or logical trees), Petri-Nets are particularly suited to representing in a natural way logical interactions among parts or activities in a system. Typical situations that can be modelled by Petri-Nets are synchronization, sequentiality, concurrency and conflict. The theory of Petri-Nets originated from the doctoral thesis of C.A. Petri in 1962 (Petri, 1962). Since then, the formal language of Petri-Nets has been developed and used in many theoretical and applicative areas. Petri-Nets used for modelling real systems are sometimes referred to as condition/events nets. Places identify the conditions of the parts of the system (working, degraded, or failed), and transitions describe the passage from one condition to another (end of a task, failure, or repair). An event occurs (a transition fires) when all the conditions are satisfied (the input places are marked) and give concession to the event. The occurrence of the event modifies wholly or in part the status of the conditions (marking).

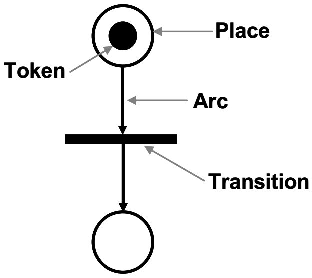


Figure 4.2: Schematic representation of a Petri-Net

The number of tokens in a place can be used to identify the number of resources which are in the condition denoted by that place. Petri-Nets have advantages over the Markov model because of their ability to handle stochastic transition rates. Petri-Net models were constructed using the software tool GRIF.

The statistical characteristics of R&M parameters contribute to uncertainty in LCC. The reason for this is that the times and conditions for these types of events are so complex that they cannot be predicted with a fair degree of accuracy. Therefore, it was decided to explore a methodology that combines the use of design of experiment (DoE) principles with Monte Carlo simulation to estimate the uncertainty involved with LCC. DoE was developed in the twentieth century to increase the effectiveness and efficiency of experimentation. However, for experiments to be effective and lead to correct conclusions there are a number of requirements that must be fulfilled (Coleman and Montgomery, 1993). For example, the response must be measurable and be correlated to the purpose of the experiment. Furthermore, even though not an absolute necessity, the power of statistical operations will be greater if the response is continuous and preferably also normally distributed. The responses of this study are the point estimate for LCC of the track and its related uncertainty, which both are continuous, but not necessarily normally distributed. The following are valid for the present study.

1. The factors that are tested in the experiment are R&M-parameters, which all are continuous and numeric. They are also measurable, controllable, and deemed important for the selected responses.

2. The factors that are not under investigation can easily be held constant, since the study is analytical and not empirical. These factors are the cost factors not directly related to R&M. Hence, no randomization is considered necessary.

Since the study is analytical there are no major economical constraints. Hence, the design is mostly dependent upon the number of R&M parameters that are to be investigated. In order to fulfil the purpose of this study, a two-level factorial design is considered valuable. However, in order to reduce the number of runs, a fractional factorial design is considered sufficient. The analysis is supported by the software tool STATGRAPHICS, which provides suitable tables and graphs for presentation.

The probability distribution of LCC can be found by the use of Monte Carlo simulation. A Monte Carlo simulation is effectively a random number generator that creates values for each R&M parameter. Values are chosen within specified ranges of each parameter and with a frequency proportional to the shape of probability distribution associated with each R&M parameter. The proposed methodology helps in determining the variable costs associated in LCC estimation.

5 Summary of the Appended Papers

This chapter presents a summary of the five appended papers. Each paper makes its own contribution towards the research questions and reports the findings of the case studies. For detailed information, the reader is referred to the appended papers.

Paper I discusses a model for setting the availability target for railway infrastructure. The availability of the infrastructure affects the capacity and the punctuality of the infrastructure. Therefore, setting the availability target should be based on the capacity and punctuality requirements. The objectives of the paper are to develop an approach to i) estimating the capacity of the infrastructure based on the design and operational characteristics and evaluating the influence of the infrastructure availability on the required capacity and ii) estimating the volume of primary and secondary delay due to failures and maintenance of the infrastructure. To achieve these objectives, an example is presented with parameters drawn from failure, maintenance and traffic data. The model was developed using Petri-Nets and Monte Carlo simulations. The simulation results show the effect of the availability of different systems of the infrastructure on the train delays and the infrastructure capacity.

Paper II proposes a model for assessing the safety of the railway track by estimating the probability of derailment. Models for the probability of derailment are developed based on undetected rail breaks and poor track quality using Petri-Nets and Monte Carlo simulations. The paper calculates the probability of undetected rail breaks on the track and the probability of the track quality index falling below the maintenance limit for the track at a given age. The models consider different inspection data and maintenance data as their parameters. The models also calculate the risk based inspection intervals based on the safety requirements of the track. Moreover, the paper discusses the difference in the probabilities between the old and the newly laid tracks of Banverket. It has been shown in the paper how the frequency of track inspections and track quality measurements affect the probabilities. The reduction of these probabilities decreases the risk of derailment. This model will help the infrastructure managers to estimate additional maintenance investment to increase the safety performance of the track to a desired level.

The purpose of **Paper III** is to estimate the availability of the DC track circuit in scheduled and condition based maintenance regimes. The models developed in the paper can estimate the optimum inspection interval for the track circuit subjected to a specific availability requirement. If a system is undergoing a scheduled maintenance regime, the residual life left in the system after the detection of degradation needs to be calculated if we need to estimate the point of failure. The paper discusses a framework for treating the degradation data of the track circuit, based on inspection remarks, to calculate the residual life left in the track circuit. In the case of the condition based maintenance regime, the paper estimates the probability of degradation detection of the condition monitoring device to achieve a certain availability target. The data used in the paper are taken from a specific line section of Banverket.

Paper IV shows that achieving the optimal cost effectiveness is one of the significant ways to address the efficiency of a system, and involves maximising the availability and minimising the life cycle cost of the system over the system life cycle. One of the important ways of maximising the cost effectiveness of the system is to optimise the maintenance policy. This paper demonstrates the estimation of the cost effectiveness of an ERTMS (European Rail Traffic Management System). The degradation and repair process of the system is modelled by Petri-Nets. The model considers systems that experience degradations and are subjected to imperfect maintenance. The results show the effects of the maintenance factor, detectability, inspection interval and deferred maintenance time on the cost effectiveness of the system.

Paper V presents a methodology for estimation of the uncertainty linked with LCC, by a combination of design of experiment (DoE) and Monte Carlo simulation. The proposed methodology is illustrated by a case study of Banverket. LCC is being used as a tool to help in making effective maintenance decisions. However, there are various uncertainties associated with estimation of the LCC. The paper investigates more the uncertainties caused by technical parameters, i.e. reliability and maintainability parameters. The uncertainty in reliability and maintainability parameters exist because of their probabilistic nature, which contributes to the uncertainty in LCC estimation. The simulations are used to make the deterministic LCC equations probabilistic. DoE is applied to provide guidance as to how the R&M parameters should be varied in a systematic way. The paper also illustrates cost models for different maintenance and renewal actions carried out on track.

6 Discussion and Conclusions

In this thesis the research emphasis is placed on the applicability of RAMS and life cycle cost analysis to the development of maintenance decision models for railway infrastructure. The thesis also examines how different maintenance strategies affect the RAMS requirements of the infrastructure. The overall goal is to achieve the RAMS levels stipulated for the railway infrastructure at a lower maintenance cost. The thesis demonstrates approaches and models for achieving a more cost effective maintenance strategy for the railway infrastructure.

Railway infrastructure is a complex system. An important aspect of the rail infrastructure is that the assets have a long useful life. Consequently, once they are installed, it is very difficult to modify the initial design. Therefore, the performance of the infrastructure depends on the maintenance and renewal decisions taken during its life cycle. In many countries, the restructuring of railways and increasing efficiency requirements are causing a changing environment for infrastructure management. The responsibility for parts of the railway system is often handed over to different players. In order to guarantee optimal long-term results for the railway systems, the effects of decisions should be systematically evaluated. The infrastructure manager, who is responsible for the design, construction, maintenance, renewal and upgrading of the infrastructure, has a clearly defined role and is confronted with the increasing performance of its associated partners. Due to increases in the operation and maintenance costs, infrastructure managers are compelled to optimise their budget, while reliability and availability have to be increased without endangering traffic safety. A systematic approach is needed by the infrastructure manager for guaranteeing the defined levels of performance. Since, in the current scenario, most of the maintenance and renewal decisions are based on past experience and expert estimations, a need for an LCC approach arises. A life cycle costing approach combined with RAMS analysis will provide a way to optimise the maintenance strategy, considering the short term budget requirements as well as the long term costs of ownership. As discussed earlier, efficient maintenance decisions always try to achieve the RAMS targets of the system. Appropriate estimation of the RAMS target for the infrastructure is important, as it influences the maintenance strategy as well as the maintenance investment over a period of time. Fig. 6.1 illustrates the RAMS parameters on different hierarchy levels. The RAMS hierarchy has been divided into three levels, i.e. the infrastructure level, system level and component level.

The RAMS parameters indicate the business characteristics of the infrastructure at the infrastructure level, the technical characteristics at the system level (e.g. track system, signalling system, etc.) and the failure characteristics at the component level (e.g. rail, track circuit, etc.). Thus, the availability and safety targets are generally estimated at the infrastructure level, whereas the reliability and maintainability targets are estimated at lower hierarchy levels.

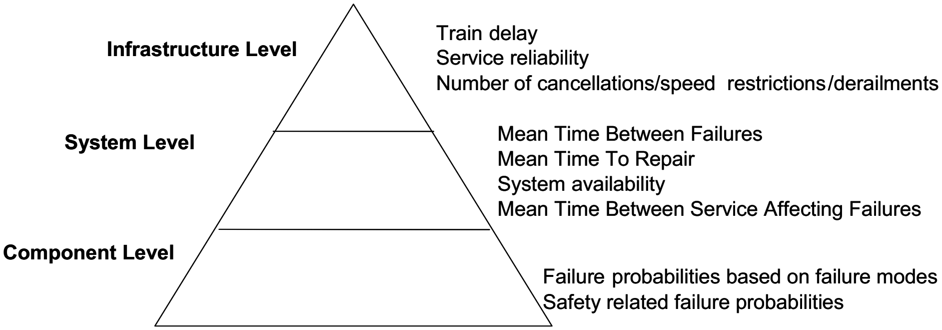


Figure 6.1: RAMS hierarchy within track infrastructure

This necessitates the estimation of the availability and safety targets to achieve the required quality level of the infrastructure. An approach to estimating the availability target for the infrastructure is proposed in Paper I. Paper I deals with the estimation of the availability target of the railway infrastructure to achieve the capacity and punctuality requirements. An example delineating specific design and operating conditions of the infrastructure is presented to estimate the availability target using Petri-Net models. The relationship between availability, capacity and punctuality will help infrastructure managers to identify the availability target for the specific railway infrastructure section when the punctuality and capacity requirements are known. The paper also outlines the operating and design parameters along with the reliability and maintainability parameters that need to be considered when developing this relationship. The results are very dependent on the timetable for the trains in place. A proper setting of the availability target will allow the infrastructure manager to avoid penalties due to delay and lower capacity. In this context, another tool of interest is TrainPlan (see Nilsson, 2006). The system TrainPlan is a tool for planning timetables and available resources and is used for long time timetable construction. The system contains information about each train, such as, train number, description of the train, departure time, arrival time, etc. However, this tool does not calculate reliability or availability of the system explicitly.

In paper I the authors have defined railway capacity under three categories: inherent, achieved and operational capacity. The inherent capacity is the maximum capacity that a railway network can achieve. It is the number of trains that could run over a line or route, during a specific time interval, in a strictly perfect environment, with the trains running permanently and ideally at minimum headway. The inherent capacity is based on the infrastructure design. The achieved capacity is calculated under more realistic assumptions, which are related to the level of expected punctuality. It is the capacity that can permanently be provided under normal operating conditions. It is usually around 60–75% of the inherent capacity (UIC, 2004). Banverket (Swedish National Rail Administration) (Banverket, 2001) indicates a lack of capacity when the capacity utilization is above 80%, as higher capacity utilization leads to more delays of unexpected durations. The achieved capacity is the most significant measure of the track capacity, since it relates the ability of a specific combination of infrastructure, traffic, and operations to move the largest volume within an expected service level. The service level represents the punctuality level for the infrastructure based on a specified time table. If we change the time table, the achieved capacity also changes. The achieved capacity can be defined as the maximum capacity for a specified time table with defined operational headway where as the inherent capacity is the maximum capacity when there is no time table in place and the trains run at minimum (safety) headway. The operational capacity is less than the achieved capacity. This is the case if there is a prolonged shortage of facilities, e.g. due to accidents or weather conditions, but more generally due to failures in the infrastructure which disrupts the train operations.

The safety of the infrastructure is defined at the infrastructure level. Traditionally the safety performance of the railway infrastructure is measured as the number of accidents/million train kilometres or the number of derailments/million kilometres of track. Broadly, performance indicators are classified as leading or lagging indicators. A leading, lead, or prospective indicator is a performance driver. The outcome measure itself is simply the lagging, lag, or retrospective indicator. Leading and lagging indicators can also relate to strategy or goals, and therefore it is important not to mix means and ends. These safety indicators are lagging indicators which only represent the current safety level of the track. If the infrastructure manager wants to improve the safety of the track in the future, it needs to have a lead indicator, e.g. the probability of derailment. Paper II depicts a model for estimating the probability of derailment for the railway track. The model primarily calculates the probability of undetected rail breaks and the probability of the track quality level falling below the maintenance limit. The estimated probabilities are age-dependant i.e. dependant on the tonnage that has passed over the track. The model also considers the influence of the inspection interval on the probabilities. If the infrastructure manager wants to achieve a certain safety performance or target on the aging track at some point of time, this model can compute the amount of investment in inspection that the infrastructure manager needs to make. Paper I and II have answered the first research question.

RAMS analysis can further be applied to the effective maintenance planning of the railway infrastructure to achieve the safety and availability targets. The effect of the inspection interval on the safety level of an aging infrastructure is demonstrated in Paper II. Paper III discusses a model for estimating the availability of the track circuit in two maintenance regimes, i.e. scheduled maintenance and time based maintenance. The model provides a comparison of system availability between the two maintenance regimes. A reduction of the unavailability caused by track circuit failures must be addressed using an assured methodology. This methodology usually requires either an optimisation or a complete overhaul of the maintenance processes in place. This starts with understanding the causes of track circuit failures, which can either be within the track circuit itself, e.g. component or systemic failure, or result from a failure of part of the infrastructure, such as poor track or related components. If a system is undergoing a scheduled maintenance regime, the residual life left in the system after the detection of degradation needs to be calculated if we need to estimate the point of failure. A framework for treating the degradation and failure data of the track circuits is presented in Paper IV. The paper demonstrates the effect of the inspection interval on the system availability in the case of the scheduled maintenance regime. If an infrastructure manager is following a scheduled maintenance regime, this model will be helpful in estimating the optimal inspection interval for achieving the specified availability target described in Paper I. A similar model has been developed in Paper IV to show the effect of the detection probability of the condition monitoring device on the system availability. Paper II and III have answered the second research question.

As described in Figure 2.7, different maintenance strategies can achieve the RAMS targets specified for the infrastructure. However, the cost effective maintenance strategy is the one that provides the lowest life cycle cost. The application of both RAMS and LCC analyses to realize a cost effective maintenance policy is depicted in Paper V in the form of a case study. The third research question is answered in Paper IV. A Petri-Net model is developed to estimate the cost effectiveness of the system. The model considers the different degradation and maintenance states of the system. An imperfect maintenance scenario is considered in this model. Kijima models are used to estimate the virtual age of the system in the case of an imperfect maintenance. Sensitivity analysis has been performed on different maintenance parameters to maximise the cost effectiveness of the system. The model focuses on achieving higher availability targets at a lower life cycle cost. The estimation of the life cycle cost should always consider the uncertainties associated with it.

An approach to estimating the uncertainties connected with the life cycle cost is described in Paper V. The paper deals with two different levels of uncertainty associated with the LCC of the track infrastructure. The level I uncertainty concerns costs due to penalties imposed by traffic operators on the infrastructure manager due to such factors as train delay, traffic disruption, or derailment. These anomalies can be caused by planned or unplanned maintenance actions, as well as a lack of necessary maintenance. Hence, the resulting costs are related to decisions about maintenance actions and can be estimated by the probabilistic assessment of train delay, derailment, or traffic disruptions, considering the technical and operational characteristics of the track, as well as the maintenance actions. The level I uncertainty can also be viewed as belonging to the external risk of the LCC analysis, where the costs should be included to make the LCC analysis more effective. However, there is also the level II uncertainty, which is the internal risk associated with LCC. The level II uncertainty pertains to the variable contribution to the total LCC originating from the uncertainty in the reliability and maintainability parameters. For better estimation of uncertainty in LCC, this paper outlines a methodology based on a combination of Monte Carlo simulation and DoE. This combination provides the possibility of identifying parameters that are influential on the LCC estimation and its variability. The uncertainty in the LCC is presented as variable costs with associated distributions. Paper V has answered the fourth research question. When the variable costs are added to the LCC, it becomes more robust. Hence, the decision-makers are helped in making more effective decisions about maintenance policy by considering the LCC. The approach and models developed by this research can be adapted to other systems in the railway infrastructure with suitable modifications.

6.1 Research Contributions

The research presented in this doctoral thesis has focussed on the applicability of RAMS and LCC analysis to the development of maintenance policies for the railway infrastructure. The literature study shows that the use of RAMS and LCC methodologies for railway infrastrcuture is still in its initial phase. Furthermore, it has been found that these methodologies have a limited use in developing a maintenance policy for the infrastructure.

The research contributions can be listed as follows:

1. The development of an approach to estimate the availability target for the railway infrastructure considering capacity and punctuality requirements (Paper I)

2. The development of a model to assess the safety level of the infrastructure and estimate the maintenance investments based on the safety target (Paper II)

3. The development of models to estimate the availability of DC track circuit subjected to different maintenance policies and predict the residual life of the system (Paper III)

4. The development of a model to estimate the cost effective maintenance policy for a signalling system subjected to degradation and imperfect maintenance (Paper IV)

5. The development of an approach to determine the uncertainty in LCC estimations (Paper V)

6.2 Scope of the Future Research

In summary, based on the research conducted, the following areas are suitable for further research:

- The development of a more robust model for railway infrastructure addressing the interactive effects of the different sub-systems in the infrastructure. On the infrastructure level, the availability and safety are dependent upon the availability and safety of the different sub-systems of the infrastructure. The degradations and maintenance of one sub-system affect the RAMS characteristics of the other sub-systems. In order to achieve the RAMS target at the infrastructure level, it is necessary to investigate the interactive effects of the different sub-systems.

- The implementation of the models on the different systems of the railway infrastructure. The thesis presents different approaches and models for the maintenance decision support of the railway infrastructure. These models can be implemented on the different systems of the infrastructure to develop maintenance policies.

References

Andersson, M. (2002). *Strategic Planning of Track Maintenance - State of the Art,* Research Report, Division of Urban Studies, Department of Infrastructure, Royal Institute of Technology (KTH), ISSN: 1651-0216.

Ascher, H. and Feingold, H. (1984). *Repairable system reliability: Modelling, Inference, Misconceptions and Their Causes,* Marcel Dekker Inc., New York, ISBN 0-8247-7276-8.

Asiedu, Y. and Gu, P. (1998). Product life cycle cost analysis: state of the art review, *International Journal of Production Research*, 36(4), 883-908.

Blanchard, B.S. and Fabrycky, W.J. (1998). *Systems Engineering and Analysis,* 3rd ed., Upper Saddle River, NJ: Prentice-Hall.

Blischke, W.R. and Murthy, D.N.P. (2003). *Case Studies in Reliability and Maintenance*, John Wiley & Sons, USA.

Burström, B., Ericsson, G. and Kjellsson, U. (1994). Verification of Life-Cycle Cost and reliability for the Swedish high speed train X2000, *In Proceedings of Annual Reliability and Maintainability Symposium(RAMS)*, Los Angeles, 24-27 January, 166-171

Banverket, (2001). *Calculation manual: Aid for cost-benefit analysis in the railway sector*, Handbook BVH 706.00, Banverket, Borlänge, (in Swedish).

Cannon, D.F., Edel, K.O., Grassie, S.L. and Sawley, K. (2003). Rail defects: an overview, *Journal of Fatigue & Fracture of Engineering Materials & Structures*, 26, 865-886.

Coleman, D.E. and Montgomery, D.C. (1993). A Systematic Approach to Planning for a Designed Industrial Experiment, *Techometrics*, 35(1), 1-27.

Cooper, D.R. and Schindler, P.S. (2006). *Business Research Methods,* 9th ed., McGraw-Hill Companies, Inc., Singapore, ISBN 0-07-124430-1.

CRMA (1998). *Cost, Reliability, Maintenance, and Availability,* Final Summary Report
Available Online: http://cordis.europa.eu/transport/src/crmarep.htm#1
Accessed on: 20/10/2007.

Dane, F.C. (1990). *Research method*, Brooks/Cole Publishing Company, California.

Diamond, S. and Wolf, E. (2002). *Transportation for the 21st century,* TracGlide Top-of-Rail Lubrication System, Report from Department of Energy, USA.

EN 50126 (1999) *The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) for Railway Applications,* European Committee for Electrotechnical Standardization (CENELEC), Brussels, Belgium.

Esveld, C. (2001). *Modern railway track,* 2nd ed., Delft: MRT-Productions, Germany.

Fazio, A.E. and Prybella, R. (1980). *Development of an Analytical Approach to Track Maintenance Planning,* Transportation Research Record 744, National Research Council, National Academy Press, Washington, D.C., USA, 46-52.

Ferreira L. (1997). Rail track infrastructure ownership: investment and operational issues, *Transportation*, 24(2), 183-200.

IEC 62278 (2002). *Railway applications - specification and demonstration of reliability, availability, maintainability and safety (RAMS*), International Electrotechnical Commission, Geneva.

IEV 191-01-07 (2007). *Dependability and Quality of Service/Item Related Performance,*
International Electrotechnical Vocabulary, Online Database.
Available online: http://std.iec.ch/iec60050
Accessed on: 25/10/2007.

IMPROVERAIL (2003). *Improved Tools for Railway Capacity and Access Management*
Available Online: http://ec.europa.eu/transport/rail/research/improverail_en.htm
Accessed on: 20/10/2007.

INNOTRACK (2009). *Innovative Track Systems*
Available Online: http://www.innotrack.eu/
Accessed on: 16/06/2009.

Jovanovic, S. (2006). Railway Track Quality Assessment and Related Decision Making*, In Proceedings of AREMA annual conferences*, Kentucky, 17-20 September.

Kalousek, J., Sroba, P. and Hegelund, C. (1989). Analysis of rail grinding tests and implications for corrective and preventative grinding, *In Proceedings of the 4th International Heavy Haul Railway Conference*, Brisbane, 11-15 September, 193-204.

Karlsson, V. (2005). An overall view of maintenance, *European Railway Review*, 11(3), 11-17.

Kijima, M. (1989). Some results for repairable systems with general repair, *Journal of Applied Probability,* 26, 89-102.

Kumar, S. (2008). *A Study of the Rail Degradation Process to Predict Rail Breaks,* PhD Thesis, Luleå University of Technology, Luleå, Sweden, ISSN: 1402-1544; 2008:15.

Kumar, U. and Akersten, P A. (2008). Availability and maintainability, *In Encyclopedia of Quantitative Risk Analysis and Assessment*, ed. / E. Melnik and B. Everitt, Wiley, 77-84.

Kumar, U. (1989). *Maintenance Strategy for Mechanized Mining Systems,* Licentiate Thesis, Luleå University of Technology, Luleå, Sweden, ISSN: 0280-8242; 1989:01L.

Larsson, D. (2002). *Verification of the DeCoTrack: a prediction model for degradation cost of track*, Research Report, Luleå University of Technology, Luleå, Sweden, ISSN: 1402-1528; 2002:18.

LibeRTiN (2005). *Light rail thematic network*
Available online: http://www.libertin.info/
Accessed on: 20/10/2007.

LICB (2007). *Lasting infrastructure cost benchmarking*
Available online: http://www.uic.asso.fr/infra/article.php3?id_article=13
Accessed on: 25/10/2007.

Madu, C.N. (2005). Strategic value of reliability and maintainability management, *International Journal of Quality and Reliability Management*, 22 (3), 317-328.

Magel, E. and Sroba, P. (2007). Workshop on Rail Grinding Practices, International Heavy Haul Association (IHHA), Kiruna, Sweden, 8-10 June.

Markeset, T. and Kumar, U. (2004). Dimensioning of Product Support: Issues, Challenges, and Opportunities, *In Proceedings of Annual Reliability and Maintainability Symposium (RAMS)*, Los Angeles, 26-29 January, 565- 570

Markeset, T. and Kumar, U. (2003). Integration of RAMS and risk analysis in product design and development work processes: a case study, *Journal of Quality in Maintenance Engineering*, 9(4), 393-410.

MODTRAIN (2007). *Innovative Modular Vehicle Concepts for an Integrated European Railway System*
Available Online: http://www.modtrain.com/
Accessed on: 25/10/2007.

Neuman, W.L. (2003). *Social Research Method,* 5th ed., USA.

Nilsson, K. (2006). *Alternativa system för tågstyrning på Malmbanan,* Forskningsrapport, Avdelningen för industriell logistik, Luleå Tekniska Universitet, ISSN:1402-1528; 2006:01, (In Swedish).

Petri, C. (1962). *Kommunikation mit Automaten*, PhD thesis, University of Bonn, Germany, (In German).

ProM@in (2003). *Progress in Maintenance and Management of Infrastructure*
Available Online: http://www.promain.org/
Accessed on: 15/08/2007.

Putallaz, Y. (2003). Strategic M&R policy of a railway corridor taking into account the value of capacity, *In Proceedings of World Congress on Railway Research*, Edinburgh, 28 September-1 October.

Reliasoft Corporation (2003). Weibull++6 software, version 6.0.11, web: www.weibull.com

REMAIN (1998). *Modular System for Reliability and Maintainability Management in European Rail Transport*
Available Online: http://cordis.europa.eu/transport/src/remainrep.htm
Accessed on: 15/10/2007.

Ringsberg, J.W. (2001). Life prediction of rolling contact fatigue crack initiation*,* *International Journal of Fatigue*, 23, 575-586.

Stalder, O. (2001). The life cycle costs (LCC) of entire rail networks: an international comparison*,* *Rail International*, 32 (4), 26-31

Sullivan, T.J. (2001). *Methods of Social Research,* Harcourt, Inc., USA, ISBN: 0-15-507463-6.

Sumser, J.R. (2000). A *Guide to Empirical Research in Communication,* Sage Publications Inc, ISBN: 0761922229.

Swier, J. (2004). *Exposing the Infrastructure Cost Drivers*, Railway Gazette International, Sutton, Surrey, UK.
Swier, J. and Luiten, T. (2003). *Outsourcing in The Netherlands; the sector process model,* Innovations for a Cost Effective Railway Track, ProM@in, 3, 20-26.

UIC (2004). *Capacity*, UIC Leaflet 406, International Union of Railways, Paris.

UNIFE (2001). *Life Cycle Cost Working Group*
Available Online: http://www.unife.org/workgroups/lcc_2.htm
Accessed on: 21/10/2007.

URBAN TRACK (2007). *Urban Rail Infrastructure in a harmonised Europe*
Available Online: http://www.urbantrack.eu/
Accessed on: 20/10/2007.

Vatn, J. (2002). *A Life Cycle Cost model for prioritisation of track maintenance and renewal,* Innovations for a Cost Effective Railway Track, Progress in Maintenance and Management of Infrastructure (ProM@in), 2, 21-25.

Wilson, A. (1999). *Asset Maintenance Management*, Alden Press, Oxford, UK.

Veit P. (2003). *Some thoughts concerning outsourcing of track maintenance,* Innovations for a Cost Effective Railway Track, ProM@in, 3, 14-19.

Zarembski, A.M. (1998). *Development and Implementation of Integrated Maintenance Planning Systems,* Transportation Research Board Annual Meeting, Washington D.C. U.S.

Zhao, J., Chan, A.H.C., Roberts, C. and Stirling, A.B. (2006). Assessing the Economic Life of Rail Using a Stochastic Analysis of Failures*, In Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 220 (2), 103-111.

Zoeteman, A. (2006). Asset maintenance management: state of the art in the European railways*, International Journal of Critical Infrastructures*, 2(2/3), 171-186.

Zoeteman, A. (2001). Life cycle cost analysis for managing rail infrastructure, concept of a decision support system for railway design and maintenance*, European Journal of Transport and Infrastructure Research*, 1(4), 391 – 413.

Zoeteman, A. and Esveld, C. (1999). Evaluating Track Structures: Life Cycle Cost Analysis as Structured Approach*, In Proceedings of World Congress on Railway Research*, Tokyo, 19-23 October.

APPENDED PAPERS

Paper I

Availability target of the railway infrastructure: an analysis

Patra, A. P., Kumar, U. and Larsson-Kråik, P-O. (2010). Availability target of the railway infrastructure: an analysis. *Accepted for publication in Proceedings of Reliability and Maintainability Symposium (RAMS)*, San Jose, USA, 25-28 January.

Availability target of the railway infrastructure: an analysis

Ambika P. Patra, MTech, Luleå University of Technology
Uday Kumar, PhD, Luleå University of Technology
Per-Olof Larsson Kråik, PhD, Swedish National Rail Administration (Banverket)

Key Words: Railway Infrastructure, Reliability, Maintainability, Availability, Capacity, Punctuality

SUMMARY & CONCLUSIONS

The railway has been accepted as one of the most environmentally friendly modes of transport for goods and passengers. However, the railway sector is striving to increase its capacity to meet the growing demand for the transport of goods and passengers with a high level of punctuality in its services. Higher availability requires the effective operation and maintenance of infrastructure, often necessitating the implementation of cost-effective preventive maintenance strategies. Therefore, a higher availability target means higher maintenance investment. However, the question of setting the availability target for the infrastructure is not easy, as it involves many influencing decision parameters, apart from a good understanding of the network configuration and traffic density. Railway networks that have a smaller number of trains and a low punctuality requirement do not require higher availability targets. The aim of this paper is to estimate the availability target for railway infrastructure based on the capacity and punctuality requirements of infrastructure managers and train operating companies. The objectives of the paper are to develop an approach to i) estimating the capacity of the infrastructure based on the design and operational characteristics and evaluating the influence of infrastructure availability on the required capacity and ii) estimating the volume of primary and secondary delay due to failures and maintenance of the infrastructure and establishing the relationship between availability and punctuality requirements. To achieve these objectives, an example is presented with parameters drawn from failure, maintenance and traffic data. Finally, a model has been developed in Petri-Nets to establish a relationship between availability, capacity and punctuality. Monte Carlo simulation is used to establish the relationship. The simulation results illustrate the effect of infrastructure availability on train delays and capacity.

1 INTRODUCTION

Rail traffic is the most important form of public traffic in Europe as the density of the railway network is very high compared to the other parts of the globe. To be in competition with other modes of transportation, railway traffic must be quick, comfortable, cheap and primarily safe. There have been contractual agreements concerning the targeted level of reliability and punctuality in the performance regime within the railway sector. The business needs of railway infrastructure can be defined as lower ownership costs, interoperability, enhanced safety, improved punctuality, increased capacity and reduced journey times. The availability of railway infrastructure plays a significant role in attaining a higher capacity and punctuality level of the infrastructure. The required level of availability determines the amount of maintenance investment in the infrastructure over a period of time. A higher availability target requires the effective operation and maintenance of infrastructure, often necessitating the implementation of cost-effective preventive maintenance strategies in combination with effective supply chain management. However, the question of setting the availability target for the infrastructure is not easy, as it involves many decision parameters, apart from a good understanding of the network configuration and traffic density. Railway networks that have a smaller number of trains and a low punctuality requirement do not require higher availability targets. In order to estimate the availability target, it is necessary to understand the capacity and punctuality requirements of the railway infrastructure. Section 2 illustrates the fundamentals of capacity and punctuality. A model has been developed using Petri-Nets to establish a three-way relationship between availability, capacity and punctuality. Section 3 describes the model with an example. Finally discussions and conclusions are described in Section 4.

2 RAILWAY CAPACITY AND PUNCTUALITY

An efficient utilization of the existing railway infrastructure is an essential component of a high-quality transportation system and has become a central task for railway infrastructure managers. Line capacity is, in essence, what the infrastructure managers have to sell as their final product. Although capacity seems to be a self-explanatory term in common language, its scientific use may lead to substantial difficulties when it is associated with objective and quantifiable measures. It is a complex term that has numerous meanings and for which numerous definitions have been given. In [1] it is stated that capacity as such does not exist. Railway infrastructure capacity depends on the way in which it is utilized. However, in [2] it is stated that capacity is a measure of the ability to move a specific amount of traffic over a defined rail line with a given set of resources under a specific service plan. As illustrated in Figure 1, capacity is a balanced mix of the number of trains, the stability of the timetable, the average speed achieved and the heterogeneity (mixed traffic with different train speeds) of the train system. It is, for instance, possible to achieve a high average speed on a railway network by having a high heterogeneity – a mix of fast and slower trains. However, the cost of maintaining a high average speed with a high heterogeneity makes it difficult to run a great number of trains with a high stability (punctuality) than if all the trains ran at the same speed. If one wants to run more trains, it is necessary to operate with less mixed traffic and thereby have a lower average speed, as in the case of metro systems.

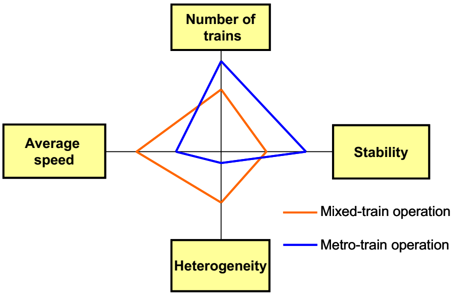


Figure 1 – Capacity balance [1]

Railway capacity has different values based on different criteria. In this paper the authors have defined railway capacity under three categories: inherent, achieved and operational capacity. The inherent capacity is the maximum capacity that a railway network can achieve. It is the number of trains that could run over a line or route, during a specific time interval, in a strictly perfect environment, with the trains running permanently and ideally at minimum headway. The inherent capacity is based on the infrastructure design. The achieved capacity is calculated under more realistic assumptions, which are related to the level of expected punctuality. It is the capacity that can permanently be provided under normal operating conditions. It is usually around 60–75% of the inherent capacity [1]. Banverket (Swedish National Rail Administration) [3] indicates a lack of capacity when the capacity utilization is above 80%, as higher capacity utilization leads to more delays of unexpected durations. The achieved capacity is the most significant measure of the track capacity, since it relates the ability of a specific combination of infrastructure, traffic, and operations to move the largest volume within an expected service level. The service level represents the punctuality level for the infrastructure based on a specified time table. If we change the time table, the achieved capacity also changes. The achieved capacity can be defined as the maximum capacity for a specified time table with defined operational headway where as the inherent capacity is the maximum capacity when there is no time table in place and the trains run at minimum (safety) headway. The operational capacity is less than the achieved capacity. This is the case if there is a prolonged shortage of facilities, e.g. due to accidents or weather conditions, but more generally due to failures in the infrastructure which disrupts the train operations.

Railway capacity very much depends on the headway time between the trains, i.e. both the safety headway and the operational headway time. Figure 2 illustrates the safety headway time between the trains. It is dependant on the distance that the trains maintain for safe operations on the track. As shown in the figure, the safety headway time is the summation of the travel time, braking time, release time and operating time. The travel time is the time required to cover the distance between two signals. The area between two signals is called a block and is controlled by a track circuit.

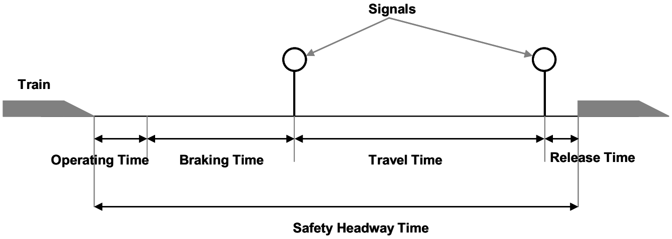


Figure 2 – Safety headway diagram

At any given time only one train can occupy a block section. The travel time depends upon the distance between the signals and the speed of the trains. The braking time depends on the braking distance i.e. the distance required to stop before a signal. It is calculated by considering the train speed and deceleration. The release time is the time required for the entire length of train to cross the signal. This depends on the length of the train and the speed of the train. The operating time is a safety time and is fixed by the infrastructure managers. The inherent capacity of a double track line depends on the safety headway between the trains. For example, if the safety headway is five minutes, the inherent capacity per track section will be 12 trains per hour. However, as discussed earlier infrastructure managers consider buffer time to accommodate delays. Moreover, the requirements of stakeholders set the operational headway time between the trains more than the safety headway. The operational headway is the actual time between two consecutive trains as per the train timetable.

Punctuality is defined differently by different infrastructure managers across the globe. A train in Sweden is considered punctual if it is less than five minutes off schedule at a station, otherwise it is delayed. Train delays may be classified into two major categories: primary delays and secondary delays. Primary delays are the delays undergone by the trains passing over a disturbed track section. Primary delays are the time differences between the normal and the disrupted journey. Secondary delays are delays of follower trains, which will not undergo the totality of primary delay, but which will undergo a delay because the previous train is delayed. This kind of delay happens when a failure is close to being restored. The principles of primary and secondary delay are illustrated in Section 3.

3 AVAILABILITY TARGET ESTIMATION

Railway infrastructure consists of various sub-systems, such as the track system, the signalling and telecommunication system, and the power system. Each of these sub-systems contributes to the infrastructure availability. As discussed earlier, the railway infrastructure availability influences the capacity and punctuality of train operations. Therefore, when estimating the availability target, the capacity and punctuality requirements of the infrastructure must be considered. Different failure modes in the railway infrastructure induce different amounts of delay in railway network based on speed restrictions. The amount of delay depends on the occurrence rates and repair times of the failure modes. Figure 3 depicts the speed profile of trains due to a track circuit failure. A failure in a track circuit turns the signal red for that particular block and trains pass at a reduced speed over that block until the failure is rectified. This section develops a model which estimates the availability target. An example is provided to illustrate the model.

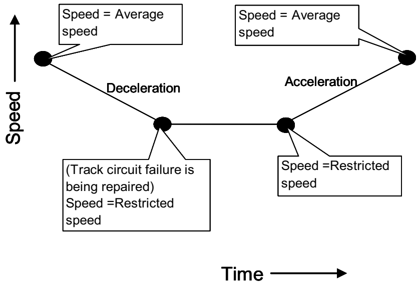


Figure 3 – Speed profile of trains due to failure of a track circuit

3.1 Example

Let us consider a double track railway line between two main stations. This line has multiple intermediate stations. Trains run with a uniform operational headway (OH) of 15 minutes and a safety headway (SH) of 5 minutes. All the trains that run on the track have the same speed pattern. As discussed earlier, the capacity and punctuality will be estimated for the line section between two adjacent stations. We consider the failures of three sub-systems that occur in this specific line section. The reliability and maintainability details of these sub-systems are given in Table 1. Occurrences of failures induce primary and secondary delay in the railway network as illustrated in Figure 4 & 5. When a failure occurs, trains reduce their speeds over the affected area (see Figure 3) and arrive late compared to their specified arrival time. The time difference determines the primary delay (PD). It is calculated by the kinematics equations of motion considering distance, acceleration, deceleration and speed. The number of trains that will be disrupted by primary delay is given by:

N_{PD} = Mean Down Time (MDT)/Operational headway (OH)

N_{PD} is an integer.

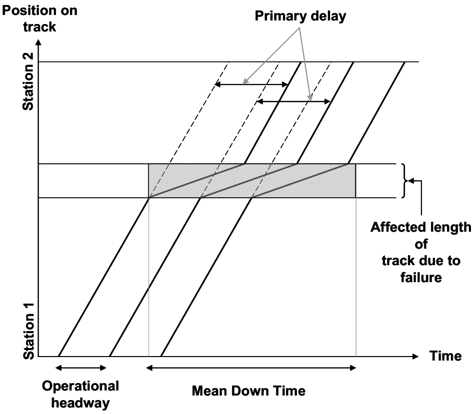


Figure 4 – Illustration of primary delay

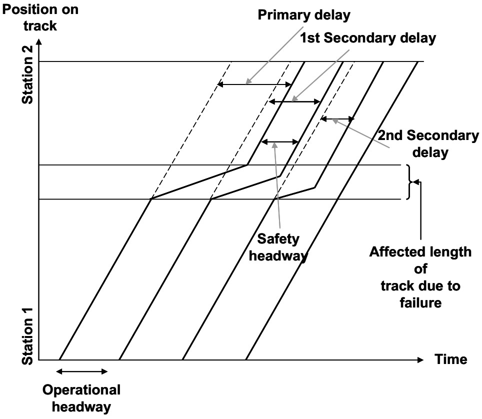


Figure 5 – Illustration of secondary delay

Secondary delay will occur if primary delay > (Operational headway – Safety headway). In this case the following train will have to slow down to keep the minimal distance (SH) from the last primarily delayed train. The secondary delay undergone by the 1st following train is given by:

$$SD_1 = PD - (OH - SH)$$

If $SD_1 > (OH - SH)$, the 2nd following train will be subjected to secondary delay. The secondary delay undergone by the 2nd following train is given by:

$$SD_2 = SD_1 - (OH - SH) = PD - 2*(OH - SH)$$

Secondary delays can be more generally expressed as:

$$SD_i = PD - i*(OH - SH)$$

The number of trains that will be disrupted by secondary delay is given by:

S_{PD} = Primary delay/(Operational headway – Safety headway)

S_{PD} is an integer.

3.2 Petri-Net model for studying the relationship between availability, capacity and punctuality

The Petri-Net model (see Figure 6) has been developed for estimating the relationship between availability, capacity and punctuality on the line between two stations for the example described above. The reliability and maintainability parameters for the model are given in Table 1. As illustrated in Figure 6, places 1, 3 and 5 denote the working states of sub-systems 1, 2 and 3 respectively where as places 2, 4 and 6 denote the failed states. The transitions between these places operate according to the failure rates and mean down time of the sub-systems. The failure rates in this case are assumed to be following exponential distributions. Whenever any sub-system fails, the infrastructure goes to a failed state and it is restored to the working state depending on the mean down time of the sub-system. This is illustrated by places 12, 13 and 14. The firings of the transitions between these places occur at any time depending on the failures of the sub-systems. It needs to be mentioned that in this model we have assumed only three sub-system failures that affect the capacity and punctuality of the railway network. In other cases the number of sub-systems can be more dependent on the specific railway track design criteria.

Sub-system	Failure Rate-FR (per minute)	Mean Down Time-MDT (in minutes)	Primary Delay-PD (in minutes)
Sub-system 1	6.00E-05	75	12
Sub-system 2	1.00E-04	60	10
Sub-system 3	1.50E-04	45	8

Table 1 – Reliability and Maintainability data for sub-systems

Places 7 to 9 describe the movement of trains between two stations. The transitions between these places consider the fundamentals of primary and secondary delay illustrated in Section 3. Trains start at station 1 (place 7) and reach station 2 (place 9). Place 9 keeps account of all the primary and secondary delay that the trains undergo in the case of sub-systems failures. Places 10 and 11 calculate the capacity of the track section between station 1 and 2 depending on the arrival time of the trains at place 9. Place 10 calculates the number of trains that reach station 2 every hour. In this model we calculate the average capacity (trains/hour) of the track section over a period of time.

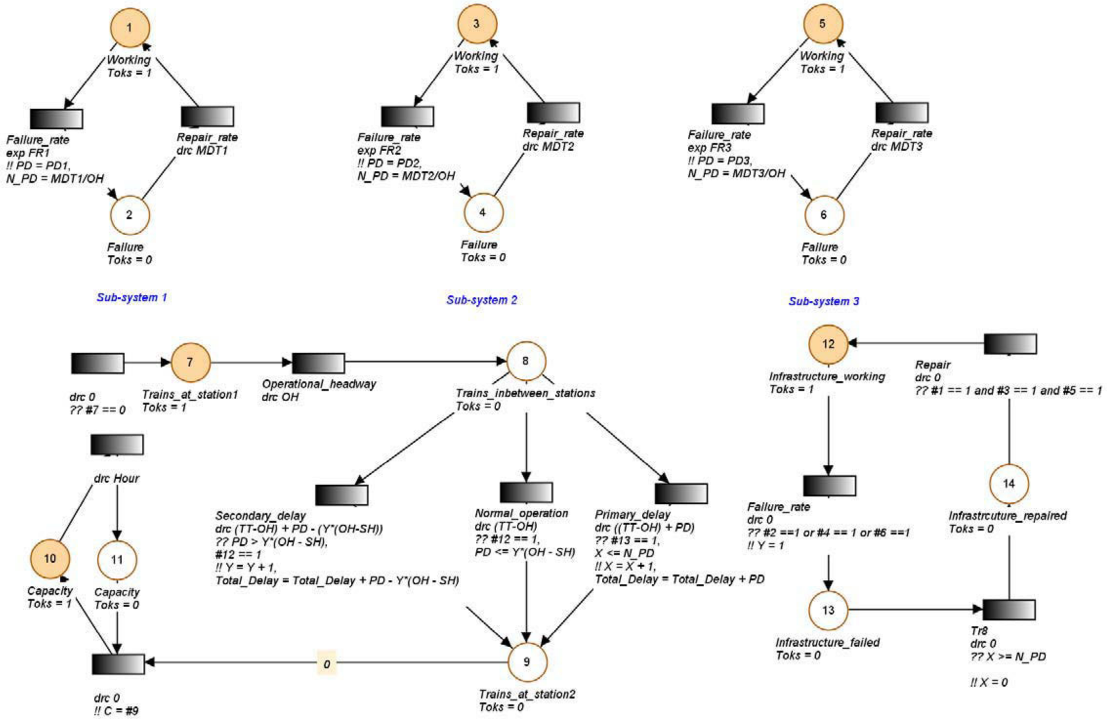


Figure 6 – Petri-Net model

The model also calculates the cumulative delay that the trains can undergo over a period of time on that track section. It is also considered that the trains take 25 minutes to travel between the stations i.e. the travel time (TT) is 25 minutes. The model will enable us to estimate the effects of the track system availability on the capacity and punctuality of that particular track section. The fundamentals of Petri-Net modelling can be found in [5].

Trains are operated for 18 hours a day and preventive maintenance is carried out during the rest six hours. Therefore, the availability of infrastructure considers only corrective maintenance. By performing Monte Carlo simulations on the Petri-Net models, we obtained the capacity variations over a period of one month (18x60 hours). Figure 7 illustrates the estimated operational capacity over a period of one month.

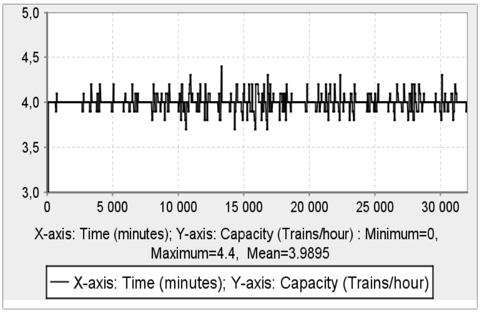


Figure 7 –Operational capacity over a period of one month

As shown in Figure 7, when a failure occurs, the train operation is disrupted until the failure is corrected and hence the capacity decreases. However, after the failure is corrected the operational capacity of the track increases because the delayed trains arrive in the same hour as the non-delayed trains. The achieved capacity in normal operation is expected to be 4 trains/hour as the operational headway is 15 minutes. However, due to failures, the mean operational capacity over a period of one month is 3.9895 trains/hour, which almost satisfies the achieved capacity requirement. Similarly, the cumulative delay that the trains will undergo over a period of 1 month is illustrated in Figure 8. The total delay occurring in one month is 355.6 minutes, which comprises of both primary delay and secondary delay. The average availability over that period is estimated to be 0.9826. In order to estimate the target availability, we need to perform sensitivity analysis on capacity and punctuality by changing the values of availability; i.e. changing the values of reliability and maintainability.

A relationship between the capacity, punctuality and availability is illustrated in Figure 9. As shown in the figure, the delay decreases with increases in availability value. However, in this particular case, the capacity remains constant with the change in availability values. This is due to fact that the timetable specified in the example absorbs all the delays because of the buffer embedded into it.

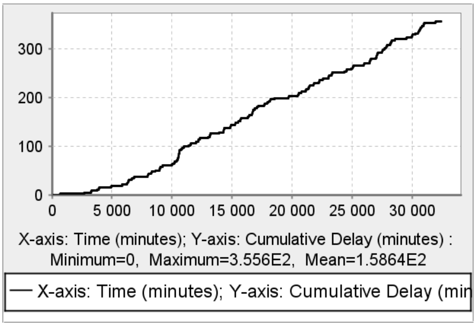


Figure 8 –Cumulative delay over a period of one month

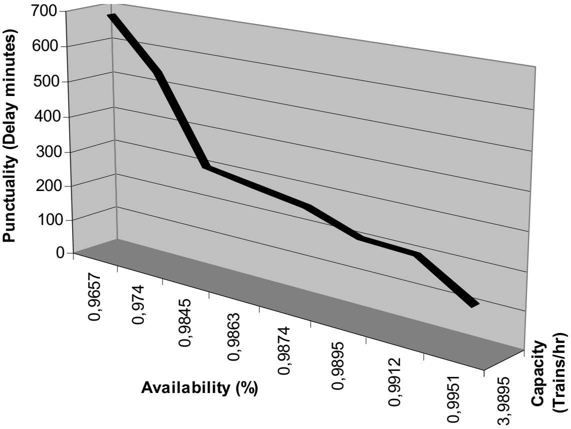


Figure 9 – Relation between capacity and punctuality with availability

In this particular example the operational capacity is almost equal to the achieved capacity for all values of availability. There might be a case when we reduce the availability to a much lower level then operational capacity will fall below than the required capacity. However, in this case we can estimate the availability target for the line section e.g. if the delay requirement should be less than 250 minutes, then our availability target should be 0.988. We estimated this availability target for a line section between two adjacent stations. In a railway network there can be many stations. The highest availability target that we estimate for any section will determine the availability target for the whole railway network. When we determine the availability target for a specific railway network, the infrastructure manager can estimate the maintenance investment in that particular network over a period of time to achieve that availability target.

4 DISCUSSIONS

The increasing complexity of modern technical systems has resulted in high reliability, maintainability and availability requirements. These requirement need to be met by the system owner to remain competitive. Setting these requirements is difficult when the systems have many stakeholders. Railway infrastructure is one of these systems. The system availability of the railway infrastructure directly affects the punctuality and capacity of the railway network. Failure to meet the capacity and punctuality requirements of the railway network incurs a penalty for the infrastructure manager who owns the system. Therefore, the infrastructure manager needs to estimate the availability target that it intends to achieve so as to meet the capacity and punctuality requirements. The availability target also enables the infrastructure manger to estimate the maintenance investment over a period of time. In this paper we have developed a model for estimating the availability target of the railway infrastructure in Petri-Net. The model considers the design, operation and failures of the infrastructure to derive a relationship between availability, capacity and punctuality. The relation is derived by means of an example presented in this paper. When the capacity and punctuality requirements are known, the infrastructure manager can estimate the availability target for this particular scenario explained in the example. However, the results show that the operational capacity does not change with the change in the system availability. This is due to the fact that the operational headway is kept large enough in the train timetable to absorb the delays, to keep the operational capacity of the network close to the achieved capacity. In general, if the operational availability changes, the operational capacity has to change. It actually does change for the hour during which a failure occurs (see Fig. 7); but during the next hour, the operational capacity increases as the delayed trains arrive along with the scheduled trains and we estimate the operational capacity as the average of all the hours. However, if we lower the operational headway, we can observe that the operational capacity will change with the change in system availability. This model will help the infrastructure manager to estimate the availability target of the railway infrastructure based on the capacity and punctuality requirements. A proper setting of the availability target will allow the infrastructure manager to avoid penalties due to delay and lower capacity. This will also help to estimate the total maintenance investment that the infrastructure manager needs to make over a period of time.

REFERENCES

1. UIC "Capacity", UIC Leaflet 406, International Union of Railways, Paris, 2004
2. M. Abril, F. Barber, L. Ingolotti, M. A. Salido, P. Tormos, and A. Lova, "An assessment of railway capacity", *Transport Research Part E: Logistics an Transportation Review*, (Sept) 2008, pp. 774-806
3. Banverket, "Calculation manual: Aid for cost-benefit analysis in the railway sector", Handbook BVH 706.00, Banverket, Borlänge, 2001 (in Swedish)
4. B. Nyström, "Use of Availability Concepts in the Railway System", *International Journal of Performability Engineering*, Vol. 5(2), 2009, pp. 103-118
5. T. Murata, "Petri Nets: Properties, Analysis and Applications", *Proceedings of the IEEE*, Vol. 77(4), 1989, pp. 541 – 580.

BIOGRAPHIES

Ambika Prasad Patra
Luleå Railway Research Center
Division of Operation and Maintenance Engineering
Luleå University of Technology
SE-97187 Luleå, Sweden

e-mail: ambika.patra@ltu.se

Ambika Prasad Patra is a PhD candidate at the Division of Operation and Maintenance Engineering, Luleå University of Technology. His research topics are RAMS and LCC analyses for railway systems. He received his MTech degree in 2002 in Reliability Engineering from the Indian Institute of Technology, Bombay, India.

Uday Kumar, PhD
Luleå Railway Research Center
Division of Operation and Maintenance Engineering
Luleå University of Technology
SE-97187 Luleå, Sweden

e-mail: uday.kumar@ltu.se

Dr Uday Kumar is Professor of Operation and Maintenance Engineering at Luleå University of Technology, Luleå, Sweden. He also is the Director of Luleå Railway Research Centre (www.jvtc.ltu.se) at the same university. His research interests are maintenance engineering and management, reliability and maintainability analysis, product support, life cycle costing, risk analysis, system analysis, etc. He is also a member of the editorial board and a reviewer for many international journals. He has published more than 150 papers in international journals and conference proceedings.

Per-Olof Larsson Kråik, PhD
Swedish National Rail Administration,
SE-97102, Luleå, Sweden

e-mail: per-olof.larsson-kraik@banverket.se

Dr Per-Olof Larsson-Kråik co-ordinates research and development activities for Banverket (the Swedish National Rail Administration). Dr Larsson-Kråik has written over 50 articles and case studies in the areas of tribology, reliability, maintainability, logistics, and supportability. He has carried out several research and consultancy projects for several organizations in Europe and Australia. Dr Larsson-Kråik's research interests include systems engineering, rail wheel (w/r) interaction, w/r lubrication and strategic operation and maintenance decision making in an organisation.

Paper II

Assessment and improvement of railway track safety

Patra, A. P., Kumar, U. and Larsson-Kråik, P-O. (2009). Assessment and improvement of railway track safety. *Proceedings of 9th International Heavy Haul Conference (IHHA)*, Shanghai, China, 22-24 June.

Assessment and Improvement of Railway Track Safety

Ambika Prasad Patra[1*], Uday Kumar[1], P-O Larsson-Kråik[2]
[1] Luleå Railway Research Center, SE-97187, Luleå, Sweden
[2] Swedish National Rail Administration, SE-97102, Luleå, Sweden
E-mail: ambika.patra@ltu.se, uday.kumar@ltu.se, per-olof.larsson-kraik@banverket.se

Summary: In this paper, an approach has been developed to assess safety of the railway track by estimating the probability of derailment. Models for probability of derailment are developed based on undetected rail breaks and poor track quality using Petri-Nets and Monte Carlo simulations. The Effect of inspection intervals on lowering the probability has been analysed. The performance of the model is illustrated by an example from a track section of the iron ore line of Banverket (Swedish National Rail Administration).

Index Terms: Railway track safety, Maintenance, Petri-Net modelling

1. INTRODUCTION

Safety is the most important attribute of railway quality of service and operation. Infrastructure managers always try to reduce the number of potential risks areas that can lead to train accidents. Railway operations, that do not pay attention to reduce potential risks, face severe consequences. Risk has multi-fold consequences i.e., high society cost for compensation of deaths and injuries, material damages to infrastructure and rolling stock, cost of delay due to accidents and cost of damage to the environment. Proper maintenance planning helps infrastructure managers to reduce the potential risks. To study the effect of maintenance a case study was developed for a specific track section of the iron ore line of Banverket (Swedish National Rail Administration).The studied ore line Malmbanan, completed and operational by 1902, was later electrified with 15 kV 16 2/3 Hz, completed on 19th of January 1915, and remains so to this day. The studied track is a 1435 mm gauge. The track configuration is an electrified single track using block system. The signalling system (Low voltage DC track circuit), with its traffic control safety mechanisms, will detect any deviation that could be linked to a rail failure. However, the signalling system is not used as a maintenance planning/identification tool; it is a safety system for operating trains. Visual inspection is carried out separately by rail inspectors according to an inspection plan, recorded in a report and stored in a database. Visual inspection may also be carried out in an unplanned manner by the inspector to check the track condition between planned inspection intervals. Finally, track inspectors are obligated to report if they detect any deviation from normal rail condition, as they perform their daily maintenance work along the track.

Maintenance plays a vital role in improving safety performance of the track. In this paper the authors have tried to establish the relationship between maintenance intervals and safety performance. Petri-Net models have been used to model the safety performance of the track. The developed model in this work will estimate the maintenance investment required to achieve a specific safety level at a given point of time.

2. MODELLING TRACK SAFETY

Rail infrastructure consists of various sub-systems like track system, signalling and tele-communication system, and power system. Each of these sub-systems contributes to infrastructure safety.

Infrastructure managers use different indicators to measure the safety of the infrastructure e.g. number of accidents/per million train kilometres. Safety of the track is measured as number of derailments/per million kilometres as failures in track system leads to derailments not collisions. Broadly, performance indicators are classified as leading or lagging indicators. A leading, lead, or prospective indicator is a performance driver. The outcome measure itself is simply the lagging, lag, or retrospective indicator. Leading and lagging indicators can also relate to strategy or goals, and therefore it is important not to mix means and ends. These safety indicators are lagging indicator which only represents the current safety level of the track. If the infrastructure manager wants to improve the safety of track in future, it needs to have a lead indicator i.e., probability of derailment. Derailment because of track depends on the undetected rail breaks on track and poor track quality coupled with vehicle induced dynamic forces.

Let P_1 (t) = probability of undetected rail break on track at time t.
P_2 (t) = probability of track quality index falling below the maintenance limit at time t

Time here is expressed in Million Gross Tonnes (MGT). Probability of derailment due to rail breaks and poor track quality are given by K_1*P_1 and K_2*P_2 respectively. Factors K_1 and K_2 represent the external factors such as train speed, wheel condition, etc which induce dynamic forces. During winter time the trains have higher probability of getting wheel flats due to ice in the braking system. Wheel flats are major contributors to broken rail. In winter time the rail is in tensile stress due to low temperature which makes it more sensitive to external forces. It can be assumed that for a specific track section, K_1 and K_2 are constant because the track structure, speed ranges, climatic conditions do not change. Thus, if probability of derailment needs to be decreased, one needs to decrease P_1 and P_2. The model in this paper illustrates the effect of track inspection interval and track quality measurement interval on P_1 and P_2 respectively. The model relies on Petri-Nets and it provides dynamic means of modelling stochastic failure processes. A standard Petri-Net consists of a set of places, a set of transitions and a set of directed arcs. Directed arcs connect places to transitions and vice versa. The modelling is supported by software tool GRIF. Some of the data used in the models are taken from Banverket data bases **[1]** and some are hypothetical in nature. However, the assumed data are in close proximity to reality. The case study is based on a section of heavy haul line (10 kilometre in length) in north of Sweden.

2.1 Modelling rail breaks

There are many stresses that operate on rail and can influence rail defects and rail failure. Rail defects mainly consist of surface initiated defects, internal defects and weld defects. All these defects can potentially lead to rail breaks which are termed as rail failure **[2]**. Fig. 1 describes the development derailments due to rail defects.

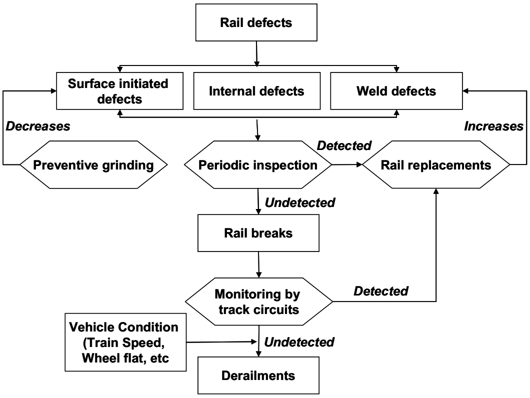


Figure 1: Logic flow of train derailment due to rail defects

Preventive grinding on the rail removes the rail surface initiates cracks in their initial phase and thereby stops their growth. Thus, formation of surface initiated rail defects such as head check, squats, etc are minimised. Internal defects (sub-surface initiated), such as shell and transverse defects, are associated primarily with heavy-haul railways. Though grinding is not used in general to remove embryonic cracks that cause shell formation, transverse re-profiling of the rail reduces stresses causing crack growth and hence shell formation **[3].** Weld defects are quite common on a heavy haul line especially the thermite welds. Until 2006 Banverket used flash butt welding to weld 40m rail to 320m and then 320 meter rails were welded by thermite welding. Today 60m rails are flash butt welded to 420 m and then 420 m rails are thermite welded. This decreases the number of thermite welds on track. In the proposed model only defects in thermite welds have been considered. Flash butt weld defects have not been considered as they show very low failure rate. On a heavy haul line, thermite welds on the rail become defective due to high axle loads in combination with cyclic loading. When a defective weld is repaired one more new weld is introduced and when a rail defect is removed additional two welds are added to the rail. Thus, the rate of thermite weld defect increases due to increase in number of welds. Periodic inspections are made to detect rail defects and remove them; however, there is a probability of detection attached to the ultrasonic inspection depending on the size of defect.

The P-F (Potential failure to Failure) interval of the rail defects is an important factor as it denotes the time interval between potential detection of rail defects till a failure (rail break) occurs **[4]**. The P-F interval for the rail defects is given in Table 1. Rail breaks are primarily detected by track circuits. However, not all rail breaks are detected by track circuits if rail breaks do not create a gap between the rails. These rail breaks remain undetected on the track until the next periodic inspection i.e., NDT or visual inspection. It is assumed in the paper that NDT inspection detects the rail breaks 100% of the time where as visual inspection detects 10% of the time. Undetected rail breaks on the track pose a serious threat to derailments. The probability of derailment should also consider the vehicle dynamics along with undetected rail breaks on track. In this paper contribution of vehicle dynamics to derailment has not been discussed. Authors have tried to reduce the probability of derailment by reducing the probability of undetected rail breaks by keeping the speed factor as constant as discussed earlier.

A Petri-Net model for estimation of undetected rail break has been developed (see Fig. 2). The model calculates the probability of undetected rail break(s) at any given point of time. Some of the parameters used by the model are given in Table 1. Table 1 illustrates the four types of defects that lead to rail break in the current study. UIC 421 is a thermite weld defect where as the other three are rail defects. The description of these defects can be found in UIC-712R i.e., catalogue of rail defects. As shown in Table 1, all these defects follow 2-parameter Weibull distribution. Detection probabilities of these defects by NDT car as well as visual inspection are also mentioned in Table 1. The rail is inspected by NDT car at every 12 MGT and visually every 0.5 MGT with the annual tonnage on the line is 24 MGT. If a rail break occurs, it is detected by track circuits. However, in case of rail breaks that do not separate the rails or rail gaps are small and are not detected by track circuits. In this model rail break detection probability of track circuits is assumed to be 0.98. Initial number of thermite welds for 10 km track is considered to be 32.

Table 1: Parameters for rail defects

Defect type	Scale parameter (η) in MGT	Shape parameter (β)	Detection probability by NDT Car	Detection probability by visual inspection	P-F interval in MGT
UIC 135	225	2.5	0.90	0.06	8
UIC 211	338	2.5			
UIC 2321	375	3.6			
UIC 421	333	3.1			

There can be a number of places on the rail where defects (surface initiated and internal) may occur and these places will change with respect to time. In case of weld defect (UIC 421), the number of welds will determine the number of defects. Number of potential defect locations (PD) can be calculated from the equation given below

PD (t) = cumulative number of defects in time t / cumulative probability of defects in time t

Here time is considered in terms of MGT.
As defect is following a Weibull distribution, probability of defect in time t is given by
$F(t) = 1 - \exp(t / \eta)^{\beta}$

Cumulative numbers of defects are calculated from the inspection data of the rail. After each inspection the numbers of defects found on the rail are known. These defects are the defects that are detected by the NDT as well as visual inspections. When these defects are divided by detection probability, we get the probable number of defects that may have occurred during that inspection interval. When we add these defects with the defects of previous inspection intervals, we get the cumulative number of defects.

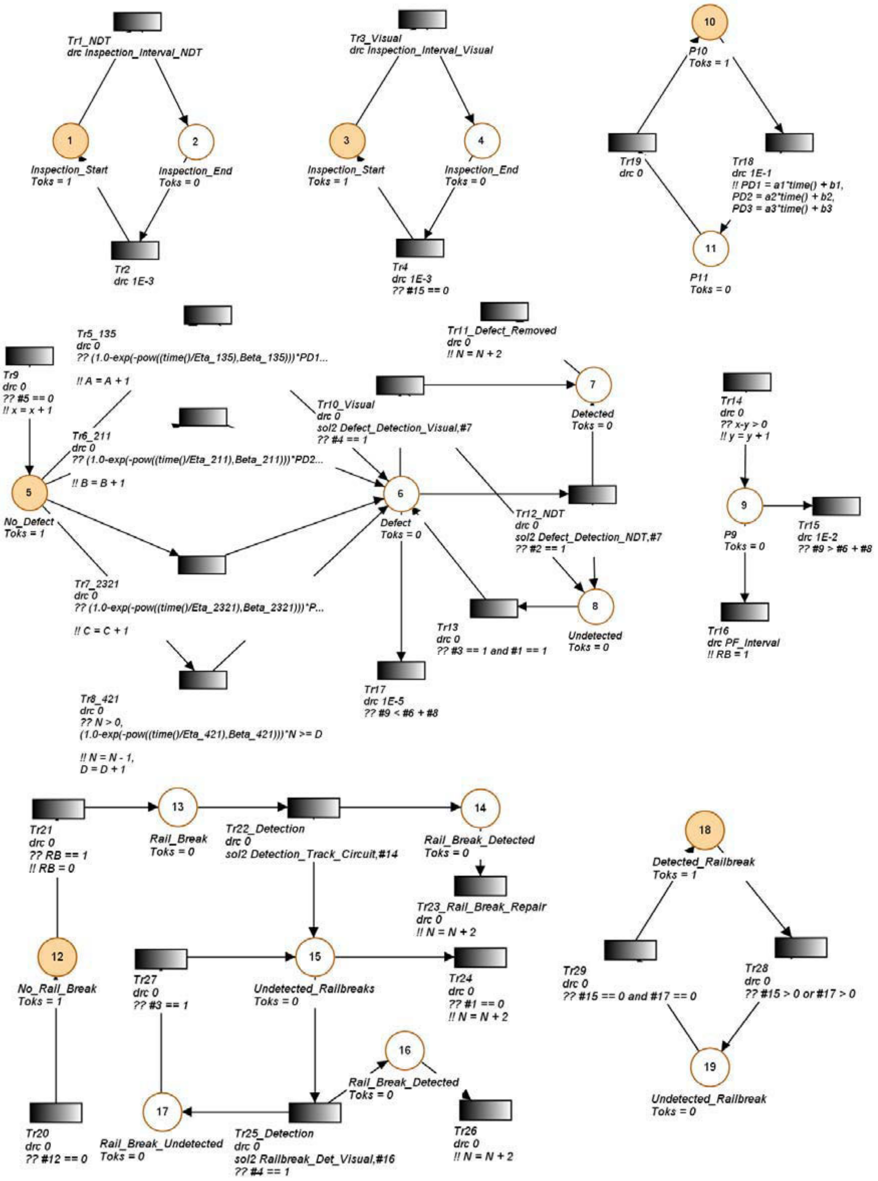


Figure 2: Petri-Net model for determination of undetected rail breaks

If we plot the values of PD with respect to time, we can get the trend for PD. When we multiply PD (t) with F (t), we can estimate the number of defects that we can expect at a given point of time. Table 2 explains the different places and transitions Petri-net model.

Table 2: Description of places and transitions of Petri-Net model for rail break

Places	Transitions
1: Start of NDT car inspection, 2: End of inspection, 3: Start of Visual inspection, 4: End of inspection	Tr1 and Tr3 fire at each inspection interval. Tr2 and Tr4 fire when the inspection is over
5: No Defect on rail, 6: Defects on rail, 7: Defects of rail that are detected by NDT or Visual inspection, 8: Defects that are undetected	Tr5, Tr6, Tr7 and Tr8 fire everytime a defect of either type occurs. Tr9 feeds the place 5 with 1 token each time a defect occurs.Tr10 fires when visual inspection is done and Tr12 fires when NDT inspection is done. Tr10 and Tr12 fire with detection probabilities mentioned in Table 1. Firing of Tr11 repairs the detected defects. Tr13 fires when defects are not detected by NDT or visual inspection. Those defects come to place 6 for next inspection.
9: Defects	Tr14 fires whenever a defect occurs. Tr15 fires when a defect is removed. Tr16 fires when an unremoved defect reaches its P-F interval. Firing of Tr16 initiates a rail break. Firing of Tr17 occurs when a rail break happens.
10 & 11: Act as counter for potential defect locations	Firing of Tr18 abd Tr19 calculate the potential defect locations for three types of defects in Table 1
12: No rail breaks, 13: Rail break, 14: Rail break detected by track circuit, 15: Rail break not detected by track circuit, 16: Rail break detected by visual inspection, 17: rail break undetected by visual inspection.	Tr21 fires when a rail break occurs. Tr20 feeds the place 12 with 1 token each time a rail break occurs.Tr22 fires with track circuit detection probability of detecting a rail break. Firing of Tr23 repairs the rail break. Tr25 fires with visual detection probability of a rail break. Firing of Tr26 repairs the rail break.Firing of Tr27 occurs when rail breaks are undeteced by rail visual inspection. Tr24 fires when NDT car inspection takes place and it repairs all the undetcted rail breaks from the track.
18: Rail break detected, 19: Rail break undetected	Tr28 fires when rail break(s) remain undetected. Tr29 fires when rail breaks are detected and repaired.

The incidences of rail defects are random in nature and the time for these defects to become rail breaks depend on the P-F interval of the defects. In this model it is assumed that if a number of rail defects occur during a period of time and remain undetected, the incidence of rail break depends on the P-F interval from the time of occurrence of the 1st defect. By performing Monte Carlo simulation on the Petri-Net model, probability of undetected rail breaks (with 90% confidence interval) with respect to increase in MGT on track has been found out (see Fig. 3). As seen in the figure, the probability of undetected rail breaks increases 5 times, when MGT increases from 200 to 300. The increase in probability is due to the fact that numbers of defects keep on increasing with increase in MGT. Thus, if proper maintenance measures are not taken with increase on accumulated tonnage on track, safety levels of the track will go down and more derailments will be expected to occur. Fig. 4 depicts the change in probability of undetected rail breaks with change in inspection interval from 12 to 6 MGT for MGT 250 to 300. Mean value of the probability was considered. While doing this sensitivity analysis all other parameters were kept constant. With increase in inspection frequency, infrastructure manager can estimate the extra maintenance investment that it has to put so that it can achieve the desired safety level.

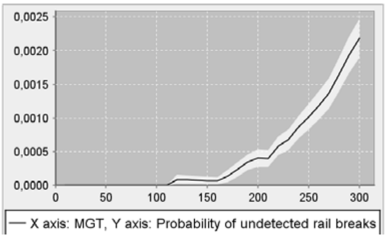


Figure 3: Probability of undetected rail breaks vs. MGT

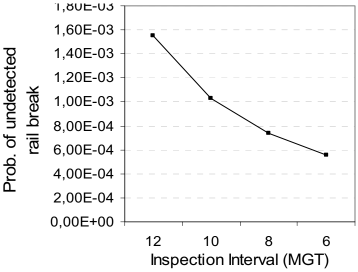


Figure 4: Change in probability of undetected rail breaks with change in inspection interval for MGT 250-300

As discussed earlier currently Banverket is reducing number of thermite welds by welding 420 m rails instead of 320 m. Also the manufacturing quality of rails has increased considerably which reduces defects like tache ovale (UIC 211). Fig. 5 illustrates the probability of rail break in the current practice. As these are newly laid rails, the defect statistics are not obtained yet. Thus, the same potential defect locations that were considered for old rails have also been considered in Fig. 5.

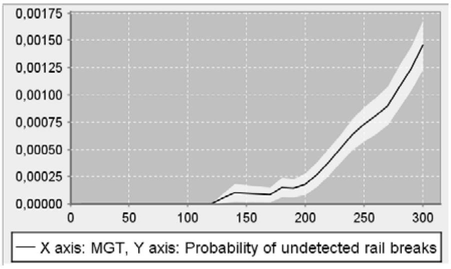


Figure 5: Probability of undetected rail breaks vs. MGT for newly laid rails

If we compare the probability of Fig. 5 with Fig. 3 at 250 MGT, we can see that in the newly laid rails there has been 25% decrease in probability.

2.2 Modelling track geometry deterioration

Track geometry deteriorates primarily due to the influence of dynamic loads exerted by vehicles. Continuous measurements of track geometry are necessary in order to make decisions on maintenance. Banverket has a number of condition indices to describe the condition of their infrastructure facilities. The main condition indices are known as K-value and Q-value. These are calculated from detailed inspection car measurements of the track. The inspection car measures relative rail position (lateral and vertical), rail profile and rail gauge. The Q-value is a weighted index of the standard deviation of two inspection car measures calculated as deviation from geometric comfort limits set for specific track class. The Q-value is calculated per kilometre track as:

$$Q = 150 - 100\left[\frac{\sigma_H}{\sigma_{H\,\lim}} + 2.\frac{\sigma_S}{\sigma_{S\,\lim}}\right]/3$$

where σ_H and σ_S are the average standard deviation of height and interaction on the section measured. The standard deviation for interaction is calculated as a combined effect from cant and side position of the rail. σ_{Hlim} and σ_{Slim} are the comfort limits for a given track class. Track class classifications are based on the speed of the train. Banverket uses the following levels for a specific class of track **[5]**.

- A: New built or recently adjusted track
- B: Lower quality limit. It states target value for maintenance actions. The track irregularities should normally be adjusted before this level attains. This limit is often related to comfort aspects.
- C: This limit should not be exceeded. The track irregularity must be corrected as soon as possible. Reduced speed limits should be taken into consideration until the irregularities have been corrected.

In this paper Q value for maintenance limit for the track is taken as 82. If the Q value falls below maintenance limit and tamping is not performed then probability of derailment increases. Q value is measured by the measuring wagon at every 24 MGT. Table 3 provides the Q value with passing tonnage and the time when tamping was preformed on track.

Table 3: Data of track quality measurement and tamping

Measurement (MGT)	Track Quality Index (Q)	Measurement (MGT)	Track Quality Index (Q)
24	95	144	79
48	88	168	92
72	81	192	85
96	94	216	78
120	86		
Tamping: 72 MGT, 144 MGT, 216 MGT			

By treating the data provided in the Table 3, the slope of Q value with MGT and the effectiveness of tamping were calculated. These values were used in the Petri-Net model described in Fig.6. Table 4 describes the different places and transitions mentioned in Fig. 6. Fig. 7 illustrates the change of Q value with passing MGT where as Fig. 8 depicts the probability (P_2) of Q value below maintenance (tamping) limit.

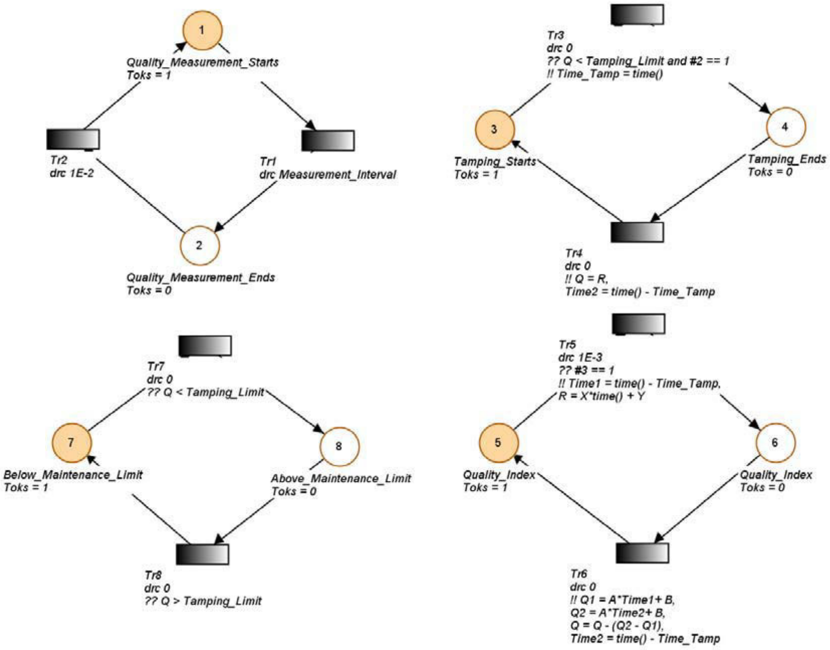


Figure 6: Petri-Net model for determination of probability of track quality index exceeding maintenance limit

Table 4: Description of places and transitions of Petri-net model for track quality

Places	Transitions
1: Start of measurement, 2: End of measurement	Tr1fires at each measurement interval. Tr2 fires when the measurement is over
3: Start of tamping, 4: End of tamping	Tr3 fires when Q value is below tamping limit and it is detected by the measuring wagon. Tr4 fires when the tamping is over.
5 & 6 : Act as counter for quality index	Firing of Tr5 and Tr6 calculate the track quality index at any given point of time.
7: Q value below maintenance limit, 8: Q value above maintenance limit	Tr7 fires when Q value falls below maintennace (tamping) limit and tamping is yet to be carried out. Tr8 fires when tamping is done and Q value is above maintenance limit.

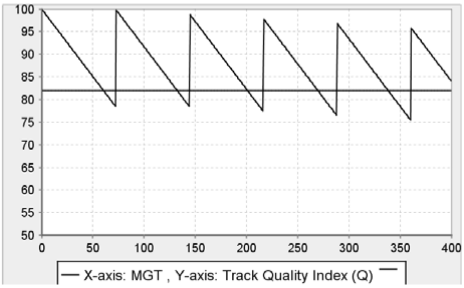


Figure 7: Q- value vs. MGT

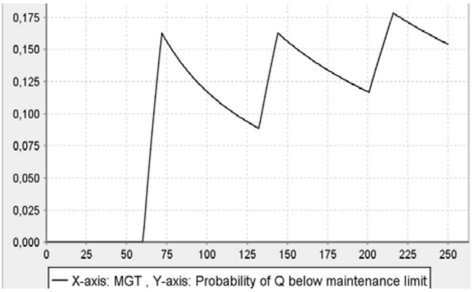


Figure 8: Probability of Q value below maintenance limit vs. MGT

It can be seen in Fig. 8 that as soon as the tamping is done on track, the probability decreases and hence the safety performance increases. However the overall probability increases with time; Fig. 9 illustrates the change in probability of Q value below maintenance limit with change in measurement interval for MGT 250 to 300. The mean value of the probability was considered. With increase in measurement frequency, infrastructure manager can estimate the additional maintenance investment.

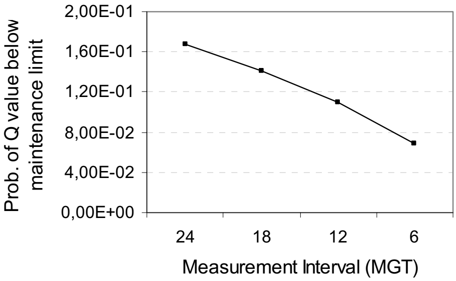


Figure 9: Change in probability of Q value below maintenance limit with change in measurement interval for MGT 250-300

3. DISCUSSION AND CONCLUSION

The safety performance of the railway track is compromised by derailments. Derailments take place due to faults on the track and/or due to bad vehicle dynamics. In these paper derailments due to undetected rail breaks and poor track quality have been described. However, derailments can also occur due to track buckling. Track buckling happens when the thermal stress in the track exceeds the track lateral resistance. Lateral resistance of the track changes due to maintenance work on the track. After each maintenance action, if the track lateral resistance is not restored to the original value, buckling may take place. The probability of buckling was not considered in this paper because probability of buckling depends on the quality of maintenance work rather than maintenance frequency. The Probabilities obtained in this paper can be used as safety indicators for the track. It has been shown in the paper that how frequency of track inspections and track quality measurements affect the probabilities. Reductions of these probabilities reduce the risk of derailment. This model will help the infrastructure managers to estimate additional maintenance investment to increase safety performance of the track to a desired level.

References

[1] Kumar, S., **Reliability Analysis and Cost Modelling of Degrading Systems**, PhD Thesis, Luleå University of Technology, Luleå, Sweden, ISSN: 1402-1544

[2] Zhao, J., Chan, A. H. C. and Stirling, A. B., **Risk Analysis of Derailment Induced by Rail Breaks − a Probabilistic Approach**, Annual Reliability and Maintainability Symposium, 2006, pp.486-491

[3] Cannon, D. F. *et al.*, **Rail Defects: An Overview**, International Journal of Fatigue & Fracture engineering Materials & Structures, Vol. 26, October 2003, pp. 865-886

[4] Vatn, J., Podfillini, Z. E. and Svee, H., **A Risk Based Approach to Determine Type of Ultrasonic Inspection and Frequencies in Railway Applications**, the World Congress on Railway Research, 2003, Edinburgh, UK

[5] Lindahl, M. **Track Geometry for High-speed Railways**, TRITA – FKT Report, Royal Institute of Technology, Stockholm, Sweden, ISSN: 1103: 470X

Paper III

Availability analysis of railway track circuit

Patra. A. P. and Kumar, U. (2009). Availability analysis of railway track circuit. *Accepted for publication in Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit.*

Availability analysis of railway track circuit

A P Patra[*] and **U Kumar**

Luleå Railway Research Center, Division of Operation and Maintenance Engineering, Luleå University of Technology, Sweden

Abstract: Railways are expected to operate with ever increasing availability. The availability of railway systems and subsystems influences the overall operational availability. As the track circuit is a key component of railway signalling and control, it can contribute significantly to the loss of availability of the railway system. One way to increase the availability of the track circuit is through implementing proper a maintenance policy. In this article, Petri-Nets have been used to develop models for availability analysis of the track circuit for both time-based and condition based maintenance. A sensitivity analysis has been performed to estimate the effect of maintenance parameters on availability performance using cost-benefit analysis.

Keywords: Track circuit, Availability, Petri-Nets

1. INTRODUCTION

The railway track circuit is an electrical device installed on the track to provide an indication of train presence on a particular section of a track. The track circuit has been the most commonly used train detection device besides the axle counter. The advantage of the track circuit over the axle counter is its ability to detect rail breaks. Its correct operation is critical to obtaining dependable train operations in most main line and metro systems. The track circuit is a fail-safe device ensuring that any fault results in the signal light turning red, and in trains being prevented from entering the associated track section. However, this fail-safe property also has the potential to cause significant train delays if the system becomes unreliable. Any track circuit failure can cause significant disruption to rail services and hence can become a safety risk due to delay in the restoration of normal service. Over 12,000 track circuit failures were reported in the UK during 2004–2005, resulting in 1.5 million minutes of attributable delay. Typically, the UK railway infrastructure manager can be penalised with a penalty of £20–60 per delay minute arising from infrastructure failure **[1]**. Therefore, the ability to detect and diagnose track circuit failures in order to provide a fast response to failures/incidents has significant economic benefits.

Reduction of the un-availability caused by track circuit failures must be accomplished using an assured methodology. This methodology usually requires either an optimisation or a complete overhaul of the maintenance processes in place. This starts with understanding the causes of track circuit failures, which can either be within the track circuit itself, e.g. component or systemic failure, or result from a failure of part of the infrastructure, such as a poor track bed or related components. Railways all over the world follow a range of different track circuit maintenance policies. However, the problem is to find the cost-effective maintenance policy which will provide higher availability of the track circuit at a lower life cycle cost. To address this problem, it is important to determine the life cycle of the track circuit and the relationships between the failure distribution and the maintenance policy. The cost of each maintenance policy can be calculated by adding the direct maintenance costs and the consequential costs, e.g. the penalties due to train delay. This article provides a comparison between the track circuit availability achieved using scheduled maintenance and that achieved using condition based maintenance. The influence of the maintenance parameters on the availability in both cases is studied for track circuits currently in use across Sweden. There are different types of track circuits being used by infrastructure managers, e.g. DC track circuits, AC track circuits and audio frequency track circuits. In this article our comparisons have concerned the maintenance strategy for the DC track circuit which is used by the Swedish National Rail Administration (Banverket).

The DC track circuit is the simplest and least costly type of track circuit used for train detection. Fig. 1 presents the schematic diagram of a DC track circuit.

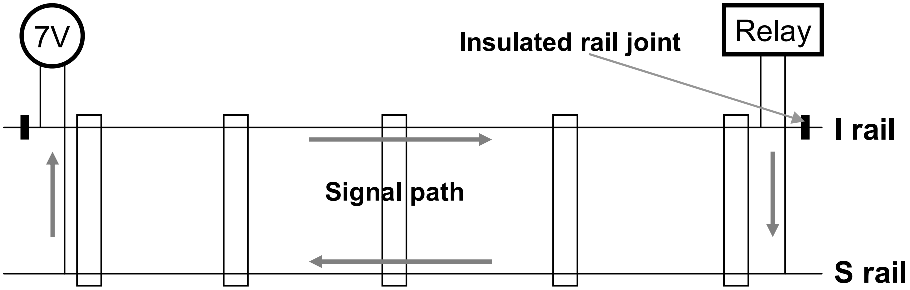


Fig. 1 Schematic diagram of DC track circuit

Banverket's the train detection system uses the two rails on a track, the common rail (S-rail) and the information rail (I-rail), to locate the position of the trains. The S-rail is continuous and is grounded and used for the track return current. The I-rail, on the other hand, has gaps with an insulated rail joint between the different sections of the rail. The polarity of each section is reversed so as to prevent the supply to one circuit from powering the adjacent circuit. The gaps isolate each section of the I-rail with a different potential. When no trains are on the line, there is a negative or positive potential between the rails. When a train moves into a new section (over a gap to a new rail section on the I-rail), the wheels complete the circuit connection and the potential drops to zero. The energized relay for the occupied section holds the voltage drops indicating where the train is on the track **[2]**. This track circuit design is fail-safe; i.e. any failure in the track circuit turns the signal to red. Thus, the safety of the track is not compromised. However, frequent track circuits failures affect the reliability and availability of the track.

To achieve a reliable system, the factors which could influence the reliability of the system need to be identified, their effects need to be assessed and the causes of these effects need to be managed throughout the life cycle of the system. Railway system reliability is influenced in three ways **[3]**:

- *System conditions*: the sources of failures are introduced internally within the system at any phase of the railway system life cycle. These failures are incurred by the design and manufacturing of the components or the system.
- *Operating conditions*: the sources of failures result from the operating system methodology. These failures are also incurred by environmental conditions.
- *Maintenance conditions*: the sources of failures are caused by maintenance actions.

The functionality of a track circuit is affected by the failure of its components, by changes in track characteristics or by track maintenance. The DC track circuit failure modes include:

- The failure of track circuit components
- Insulated rail joint failure
- Rail breaks
- Lower ballast resistance
- False alarms

Faults in track circuit components can be due to power supply failure or failure of the relay. There can be loose cable connections due to vibrations and cable discontinuities due to breaks in the cable caused by track maintenance actions, e.g. tamping. These failures are intermittent and random in nature. The failure distributions of such events are likely to follow exponential probability distributions. Insulated rail joints have shorter service lives than most other track components. The frequency of insulated joint failures is influenced by the relative and continuous weight (tonnage) of the traffic using the rails. Therefore, the frequency of failures of insulated joints is high in heavy haul. Capturing condition information about joints and taking preventive action remain challenges, as it is still proving difficult to obtain reliable trending information on the condition of joints. The failures of these joints are time-dependant and increase with time and generally follow Weibull probability distributions. Consequently, time-based inspection is recommended as an approach to detecting degradations in insulated joints and replacing them before they fail. However, it is not easy to optimise the inspection periods in a track section with a number of insulated joints installed at different times. Broken rails follow similar probability distributions to those of insulated joint failures.

The track circuit operates at a specified electrical ballast resistance. When the electrical resistance of the ballast is lower than the specified value, the current flow along the rails drops and de-energizes the relay, which makes the track circuit non-functional. This is a frequent occurrence as a result of wet or dirty ballast. Ballast cleaning is necessary to keep the ballast dry and the electrical resistance above the minimum value. Occurrences of these kinds of failures are intermittent in nature and likely to occur more during winter than in summer. False alarms where the track circuit equipment has triggered a fault but where no fault is found are phenomena which are registered as NFF. These faults occur with no prior notification and are unpredictable in nature. False alarms normally are a significant subset of track circuit failures. As track circuit failures stop the rail service operation, it is necessary to detect degradation in track circuits and repair them before they actually fail. The following sections describe the features of scheduled maintenance and condition based maintenance, the differences between the two types of maintenance, and the benefits of condition monitoring over scheduled maintenance.

2. MAINTENANCE POLICY

Infrastructure managers generally employ a scheduled (time based) maintenance regime for track circuits. The objective of this maintenance is to detect any degradation in the track circuit regularly and perform preventive maintenance to bring it back to a known acceptable state. This type of maintenance regime is expensive and time-consuming, since inspection needs to be carried out on every track circuit periodically (e.g. every 6 weeks for DC track circuits, which is the interval used by Banverket). However, sudden failures can, and do, occur between scheduled maintenance actions. In the event of such failures, urgent trackside corrective maintenance is carried out, which is costly, particularly when it has to be carried out during traffic hours. The periodic inspection of track circuits consists of measuring the voltage and current at different locations of track circuits, checking the insulated joints, inspecting the ballast condition, etc., to capture degradations (if any). If the measured current and voltage are different from the specified value, the track circuit is degraded and maintenance is carried out. However, if the occurrences of degradations of track circuits are random in nature, then it is difficult to optimise the scheduled maintenance interval in order to increase the system availability and reduce the cost.

Condition monitoring systems are designed and implemented so as to learn about system degradation and therefore reduce the maintenance burden and cost. The ability to detect degradation, identify certain incipient faults and/or provide diagnoses of failed track circuits, in a more ‘intelligent’ way, would have significant operational and economic advantages. One of the most important aspects of a robust track circuit condition monitoring system is its ability to identify degradation and failure modes, then to detect and locate a fault when it occurs, and to predict incipient failures so that potential damage can be avoided. Preventive maintenance can be performed before total failure **[4]**. The potential benefits of track circuit condition monitoring include:

- Improvement in the track circuit reliability and availability— by reducing the downtime during operational hours and hence reducing train delays.
- Improvement in the post-incident analysis — by improving the fault diagnosis capabilities, in particular by locating and classifying faults in a failed track circuit.
- Reduction in the number of ‘Tested OK on arrival’ events.
- Reduction in the scheduled down time — periodic routine maintenance tasks on the track circuit may not be required.
- Reduction in the safety risk associated with maintenance activities — by reducing the amount and length of trackside work.
- Reduction in the maintenance and replacement costs — the track circuit units can be maintained or replaced just-in-time if predictive condition monitoring can be achieved.
- Overall improvements in the track circuit performance — to provide a better understanding of the signalling network, optimise the train control system, reduce the penalty costs and increase the quality of rail services.

These are the potential advantages of track circuit condition monitoring systems. However, balanced against these advantages are the costs of establishing the requirements, of procuring, installing, “learning” and adapting the system and also of managing its own reliability. Therefore, the ability of the system to detect the probability of faults is a combination of the detection probability of the faults to be detected by the monitoring system and the reliability of the monitoring system itself.

In this paper we have modelled availability as a factor of the detectability of the monitoring system. The decision to adopt either scheduled maintenance or condition based maintenance is taken based on an evaluation of the cost-benefit analysis of both types of maintenance philosophy

3. AVAILABILITY MODELLING

Track circuit availability modelling has been performed by using Petri-Nets **[5]**. A standard Petri-Net consists of a set of places, a set of transitions and a set of directed arcs. Directed arcs connect places to transitions and vice versa. The places of Petri-Net model represent the degrading states of the track circuit and the transitions represent the time delay between the two degradation states. The modelling has been supported by the software tool GRIF. The following sections describe how estimates of availability for the two approaches to maintenance have been made.

3.1 Residual life

As discussed in the previous section, track circuit degradations and failures can occur for different reasons. A framework is presented in Fig. 2 for the treatment degradation and failure data of track circuits in order to calculate the reliability parameters. These failures constitute all the failure modes of the track circuit. As illustrated in the figure, in the present scenario track circuits are inspected at scheduled time intervals to check them for any degradation. If the track circuit is degraded, this is treated as a condition failure and preventive maintenance is carried out on the track circuit to bring the track circuit to good state. The track circuit reaches a degradation state in between two inspections, but this state can only be detected and the track circuit can only be maintained during the next inspection. In order to estimate the effect of the inspection interval on detecting a degradation of the track circuit before its functional failure, we need to calculate the time to degradation (degradation life) and the time from degradation to functional failure (residual life) of the track circuit.

Fig. 2 serves as an example of the representation of real data for the maintenance of different track circuits (TC1, TC2, etc.) in relation to time. During inspection, if a track circuit is found to be degraded, the degradation is termed as conditional failure and preventive maintenance is carried out. If the track circuit goes out of order, i.e. functional failure occurs, corrective maintenance is carried out. It is assumed in this paper that both these maintenance actions bring the track circuit to an ‘as-good-as-new’ state. It is also assumed in the paper that any manual inspection that is carried out on the track circuit at a scheduled time interval is perfect and free from any human error. This means that, if a track circuit is degraded, it is always detected during the inspection. It can be said that, when a track circuit is maintained preventively or correctively, the degradation must have occurred between the maintenance point of time and the last inspection point of time. This is shown as the occurrence of a degradation zone in the figure.

As the point of occurrence of the degradation between two inspections or between the functional failure and the last inspection is uncertain, the probability distribution of the degradation life can be estimated considering the interval data for degradation. Once we obtain the degradation life distribution, the next step is to estimate the residual life. The residual life is the remaining life in a track circuit from the point of degradation until it experiences a functional failure. When estimating the residual life of the track circuit, we should also consider undetected degradation times, which are also termed as suspended residual life.

A model has been developed in Petri-Nets to calculate the undetected degradation time (see Fig. 3). State 1 denotes that the track circuit is in good state and state 2 that it is in degraded state. Degradation of the track circuit brings the track circuit from state 1 to state 2. State 3 and 4 denote the inspection states. Inspection is carried out at a scheduled interval. A degraded track circuit remains in a degraded state until the next inspection commences. For an inspection interval of 1000 hours, by simulating the Petri-Net models we can see that the undetected degradation time is half of the inspection interval, i.e. 500 hours (see Fig. 4).

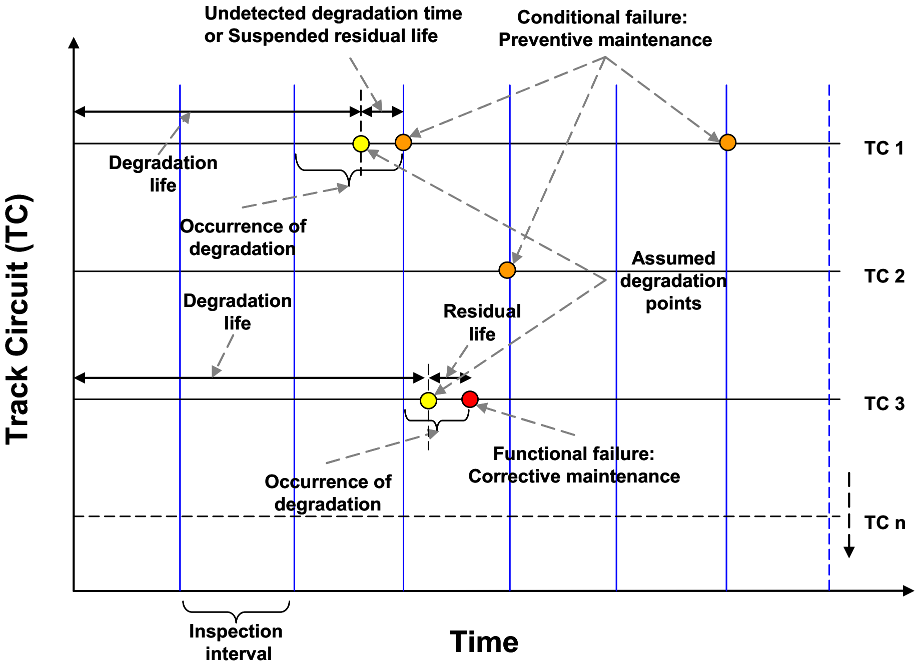


Fig. 2 An example of data representation for track circuit degradations and failures

The probability distribution of the degradation is assumed to be a Weibull distribution. It was seen that for other distributions, the undetected degradation time also remains half of the inspection interval. This result can also be inferred for the residual life calculation; i.e. the residual life is half of the time between the functional failure and the last inspection. Considering different values for the residual life and the suspended residual life (undetected degradation time), we can estimate the probability density function for the residual life.

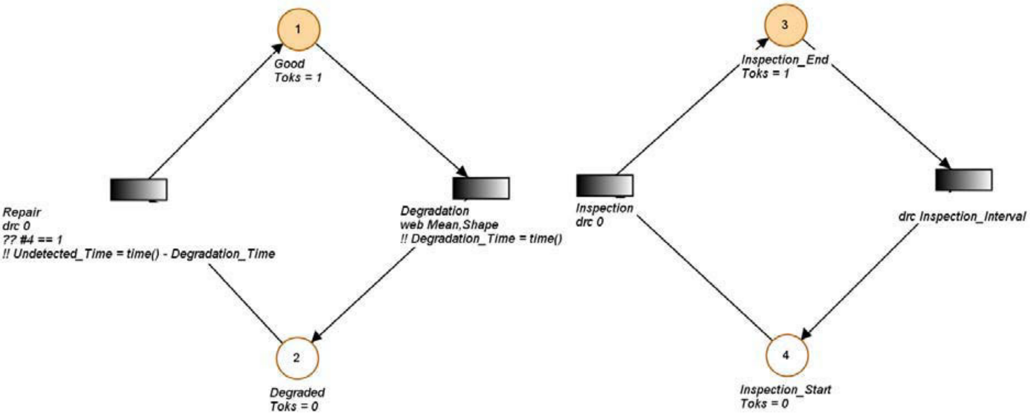


Fig. 3 Petri-Net model for estimation of undetected degradation time

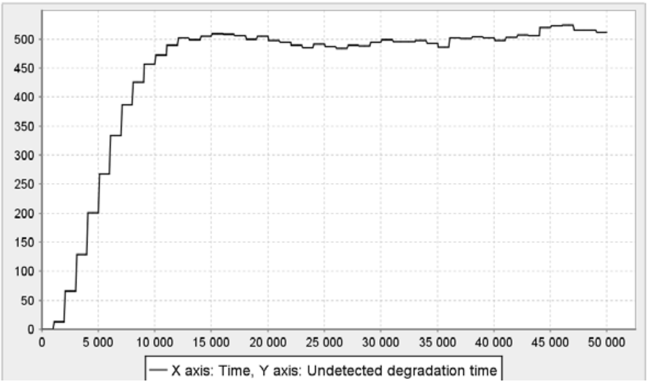


Fig. 4 Undetected degradation time

3.2 Scheduled maintenance

The availability estimation for the track circuit only considers the corrective maintenance. It does not consider the preventive maintenance, as the preventive maintenance is carried out during train-free periods; i.e. i.e. the unavailability during non-traffic hours is not included. A Petri-Net model for estimating the availability with the scheduled maintenance policy is shown in Fig. 5. State 1 denotes that the track circuit is in good state. The track circuit goes to state 2 when degradation occurs. It further goes to state 3 (failed state) depending upon the probability distribution of the residual life. Corrective maintenance in state 3 brings the track circuit back to good state. States 4, 5 and 6 are the inspection states. Inspection is carried out at every inspection interval. After the inspection is carried out, if it is found that the track circuit is in degraded state (state 2), preventive maintenance is performed on the track circuit and the track circuit goes back to good state.

For example, for each track circuit, the degradation follows a Weibull distribution with scale parameter 5000 hours and shape parameter 1.8, and the residual life distribution is a Weibull distribution whose scale and shape parameters are 700 hours and 1.2 respectively. The mean down time for both preventive and corrective maintenance is 2 hours and the inspection time is 0.5 hours. For an inspection interval of 1000 hours, the availability of the track circuit with respect to time is shown in Fig. 6. As shown in the figure, we obtain a steady state availability of 0.99985 for each track circuit. For a typical track section of 300 km, we can have an average of 200 track circuits. If we need to calculate the availability of the section due to track circuits, we must consider the delay aspects of track circuits due to their unavailability. Different indicators of availability are illustrated in **[6]**. In this article the authors define the availability of a section of track as the probability that any failure in that track section will not induce any delay to the trains. Based on this definition, the availability of the track section is given by:

$$A_s = 1- [(1-A^n)*P_d]$$

n = Number of track circuits in the section
A = Availability of each track circuit
P_d = Probability of delay

To explain how this expression is developed, let us consider the simple case of three track circuits in a track section. A is the availability and $\bar{A}$ is the un-availability of each track circuit i.e. all the track circuits are considered to be identical.

The probability of at least one of the three track circuits being un-available is given by:

$$\bar{A}_3 = \bar{A} + \bar{A} + \bar{A} - (\bar{A}* \bar{A}) - (\bar{A}* \bar{A}) - (\bar{A}* \bar{A}) + (\bar{A}* \bar{A}*\bar{A})$$

The un-availability of the track circuits is independent but not mutually exclusive (e.g. P(A + B + C) = P(A) + P(B) + P(C) – P(A)*P(B) – P(B)*P(C) – P(C)*P(A) + P(A)*P(B)*P(C))

Further, the probability of at least one of the three track circuits being un-available is also equal to one minus the joint probability that none of the track circuits are un-available.

Thus, $\bar{A}3 = 1 - [(1-\bar{A})* (1-\bar{A})* (1-\bar{A})] = 1- A^3$

If P_d is the probability of delay, then the un-availability of the track section is given by $(1-A^3)*P_d$.

Similarly, for n track circuits, the un-availability of track section will be $(1-A^n)*P_d$.

Hence, the availability of the track section is $A_s = 1- [(1-A^n)*P_d]$

If P_d is 1 (i.e. any un-availability of a track circuit in a section induces a certain delay), then $A_s = A^n$, where as, if P_d is 0 (i.e. there is no train delay in the case of any un-availability of a track circuit), then $A_s = 1$.

In this article it is assumed that any track circuit failure in a track section will induce train delay, and therefore the availability of the track section = $A^{200} = 0.99985^{200} = 0.9704$.

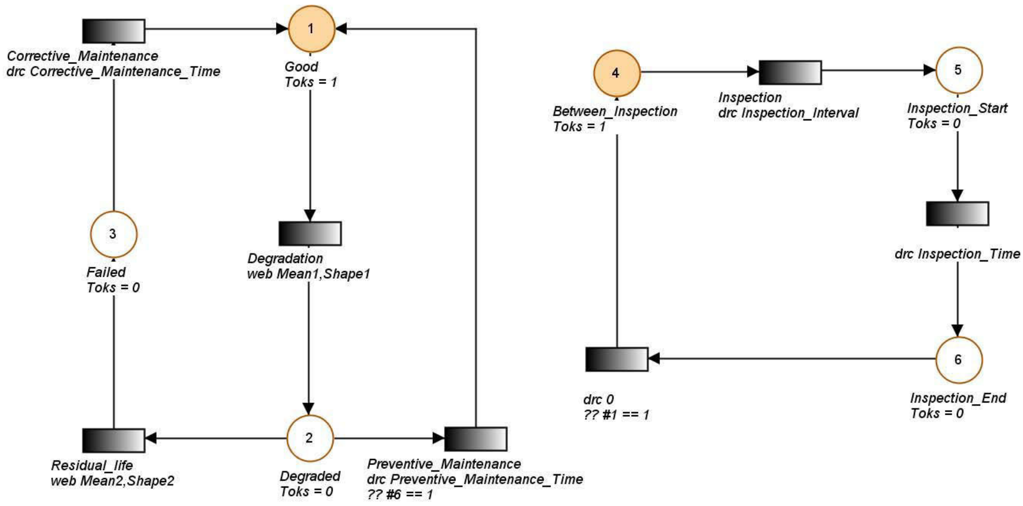


Fig. 5 Petri-Net model for estimation of availability for scheduled maintenance

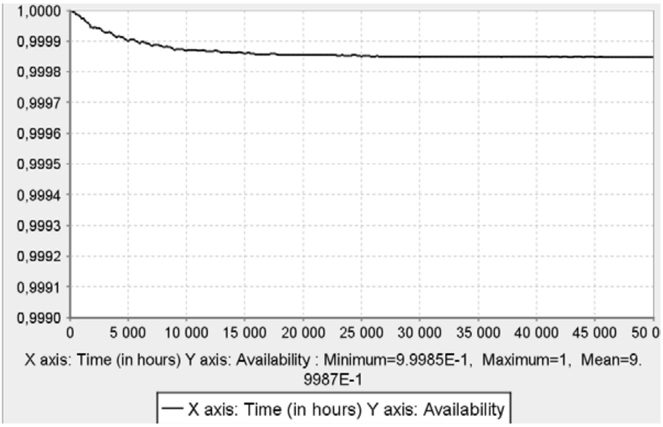


Fig. 6 Availability of track circuit for scheduled maintenance interval 1000 hours

If we want to increase the system availability, we need to perform a sensitivity analysis on the inspection interval. Fig. 7 illustrates the change in the section availability of a track circuit with a change in the inspection interval. If we want to optimise the inspection interval in terms of cost effectiveness, we need to calculate the cost of inspection as well as the cost of delay as described earlier.

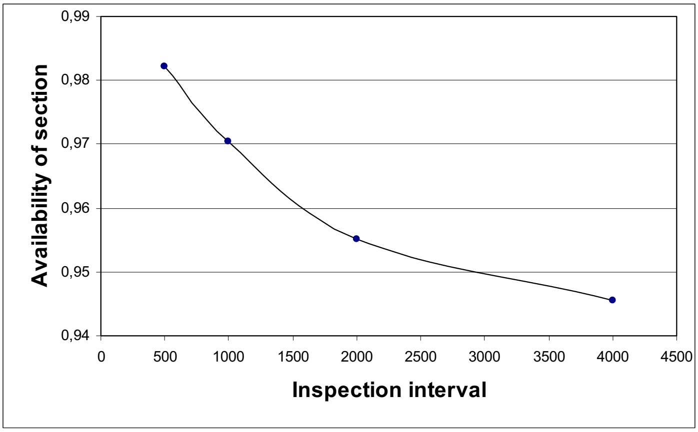


Fig. 7 Section availability vs. inspection interval

3.3 Condition based maintenance

As discussed earlier, condition monitoring on the track circuit will increase the availability of the track circuit. In this section we will illustrate the proportional increase in the availability of the same track circuit if we change the maintenance policy from scheduled maintenance to condition based maintenance. A Petri-Net model for estimating the availability with a condition based maintenance policy is shown in Fig.8. The detection probability of the condition monitoring system is considered in the model. As shown in the figure, when the track circuit is in degradation state, it can either go to detected degradation state (state 3) or un-detected degradation state (state 4), depending on the detection probability.

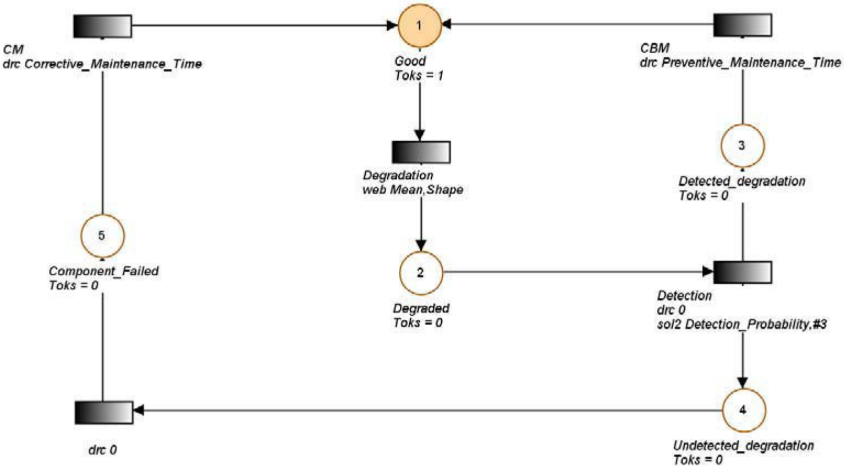


Fig. 8 Petri-Net model for estimation of availability for condition based maintenance

In this model it is assumed that the condition monitoring system detects the degradation in the track circuit all over its residual life. Therefore, the degradation life distribution in this case is the combination of the degradation life and the residual life distribution in the earlier example. If we consider the same failure and maintenance parameters as those considered for scheduled maintenance and perform a sensitivity analysis of the section availability with respect to the detection probability, we see the result in Fig. 9.

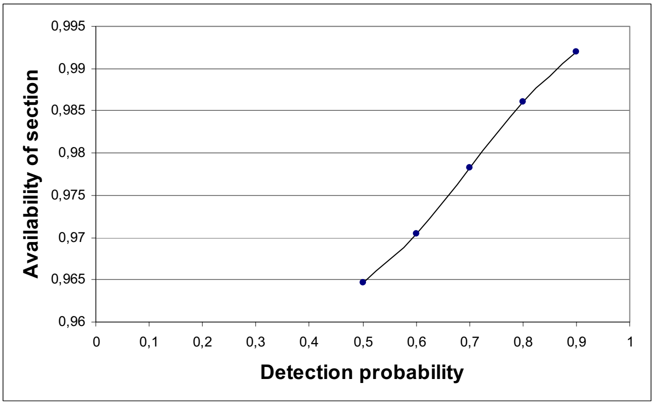


Fig. 9 Section availability vs. detection probability

As shown in the Fig. 9, one can achieve the same availability of the track section as achieved using current practice (scheduled maintenance with an interval of 1000 hours) by incorporating condition monitoring devices which have a detection probability of only 60%. However, it is very likely that condition monitoring systems have a higher detection probability than 60%. With a higher detection probability, we can achieve higher availability. However, to compare the cost effectiveness of condition based maintenance with that of scheduled maintenance, we must perform a cost-benefit analysis of each maintenance policy.

4. COST-BENEFIT ANALYSIS

As decisions on maintenance policies have a major impact on the operations of the system, the cost-effectiveness of long term design and maintenance decisions should be guaranteed. Cost-benefit analysis, an engineering economics technique, can be utilised to focus on maintenance strategies to minimise the maintenance cost in the system's life cycle, while meeting the dependability requirements. When comparing maintenance policies, the maintenance cost should include consequential costs (e.g. penalties due to train delays) in addition to direct maintenance costs. In the case of the scheduled maintenance policy, the maintenance costs include labour costs and equipment costs for inspection. If we want to change the maintenance policy to a condition based policy, we need to include the acquisition costs of monitoring systems, as well as the validation time and the costs for proving and developing the applicability of these systems and finally the cost of maintenance (i.e. corrective replacements) of the monitoring systems themselves. In both the cases we need to estimate the corresponding train delay and calculate the penalties.

Train delay consists of two types of delay, i.e. primary delay and secondary delay. Primary delay is due to some external circumstances, whereas secondary delay is caused by other trains. As secondary delay depends on the infrastructure design and train time schedule, in this paper we consider only primary delay. Failures of track circuits also directly affect the primary delay of trains. Fig 10 illustrates the primary delay incurred to trains because of track circuit failure. It is assumed that the headway between the trains is big enough not to induce any secondary delay.

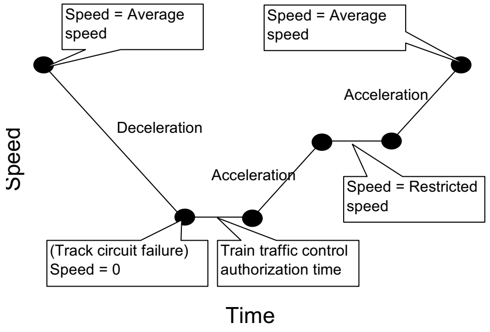


Fig. 10 Explanation of primary delay of trains

When a track circuit fails, the signal turns red and running of a train in the block section concerned is restricted. As shown in the figure, in the case of a track circuit failure, a train decelerates and comes to a standstill just before that particular block. The train then waits for the traffic control room's authorization to proceed on the affected block at a restricted speed. After the train covers that block at a restricted speed, it may accelerate to the original speed. The train is delayed by the extra time taken to cover the distance because of stopping and reducing its speed. If the track circuit is not repaired, then the following trains decelerate and cover that block section at a restricted speed. This adds further primary delay.

Consequently, the total primary delay per year due to the failure of track circuits can be calculated as:

Primary delay (minutes/year) = Failure rate of track circuit*Operation hours/year*[Primary delay of train * (Mean down time/Average headway between trains)]

In metro applications, where the headway is small, track circuit failures will give rise to significant secondary delay if the mean down time due to failure is high. The maintenance cost for each of the maintenance policies can be determined by adding the direct maintenance costs and the costs of delay. In the case of condition based maintenance, inspection costs are eliminated but we introduce additional costs for acquisitions of monitoring devices as well as maintenance of monitoring devices. The cost effectiveness of a maintenance policy can be determined by taking a ratio of the availability and the maintenance cost. The higher the cost-effectiveness, the better is the maintenance policy.

5. CONCLUSIONS

In this article, models for estimation of availability for scheduled maintenance and condition based maintenance have been developed. It has been demonstrated how the availability of the track circuit is influenced by changing the maintenance interval of the track circuit. The article also makes a comparison of scheduled maintenance and condition based maintenance based on achieving the availability of track circuit. A framework is presented for the cost of delay. Further, it is shown that cost-benefit analysis of maintenance policies can be carried out to evaluate the cost-effectiveness of each maintenance policy. Cost-effectiveness analysis will yield quantitative results to aid the decision maker with risk analysis, and provide a useful decision tool. Better cost-effectiveness of a system is achieved by higher operational availability and lower maintenance costs during the life cycle of the system.

ACKNOWLEDGEMENT

The work presented in this paper has been financially supported by the Swedish National Rail Administration (Banverket). The authors would like to acknowledge this support.

REFERENCES

1 **Chen, J., Roberts, C., and Weston, P.** Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems, *Control Engineering Practice,* 2008, **16,** 585-596.

2 **Niska, S.** *Measurements and Analysis of Electromagnetic Interferences in the Swedish Railway Systems*, PhD Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden, vol.76, 2008

3 **EN 50126** *The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) for Railways Applications,* Comité Européen de Normalisation Electrotechnique (CENELEC), Brussels, Belgium, 1999.

4 **Roberts, C., Dassanayake, H. P. B., Lehrasab, N.,** and **Goodman, C. J.** Distributed quantitative and qualitative fault diagnosis: railway junction case study. *Control Engineering Practice*, 2002, **10**(4), 419–429.

5 **Murata, T.** Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE*, 1989, **77**(4), 541 – 580.

6 **Nyström, B.** Use of Availability Concepts in the Railway System, *International Journal of Performability Engineering*, 2009, **5**(2), 103-118

Paper IV

Cost effective maintenance policy: a case study

Patra, A. P., Dersin, P. and Kumar, U. (2009). Cost effective maintenance policy: a case study. *Accepted for publication in the International Journal of Performability Engineering.*

International Journal of Performability Engineering
© RAMS Consultants
Printed in India

Cost Effective Maintenance Policy: a Case Study

AMBIKA PRASAD PATRA[1*], PIERRE DERSIN[2], UDAY KUMAR[1]

[1]Luleå University of Technology
SE-97187, Luleå, Sweden
[2]ALSTOM Transport Information Solutions
92364 Meudon-la-Forêt, France

(*Received on September 8, 2009*)

Abstract: The European Rail Traffic Management System (ERTMS) is a major venture initiated by the European Union in order to create an interoperable railway network within Europe. The manufacturers of the ERTMS strive to attain a competitive edge by demonstrating the efficiency of their systems. Achieving the optimal cost effectiveness is one of the significant ways to address the efficiency of the system and involves maximising the availability and minimising the life cycle cost of the system over the system life cycle. One of the important ways of maximising the cost effectiveness of the system is to optimise the maintenance policy. This paper demonstrates the estimation of the cost effectiveness of an ERTMS system. The degradation and repair process of the system is modelled by Petri-Nets. Failure, maintenance and cost data are used as parameters for the model. The model will be useful for systems that experience degradations and are subjected to imperfect maintenance.

Keywords: *Availability, Life Cycle Cost, Maintenance Policy, Petri-Nets*

1. Introduction

Each country in the European Union has its own railway signalling system. Each system is stand-alone and non-interoperable, and therefore requires extensive integration and engineering effort, raising the total delivery costs for cross-border traffic. This restricts competition and hampers the competitiveness of the European rail sector vis-à-vis road transport by creating technical barriers to international journeys. To fulfil the requirement of interoperability, a major industrial project named as the European Rail Traffic Management System (ERTMS) was initiated. The objective of the ERTMS is to replace the existing signalling systems with a system which will boost cross-border freight and passenger transport. This will help the countries to establish a more sustainable railway network. The ERTMS has two basic components, i.e., the European Train Control System (ETCS) and GSM-Radio (GSM-R). The ETCS is an automatic train protection system, while as GSM-R provides voice and data communication between the track and the train. There are 3 levels of the ERTMS, with Level 1 and Level 2 already in operation in Europe.

Apart from achieving an interoperable railway network, the ERTMS also increases the capacity, speeds and safety for passengers on existing lines, and at the same time reduces the maintenance costs. To be competitive and to gain the approval of infrastructure managers as well as train operating companies, the ERTMS manufacturers should demonstrate the cost effectiveness of their systems as per the stakeholders' requirement.

---

*Corresponding author's email: ambika.patra@ltu.se

The cost effectiveness of any system depends on the operational availability and life cycle cost (LCC). In order to make the system more cost effective, higher availability should be attained at a lower LCC.

However, there are numerous challenges in attaining the desired cost effectiveness of the system over a period of time, because of degradation of the systems, changes in traffic scenario etc. Optimisation of the maintenance policy is one of the major ways to attain the desired cost effectiveness of the system in the long run. All such optimisations should aim at maximum system availability and minimum life cycle costs, as well as minimum train delays for a specific traffic scenario. In this paper the authors demonstrate a model that can help to maximise the cost effectiveness of the systems. ALSTOM Transport has developed the ATLAS® platform for ERTMS application for railway operations. ATLAS® consists of various sub-systems, such as the Automatic Train Supervision (ATS) system, the interlocking system, the Automatic Train Control (ATC) system and trackside products e.g., track circuits. This paper deals with a case study where an optimum maintenance policy is developed to achieve the desired cost effectiveness of the Radio Block Centre (RBC), which is a part of the ATC track side system. The ATC sub-system consists of both trackside and trainborne components. The objective of this development work is to achieve a higher availability and a lower LCC for the system and then develop an optimum maintenance policy to achieve cost effectiveness.

As decisions on maintenance policies have a major impact on the operations of the system, the cost effectiveness of long term design and maintenance decisions should be guaranteed. Life cycle cost (LCC) analysis, an engineering economics technique, can be utilised to focus on maintenance strategies to minimise the maintenance cost in the system's life cycle, while meeting the dependability requirements. The cost effectiveness of a system can be defined as

$$\text{Cost effectiveness} = \frac{Availability}{LCC}$$

Higher cost effectiveness of the system ensures better operation of the system. The paper presents the key influential variables of maintenance policy that affect cost effectiveness. Section 2 of the paper discusses the system description of the RBC. A Petri-Net model is shown in Section 3. Section 3 also discusses the results of the model. Finally, the conclusions are stated in Section 4.

2. Radio Block Centre (RBC) System

In most of the systems which are in use at ALSTOM Transport, active redundancy has been chosen: i.e. the various units are active simultaneously, so that, in the event of the failure of one unit, the function is preserved without the need for switching on a back-up unit [1]. On ERTMS application level 2, the ETCS uses a GSM-R radio channel to exchange data between the trackside Radio Block Centre and the trains. The interlocking system reports the status of the objects controlling the routes of the trains to the RBC, which, in turn, generates the correct movement authorities for the different trains in the section. The RBC consists of different sub-systems, such as the computing channel, input/output system and cabinet. The reliability block diagrams of these systems are given in Figs. 1a, 1b, 1c and 1d. The computing channel is a 2-out-of-3 (2oo3) system. Two channels must be working at any given time for the computing channel to work. Similarly, the input/output group is a 1oo2 system.

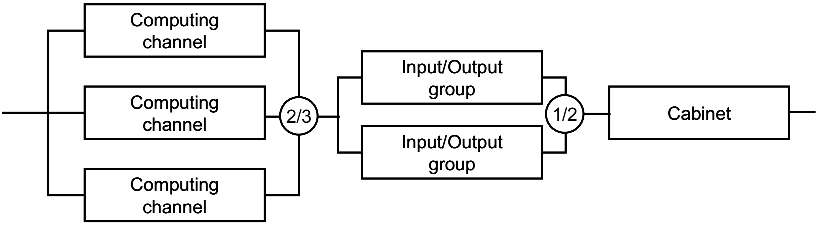


Figure 1a: Reliability Block Diagram of RBC

Figure 1b: Reliability Block Diagram of computing channel

Figure 1c: Reliability Block Diagram of input/output group

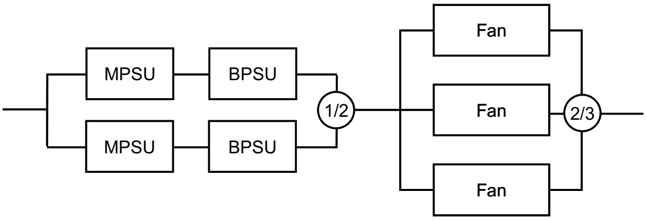


Figure 1d: Reliability Block Diagram of cabinet

Any failure in the RBC is detected by built-in-test-equipment (BITE). However, failures are detected depending upon the detection probability or fault detectability of the BITE system. When a failure is not detected, it remains in the system until the next inspection occurs. Eq. 1 provides an estimation of the undetected failure time if the fault remains un-detected till the next inspection. The relationship shown in Eq. 1 considers only the constant failure rate, but later on in the paper we develop a Petri-Net model which considers the non-constant failure rate.

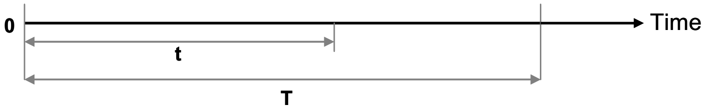


Figure 2: Undetected failure time

T = Inspection interval
t = time of occurrence of undetected failure during the interval (0, T)
T-t = duration of undetected failure time
λ = failure rate of the component following an exponential distribution

The expected un-detected failure time of the component during the interval (0, T) is given by:

$$E(T-t)=\frac{\int_0^T (T-t)f(t)dt}{\int_0^T f(t)dt}=\frac{\int_0^T (T-t)*\lambda*\exp(-\lambda t)dt}{\int_0^T \lambda*\exp(-\lambda t)dt}=(\frac{T}{1-\exp(-\lambda T)}-\frac{1}{\lambda}) \tag{1}$$

when $\lambda T << 1$, $E(T-t) = T/2$

The failure rate of a repairable component depends on the type of corrective maintenance that is applied in a range stretching from perfect maintenance to minimal maintenance. Perfect maintenance repair brings the component age to zero (i.e. the component becomes as good as new (AGAN)) whereas minimal repair keeps the component's age un-modified (i.e. the component stays as bad as old (ABAO)). In real world cases the repairs are neither AGAN nor ABAO, but are something in between. In [2], two models (type I and II) are proposed that estimate the virtual age of the component after a repair. Kijima models consider a parameter called the maintenance factor, which estimates the virtual age. If the maintenance factor is 1, the repair is ABAO and for the maintenance factor 0, the repair is AGAN. Model type I assumes that the repairs can only fix the damage incurred during the last period of operation. Thus, the n^{th} repair can only remove the damage incurred during the time between the $(n-1)^{th}$ and the n^{th} failures. Model type II assumes that the repairs fix all of the damage accumulated up to the current time. As a result, the n^{th} repair not only removes the damage incurred during the time between the $(n-1)^{th}$ and the n^{th} failures, but can also fix the cumulative damage incurred during the time from the first failure to the $(n-1)^{th}$ failure.

If the times between the failures are denoted by x_1, x_2, … $x_{n,\ t}$ the virtual age of the component after the n^{th} repair is given by

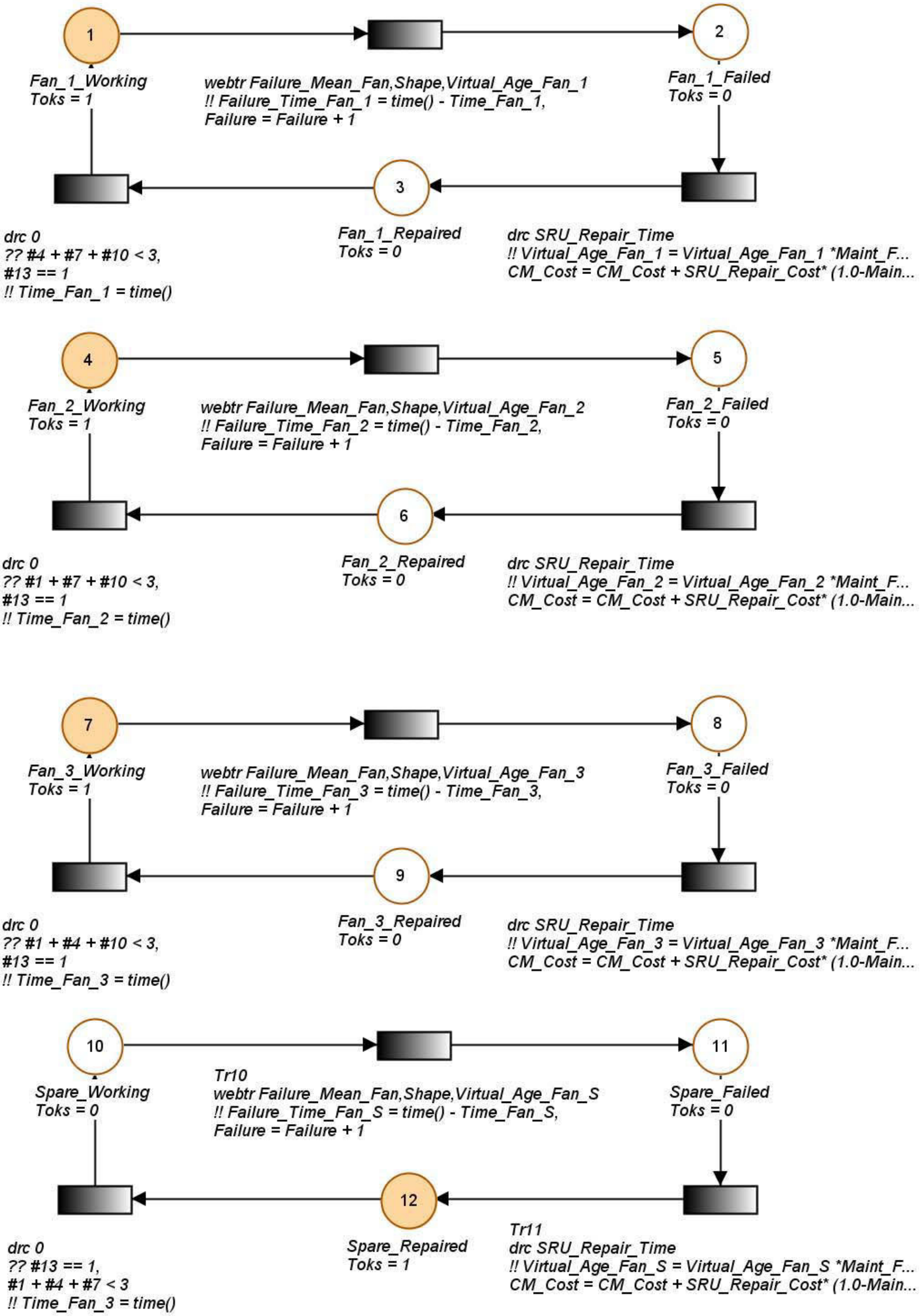
$V_n = V_{n-1}$ + (maintenance factor * x_n) →Kijima model type I
V_n = (maintenance factor*V_{n-1}) + (maintenance factor *x_n) → Kijima model type II

The maintenance factors have been calculated from the past failure times of the component by applying the Kijima models discussed above.

3. Case study on the cost effectiveness estimation of the fan system of the RBC

The Cost effectiveness of the RBC system has been estimated. In this paper we are presenting a case study on the fan system. As discussed earlier, the fan system in the RBC is a 2oo3 system. A Petri-Net (for details see [3]) model for estimation of the cost effectiveness of the fan system is illustrated in Fig. 3. The places that represent the states are depicted in Table 1. There are three fans in the system and one fan is kept as a spare in the store and all the fans are repairable in nature. As illustrated in the model, when a fan fails, it goes to the failed state. The time to failure depends on the failure probability density function of the fan. The model captures the time to failure each time a fan fails and estimates the virtual age of the fan after repair work is completed depending on the maintenance factor of the repair described earlier. If a fan fails and the failure is detected, it is removed from the system and is substituted by the fan in stock. The failed fan goes to the workshop for repair, and after repair it waits until the next fan failure occurs. Over a period of time the fans may have different virtual ages depending on the number of repairs performed on each fan. In this model, we can keep track of the number of repairs carried out on each fan, so that we can estimate the time when the next fan will fail. Now, place

13 denotes that all three fans are working. If one fan fails depending on the time explained above, the system can go to state 14 or state 15.



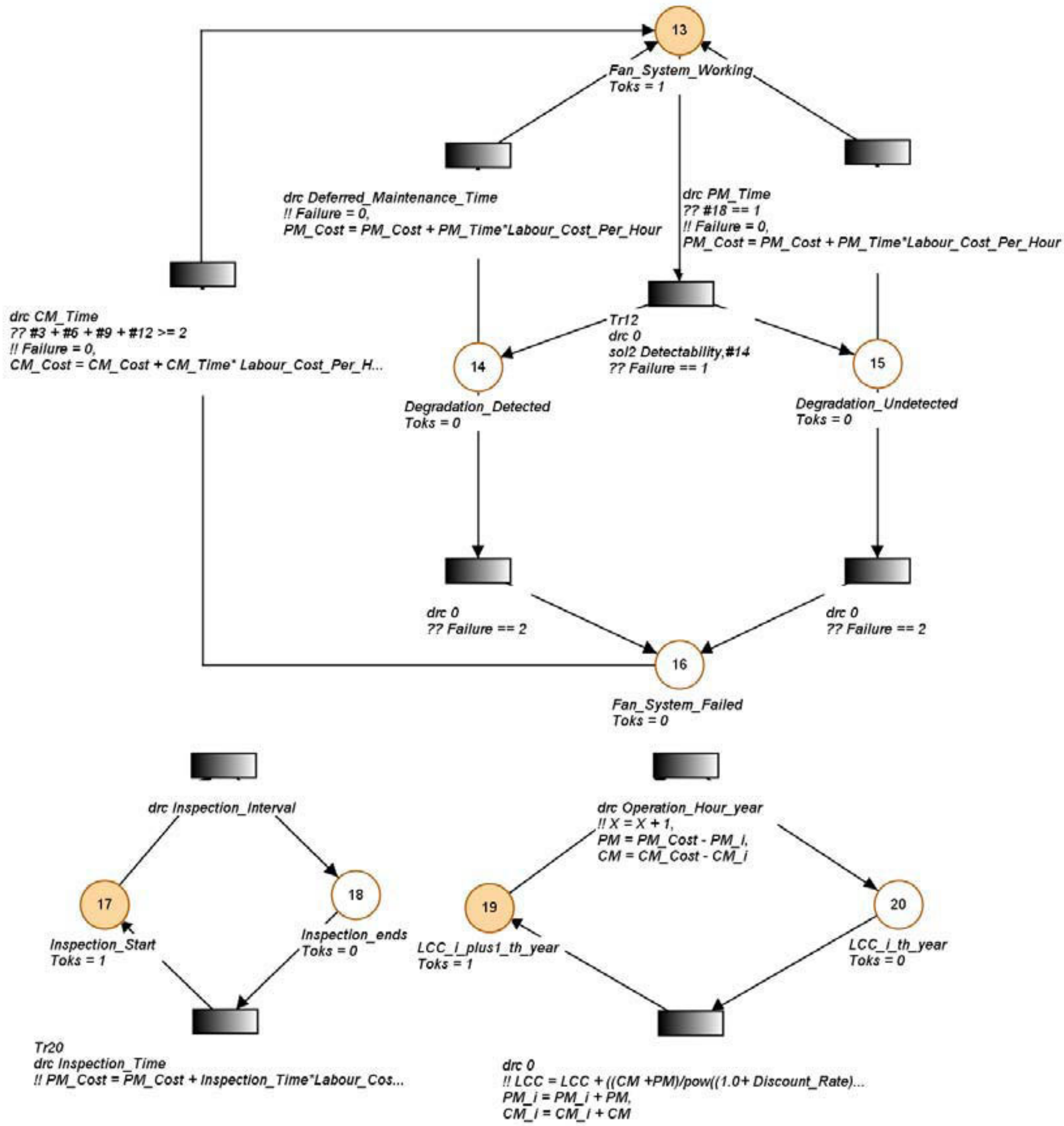


Figure 3: Petri- Net model for estimating the cost effectiveness of the fan system

If the failure is detected by the BITE system, the system will be in state 14, otherwise it will be in 15. Both state 14 and state 15 are degraded states but the system is still working because two fans are still working. If the system is in 14, it can go to 13 depending upon the deferred maintenance time. The maintenance is deferred because it is not always cost effective to stop the train operation to repair the system, and therefore the system can be repaired after the train running period. If the system goes to 15, the failure can only be detected in the next inspection, after which the system can be brought to 13. However, if the deferred maintenance time and the inspection intervals are long, then there is a probability that the system can go to the failed state if another fan failure occurs. In the case of a system failure, corrective maintenance is performed on the system to bring the system to 13. This depends on the availability of the two fans in the stock. If the fans are under repair, then the unavailability of the system increases. The parameters that are used in the Petri-Net model are illustrated in Table 2. Monte Carlo simulations are performed on the model to estimate the availability and LCC of the fan system over a period of 20 years.

Table 1: Description of places in the Petri-Net models

State description	Place	State description	Place	State description	Place
Fan 1 working	1	Fan 3 working	7	Fan system working	13
Fan 1 failed	2	Fan 3 failed	8	Fan system degraded (detected)	14
Fan 1 repaired and waiting in stock	3	Fan 3 repaired and waiting in stock	9	Fan system degraded (un-detected)	15
Fan 2 working	4	Spare fan working	10	Fan system failed	16
Fan 2 failed	5	Spare fan failed	11	Inspection starts	17
Fan 2 repaired and waiting in stock	6	Spare fan repaired and waiting in stock	12	Inspection ends	18

Table 2: Parameters used in Petri-Net model for cost effectiveness estimation

Parameters	Value
Operation hours/year	6000 hours
Inspection time	0.5 hours
Preventive maintenance time	0.5 hours
Corrective maintenance time	3 hours
Labour cost/hour	€ 40
SRU repair time	720 hours
Discount rate	4%
Fan failure (scale parameter)	20000
Fan failure (shape parameter)	2
Deferred maintenance time	9 hours
Inspection interval	6000 hours
SRU repair cost	€ 50
Maintenance factor	0.8
Detectability	0.9

Fig. 4 illustrates the cost effectiveness of the fan system with time. The step decrease in the cost effectiveness curve is due to the discounted value of the life cycle cost. After a period of time the cost effectiveness curve will be parallel to the x-axis. This is due to the fact that, after a period of time, the LCC will be constant because of the discounting of future costs.

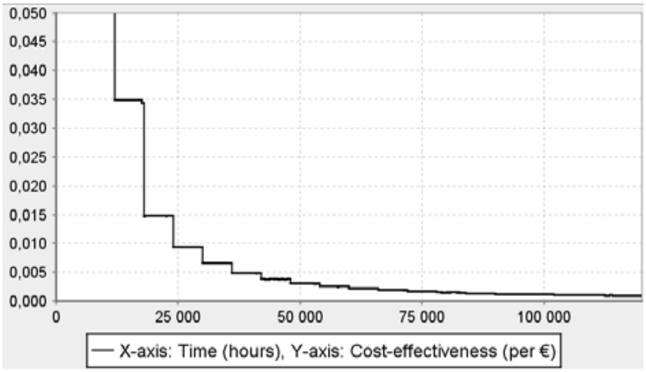


Figure 4: Cost effectiveness of the fan system with time

Sensitivity analyses have been performed on the maintenance factor, fault detectability, inspection interval and deferred maintenance time to study their effects on the cost effectiveness of the system. The results are shown in Fig. 5 and Fig. 6. The cost effectiveness decreases with an increase in the maintenance factor. That is because, with an increase in the maintenance factor, the quality of maintenance tends to be minimal maintenance, thereby increasing the number of failures. In the case of fault detectability, the cost effectiveness increases with an increase in the detectability. This analysis can help to design a better built-in-test system to achieve better cost effectiveness of the system. With an increase in the inspection interval, the probability of repairing components before they fail decreases and hence the LCC increases. With a decrease in inspection interval, the inspection costs increase and hence the LCC decreases. In this case study, as the failure rate is very low, the optimum inspection interval is obtained at a very long inspection interval. However, from a safety point of view, we cannot have a very long inspection interval. When the deferred maintenance time increases, the probability of failure from a degraded state increases. Hence, the cost effectiveness decreases. However, at the same time, an increase in the deferred maintenance time also increases the possibility of opportunistic maintenance, which decreases the overall costs.

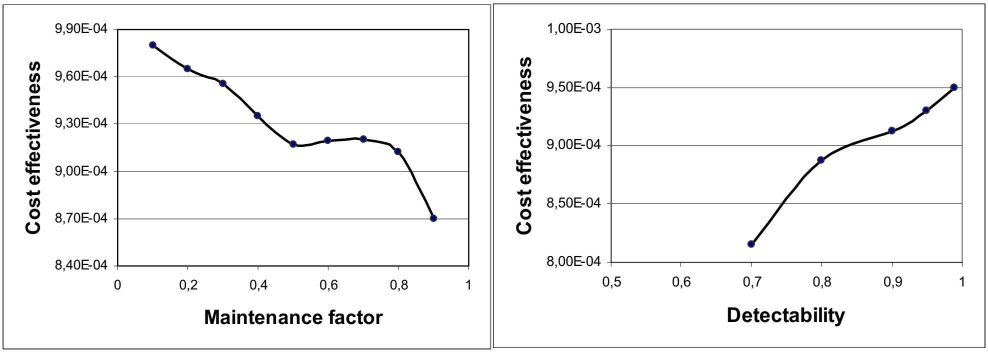


Figure 5: Effect of maintenance factor and detectability on cost effectiveness

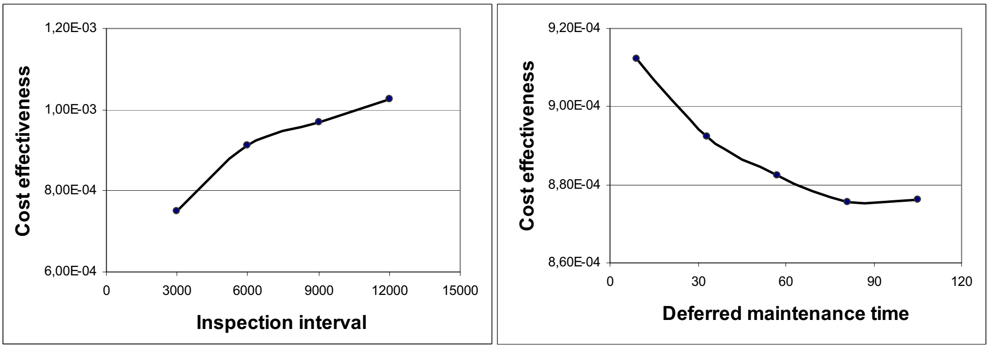


Figure 6: Effect of inspection interval and deferred maintenance time on cost effectiveness

We can also estimate the maintenance policy of the system which will ensure certain availability and fulfil LCC requirements. From Fig. 5 & Fig. 6, the optimum value for each parameter for maximising the cost effectiveness can be inferred. Further, the optimum value of combinations of the parameters for achieving the maximum cost effectiveness can also be estimated. The model can be useful for other mechanical redundant systems which are repairable in nature and subjected to degradations.

4. Conclusions

In view of the increasingly stringent availability requirements being set by the market place, the designers of complex systems have to pay close attention to test and maintenance strategies to achieve availability targets with low life cycle costs. In this paper a maintenance policy based on cost effectiveness has been developed for the fan sub-system of the Radio Block Centre (RBC). Sensitivity analysis has been performed on different maintenance parameters to maximise the cost effectiveness of the system. This maintenance policy will help the systems to achieve higher availability at lower life cycle costs over the life cycle of the systems. A Petri-Net model has been developed to calculate the cost effectiveness of these systems. Cost effectiveness analysis will yield quantitative results to aid the decision maker with risk analysis, and provide a useful decision tool. The work presented in this paper is a part of the work carried out for systems developed by ALSTOM Transport for ERTMS applications. This work will help the ERTMS manufacturers to demonstrate the sustainable benefits in terms of availability and life cycle costs to the infrastructure managers, as well as the train operating companies, in order to keep a competitive advantage.

Acknowledgement

The work presented in this paper is supported by ALSTOM Transport Information Solutions, Paris. The authors would like to acknowledge this support.

References

[1]. Dersin, P., A. Peronne, and C. Arroum. *Selecting test and maintenance strategies to achieve availability target with lowest life cycle cost*. In proceedings of Reliability and Maintainability Symposium, USA, January 2008.

[2]. Kijima, M. *Some results for repairable systems with general repair*. Journal of Applied Probability 1989; 26: 89-102.

[3]. Murata, T. *Petri Nets: Properties, Analysis and Applications*. Proceedings of the IEEE 1989; 77(4): 541 – 580.

Ambika Prasad Patra is a PhD candidate at the Division of Operation and Maintenance Engineering, Luleå University of Technology, Sweden. His research topics are RAMS and LCC analyses for railway infrastructure systems. He received his MTech degree in Reliability Engineering in 2002 from the Indian Institute of Technology, Bombay, India.

Pierre Dersin obtained his Ph.D. in Electrical Engineering in 1980 from the Massachusetts Institute of Technology (MIT), after receiving a Master's degree in O.R. in 1976, also from MIT. He worked in the late seventies on the reliability of large electric power networks, while taking part in the Large Scale System Effectiveness Analysis Program sponsored by the US Department of Energy, as an employee of MIT and Systems Control, Inc. He later joined FABRICOM (in Belgium and the U.S.A.), where he was involved with fault diagnostic systems for factory automation. Since 1990 he has been working at ALSTOM Transport, where he has occupied several positions and has mainly been involved with RAMS and maintenance. He has been R&D Manager of the Service section, and is now RAM Director in ALSTOM Transport's Information Solutions (i.e. railway signalling and communication). He has contributed a number of papers and publications in the fields of RAMS, automatic control and electric power systems. He was the chairman of UNIFE's (the Association of the European Rail Industry) LCC/RAM Working Group in 1998-2001.

Uday Kumar is Professor of Operation and Maintenance Engineering at Luleå University of Technology, Luleå, Sweden. He is also Director of Luleå Railway Research Centre at the same university. His research interests are maintenance engineering and management, reliability and maintainability analysis, product support, life cycle costing, risk analysis, system analysis, etc. He is also a member of the editorial board and a reviewer for many international journals. He has published more than 150 papers in international journals and conference proceedings.

Paper V

Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration

Patra, A. P., Söderholm, P. and Kumar, U. (2009). Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* 2009, 223 (F3), 285-293.

Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration

A P Patra*, **P Söderholm,** and **U Kumar**
Luleå Railway Research Center, Luleå University of Technology, Norbotten, Sweden

The manuscript was received on 5 May 2008 and was accepted after revision for publication on 17 October 2008.

DOI: 10.1243/09544097JRRT235

Abstract: Life-cycle cost (LCC) is used as a cost-effective decision support for maintenance of railway track infrastructure. However, a fair degree of uncertainty associated with the estimation of LCC is due to the statistical characteristics of reliability and maintainability parameters. This paper presents a methodology for estimation of uncertainty linked with LCC, by a combination of design of experiment and Monte Carlo simulation. The proposed methodology is illustrated by a case study of Banverket (Swedish National Rail Administration). The paper also includes developed maintenance cost models for track.

Keywords: reliability and maintainability, life-cycle cost, railway track, design of experiment, uncertainty

1 INTRODUCTION

Life-cycle cost (LCC) takes into account all costs associated with the life time of the system, such as operating costs, maintenance costs, energy costs, and taxes apart from capital costs. For many complex assets, the cost of maintenance plays an important role in the LCC analysis, especially for assets like track infrastructure, where the operation and maintenance phase comprises a major share of the system's life cycle. However, though most infrastructure managers today consider all the costs incurred by the system from conceptual design to disposal in their LCC calculations, there are still some issues associated with the correctness of these calculations. Some important issues are related to uncertainties in the LCC calculations.

Figure 1 illustrates two different levels of uncertainties associated with LCC of track infrastructure. Level I uncertainty is costs due to penalties imposed by traffic operators on the infrastructure manager due to such factors as train delay, traffic disruption, or derailment. These anomalies can be caused by planned or unplanned maintenance actions, but also by lack of necessary maintenance. Hence, the resulting costs are related to decisions about maintenance actions and can be estimated by probabilistic assessment of train delay, derailment, or traffic disruptions considering the technical and operational characteristics of the track, as well as the maintenance actions. Level I uncertainty can also be viewed as belonging to the external risk of the LCC analysis, where the costs should be included to make the LCC analysis more effective. However, there is also the level II uncertainty, which is the internal risk associated with LCC. The level II uncertainty pertains to the variable contribution to total LCC originating from the uncertainty in reliability and maintainability (R&M) parameters. However, the R&M parameters also indirectly impact the level I uncertainty. As conventional LCC analysis only considers point estimates of R&M parameters, it leads to an incorrect estimate of the LCC. To get a more correct estimate of the LCC, it is essential to also consider the interval estimate of the R&M parameters.

There is some research related to the stochastic nature of R&M parameters included in LCC estimation of railway infrastructure, see e.g. [**1**]. However, no published research about the estimation of the uncertainty in LCC of railway infrastructure has been found. Hence, this paper aims at describing a methodology that can be used for uncertainty estimation in railway infrastructure LCC.

**Corresponding author: Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Norbotten 97187, Sweden. email: ambika.patra@ltu.se*

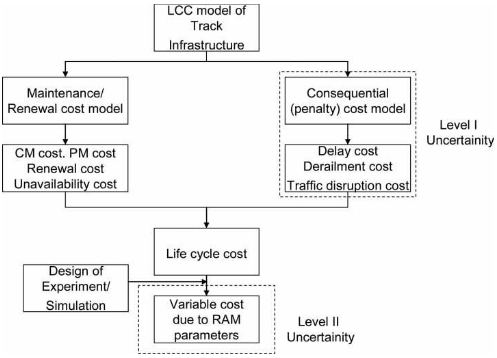


Fig. 1 Uncertainties involved with LCC modelling

2 MAINTENANCE COST MODELLING

LCC, which is generally modelled in the design phase, changes when the system enters into the operation and maintenance phase due to changes in stakeholder requirements, which makes the costs incurred during the operation and maintenance phase predominant. Maintenance costs of track infrastructure consist of preventive maintenance cost, renewal cost, and corrective maintenance cost (Table 1).

Maintenance costs are the most complex cost component of an asset during its life since maintenance is a long continuous process throughout the asset life. While the cost of any specified maintenance work on an asset can be comfortably estimated using engineering costing methodologies, estimating maintenance costs throughout the asset life is a much more sophisticated process. This is because the types of maintenance are dependent on many factors, of which the most important are asset deterioration rates, maintenance policy, and budget constraint. Maintenance schedules therefore need to be planned to enable the maintenance costs to be estimated. Maintenance costs of track must include:

(a) materials, equipment, and labour;
(b) condition monitoring and inspection;
(c) track possession time.

Table 1 Track maintenance at Banverket

Maintenance strategy	Maintenance action	Maintenance trigger
Preventive maintenance	Rail grinding Tamping Rail lubrication Ballast cleaning Track inspection	Time Condition Time Condition Time
Renewal (preventive maintenance)	Rail renewal Ballast renewal Sleeper renewal Fasteners renewal	Condition Condition Condition Condition
Corrective maintenance	Rail replacement	Failure

The maintenance process at Banverket (Swedish National Rail Administration) is divided into corrective and preventive maintenance, where the later is based on condition or time. The current strategy at Banverket is to minimize corrective maintenance and to change time-based maintenance to condition-based [**2**]. Table 1 shows the different track maintenance strategies and actions at Banverket.

In this paper, maintenance costs associated with track have been estimated separately for different curve radii as different curve radii experience different failure probabilities and magnitudes. In this paper, maintenance cost models have been developed with respect to the type of maintenance intervention summarized in Table 1. The maintenance costs have been determined as per the maintenance policy followed at Banverket. The track has been divided into different sets of curve radii (K), i.e. 0–300 m ($K = 1$), 300–450 m ($K = 2$), 450–600 m ($K = 3$), and so on. Curves with radius more than 2000 m have been considered as tangent track. The segmentations of the track have been done as per of availability of the track failure data.

A few things must be considered while performing the segmentations of a track section. The segmentation of the track must be done for a specific track section and should not be generalized. The segmentation of the track must done as per (a) the number of each individual curve existing in a track section and (b) the number of track failures occurring in each type of curve over a period of time. For example, if there are few curves of curve radii between 700 and 1000 m, it is safe to take 700–1000 m as one segment, whereas if there are a lot of curves existing of curve radii between 500 and 600 m the 500–600 m must be defined as a track segment. The same logic can be applied for the number of failures in different curve radii. If the numbers of curves as well as the numbers of failures are high in a particular segment, it can be still divided into further segments. The segmentation of track section should be specific for each studied region, as described above.

Different track maintenance and renewal costs are illustrated below.

2.1 Rail grinding cost

Grinding is the maintenance action done on the rail to control rolling contact fatigue defects. Cost due to rail grinding primarily depends on the periodicity of grinding and the number of grinding passes and is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((T_{\mathrm{g}_i} * C_{\mathrm{L}} * L_i * n_{\mathrm{g}_i}) + (C_{\mathrm{eg}} * T_{\mathrm{g}_i} * L_i * n_{\mathrm{g}_i})) * (m/m_{\mathrm{g}_i})}{(1+r)^j} \tag{1}$$

2.2 Track tamping cost

Tamping is the maintenance action done on the track to correct its alignment. Cost due to track tamping depends on the interval of tamping and is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((T_{\mathrm{ta}_i} * C_{\mathrm{L}} * L_i) + (C_{\mathrm{eta}} * T_{\mathrm{ta}_i} * L_i)) * (m/m_{\mathrm{ta}_i})}{(1+r)^j} \tag{2}$$

2.3 Rail lubrication cost

Lubrication is done on the rail to control rail wear. Cost due to lubrication depends on the number of lubricators in the curves and the cost to maintain each lubricator in terms of filling, which is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{(T_{\mathrm{clu}} * C_{\mathrm{L}} * n_{\mathrm{l}_i})}{(1+r)^j} \tag{3}$$

2.4 Ballast cleaning cost

Ballast cleaning is the maintenance action done to eliminate trapped water inside the ballast in order to restore the track quality and stiffness. Cost due to ballast cleaning primarily depends on the periodicity of ballast cleaning and is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((T_{\mathrm{b}_i} * C_{\mathrm{L}} * L_i) + (C_{\mathrm{eb}} * T_{\mathrm{b}_i} * L_i)) * (m/m_{\mathrm{b}_i})}{(1+r)^j} \tag{4}$$

2.5 Track inspection cost

Track inspection is done to detect flaws on the track that can lead to failures. The cost due to track inspection primarily depends on the interval of track inspection and is given by

$$\sum_{j=1}^{N-1}\frac{((T_{\mathrm{t}} * C_{\mathrm{L}} * L) + (C_{\mathrm{et}} * T_{\mathrm{t}} * L)) * (m/m_{\mathrm{t}})}{(1+r)^j} \tag{5}$$

2.6 Rail renewal cost

Rail renewal is done when the rail deterioration reaches maintenance or safety limits. The cost due to rail renewal is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((C_{\mathrm{r}} * L_i) + (T_{\mathrm{rr}_i} * C_{\mathrm{L}} * L_i) + (C_{\mathrm{err}} * T_{\mathrm{rr}_i} * L_i)) * (m/m_{\mathrm{rr}_i})}{(1+r)^j} \tag{6}$$

2.7 Ballast renewal cost

Ballast renewal is done when ballast deterioration reaches maintenance or safety limits. The cost due to ballast renewal is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((C_{\mathrm{b}} * L_i) + (T_{\mathrm{br}_i} * C_{\mathrm{L}} * L_i) + (C_{\mathrm{ebr}} * T_{\mathrm{br}_i} * L_i)) * (m/m_{\mathrm{br}_i})}{(1+r)^j} \tag{7}$$

2.8 Sleeper renewal cost

Sleeper renewal is done when the sleeper deterioration reaches maintenance or safety limits. The cost due to sleeper renewal is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((C_{\mathrm{s}} * L_i) + (T_{\mathrm{sr}_i} * C_{\mathrm{L}} * L_i) + (C_{\mathrm{esr}} * T_{\mathrm{sr}_i} * L_i)) * (m/m_{\mathrm{sr}_i})}{(1+r)^j} \tag{8}$$

2.9 Fastener renewal cost

Fastener renewal is done when the fastener deterioration reaches maintenance or safety limits. The cost due to fastener renewal is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((C_{\mathrm{f}} * L_i) + (T_{\mathrm{fr}_i} * C_{\mathrm{L}} * L_i) + (C_{\mathrm{efr}} * T_{\mathrm{fr}_i} * L_i)) * (m/m_{\mathrm{fr}_i})}{(1+r)^j} \tag{9}$$

2.10 Rail replacement cost

Rail replacement is done when rail breaks occur on the track. Cost due to rail break primarily depends on the probability of rail breaks and is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{N-1}\frac{((C_{\mathrm{r}} * L_{\mathrm{r}}) + (T_{\mathrm{rb}_i} * C_{\mathrm{L}}) + (C_{\mathrm{er}} * T_{\mathrm{rb}_i})) * (m/m_{\mathrm{rb}_i})}{(1+r)^j} \tag{10}$$

2.11 Track downtime cost

Downtime on the track occurs due to track possession for maintenance actions on the track. Train-free periods are usually used for planning maintenance actions, i.e. the hours between two consecutive trains. However, as the train-free periods are not long enough in most cases, this leads to train cancellations, train speed restrictions, etc., which imply penalties imposed on the infrastructure manager by the traffic operators. Preventive maintenance and renewal actions are usually planned well ahead so as not to affect the traffic. However, corrective maintenance on the track generally affects the train operation. In this case, rail breaks have been considered for corrective maintenance. Mean time to repair (MTTR) for rail break is given by

$$\frac{\sum_i f_{\mathrm{rb}_i} * T_{\mathrm{rb}_i}}{\sum_i f_{\mathrm{rb}_i}} \tag{11}$$

In this case, track possession time is calculated as the difference between MTTR and train-free period. Hence, the track downtime cost can be calculated by multiplying the track possession period with the penalty cost. Table 2 describes the R&M parameters associated with track maintenance.

3 UNCERTAINTY IN LCC

The statistical characteristics of R&M parameters contribute to uncertainty in LCC. The reason for this is that the times and conditions for these types of events are so complex that they cannot be predicted with a fair degree of accuracy. Therefore, it was decided to explore a methodology that combines the use of design of experiment (DoE) principles with Monte Carlo simulation to estimate the uncertainty involved with LCC. The area of DoE was developed in the twentieth century to increase the effectiveness and efficiency of experimentation. However, for experiments to be effective and lead to correct conclusions there are a number of requirements that should be fulfilled [**3**]. For example, the response must be measurable and be correlated to the purpose of the experiment. Furthermore, even though not an absolute necessity, the power of statistical operations will be greater if the response is continuous and preferably also normally distributed. The responses of this study are the point estimate for LCC of the track and its related uncertainty, which both are continuous, but not necessarily normally distributed. The following are valid for the present study.

1. The factors that are tested in the experiment are R&M-parameters, which all are continuous and numeric. They are also measurable, controllable, and deemed important for the selected responses.
2. The factors that are not under investigation can easily be held constant, since the study is analytical and not empirical. These factors are the cost factors not directly related to R&M. Hence, no randomization is considered necessary.

Since the study is analytical there are no major economical constraints. Hence, the design is mostly dependent upon the number of R&M parameters that are to be investigated. In order to fulfil the purpose of this study, a two-level factorial design is considered valuable. However, in order to reduce the number of runs, a fractional factorial design is considered sufficient. The analysis is supported by the software tool STATGRAPHICS, which provides suitable tables and graphs for presentation.

The probability distribution of LCC can be found by the use of Monte Carlo simulation. A Monte Carlo simulation is effectively a random number generator that creates values for each R&M parameter. Values are chosen within specified ranges of each parameter and with a frequency proportional to the shape of probability distribution associated with each R&M parameter. The proposed methodology helps in determining the variable costs associated in LCC estimation. These variable costs can be termed as the uncertainty in LCC estimation and are caused by the probabilistic nature of the R&M parameters. The LCC becomes more robust when these variable costs are included. Thus, it helps the decision makers to make more effective decisions on maintenance policy by considering LCC.

Table 2 R&M parameters associated with track maintenance

Maintenance actions	R&M parameters	
Rail grinding	T_{g_i}	m_{g_i}
Tamping	T_{ta_i}	m_{ta_i}
Rail lubrication	T_{lu}	
Ballast cleaning	T_{b_i}	m_{b_i}
Track inspection	T_{t}	
Rail renewal	T_{rr_i}	m_{rr_i}
Ballast renewal	T_{br_i}	m_{br_i}
Sleeper renewal	T_{sr_i}	m_{sr_i}
Fasteners renewal	T_{fr_i}	m_{fr_i}
Rail replacement	T_{rb_i}	m_{rb_i}
Downtime cost	T_{rb_i}	f_{rb_i}

4 CASE STUDY

The performed case study was on the iron ore line (Malmbanan) that runs from Luleå in Sweden to Narvik in Norway. The line allows 30 tonne axle load with mixed traffic. Data (Tables 3 and 4) was collected from Banverket's failure and maintenance databases (i.e. BIS, Bessy, and 0felia) that range from 1997 to 2006 with some data being collected from reference [**4**]. The study was performed on the rail replacement cost on high and low rails separately. Low rail denotes the inner

Table 3 TTF (rail break) data in MGT for curves of radius 450–600 m

High rail	Low rail
400	325
350	350
250	150
425	225
300	275
325	425
150	300
350	125
150	150
400	400
275	300
575	

Table 4 TTR (to correct rail break) data in minutes for curves of radius 450–600 m

High rail	Low rail
159	258
120	154
480	216
149	240
270	169
547	75
340	340
43	202
228	202
202	216
240	240
218	

rail (smaller radius) and high rail the outer rail (larger radius) in a curved track. The idea of separating high rail and low rail for cost estimation lies in the fact that they both have different failure deterioration due to quasi-static forces in the track curvatures.

The following assumptions were made after consultations with Banverket's track experts in the case study.

1. Average gross tonnage per year is assumed to be 25 million gross tonnes (MGT).
2. Life of track for LCC estimation is 600 MGT (24 years).
3. Discount rate is taken as 4 per cent.
4. Cost of BV50 rail (including neutralisation) is 1395 Swedish kronor (SEK)/m.
5. Average labour cost is 525 SEK/h. This includes the track worker cost, track welder cost, and inspection personnel cost.
6. Welding equipment cost is 60 SEK/h.
7. Average length of rail replacement due to rail break (L_r) is 8 m.

LCC analysis was done on curves of radius 450–600 m, with cost figures given in SEK. The time to failure (TTF) and time to repair (TTR) data obtained from the Banverket data base were analysed using probability distribution models. However, before fitting any distribution models to analyse the data, the TTF and TTR data sets were verified for (independent and identically distributed (IID) random variables) assumption using graphical method (Figs 2 to 5). This is important because if data is not independent or it has trend, then probability distribution models cannot be used for analysing the data set [**5**]. Such data sets can be modelled by the use of other non-stationary model such as power law process model, etc [**6**].

Table 5 shows the probability distribution of mean time to failure (MTTF) and MTTR for both high and low rails. The analysis was supported by the software tool Weibull++. MTTF was estimated by considering the failure events (time period to occurrence of rail break) and suspended events (no rail break has occurred) for the particular curvatures of the track. MTTR considered here comprises of the logistic time, welding time, and inspection time necessary to repair the rail breaks. A two-sided 90 per cent confidence level was considered for determining the upper limit, mean, and lower limit of MTTF and MTTR.

Table 6 shows the LCC estimation by considering DoE principles. The high and low rails were analysed separately, but followed the same design. The applied design was a screening, full factorial, two-level design with the two experimental factors MTTF and MTTR, i.e. a 2^2-design that requires four runs.

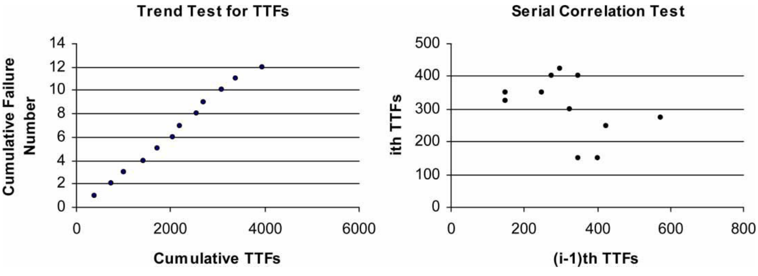


Fig. 2 Test for IID for TTFs of the high rail

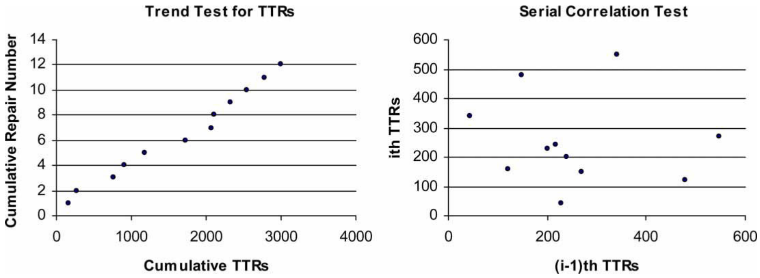


Fig. 3 Test for IID for TTRs of the high rail

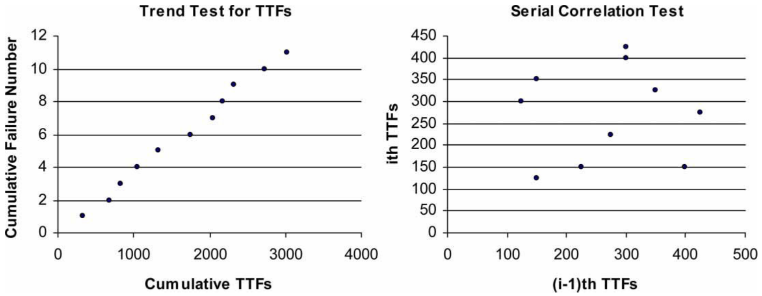


Fig. 4 Test for IID for TTFs of the low rail

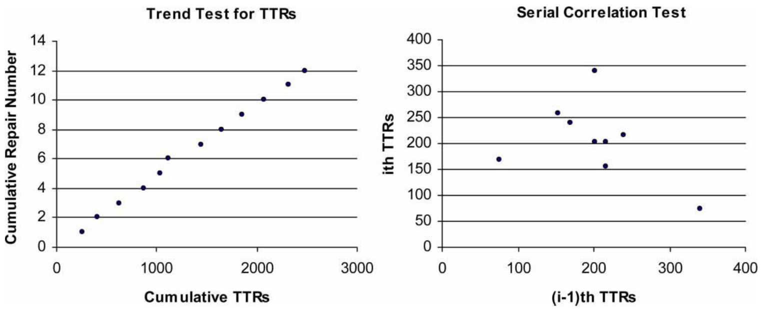


Fig. 5 Test for IID for TTRs of the low rail

These four runs were performed ten times (i.e. ten blocks with four runs in each), resulting in a total of 40 runs for high and low rails, respectively. The high and low levels for MTTF and MTTR were selected as the upper and lower limits of their distributions. The experiment contained two responses, i.e. the point estimate and the log(s^2) of LCC. The rationale for analysing log(s^2) is described in reference [**7**]. The input data was generated by Monte Carlo simulations. These data were entered into equation (10) and varied according to the experimental design summarized in Table 6.

Table 6 indicates that log(s^2) of LCC is quite stable for both high and low rails. However, changes in the levels of MTTF and MTTR do affect the variability in LCC. Since there is no interaction effect present (see Fig. 6 for example), the factors can be considered individually. An interaction between two factors means that the effects of either one cannot be judged independently. If there is an interaction between two factors, the effect of one factor on the response will depend on the setting of the other. In order to reduce the variability in LCC, one should look into the lowest value of log(s^2). The effects of variability in MTTF and MTTR

Table 5 MTTF and MTTR probability distributions for high and low rails

		High rail	Low rail
MTTF (MGT)	Probability distribution	Log normal ($\mu = 5.9933$, $\sigma = 0.2523$)	Weibull- 2 parameter ($\eta = 369.7161$, $\beta = 3.5315$)
	Upper limit	482.7	403.5
	Mean	413.6	332.8
	Lower limit	354.5	274.4
MTTR (hours)	Probability distribution	Weibull- 2 parameter ($\eta = 4.6972$, $\beta = 1.8871$)	Normal ($\mu = 3.4458$, $\sigma = 1.0296$)
	Upper limit	5.5	3.9
	Mean	4.2	3.4
	Lower limit	3.1	2.9

Table 6 LCC estimation with DoE principles

Type	MTTF	MTTR	LCC (average)	Log (s^2)
High rail	−1	−1	−1050.4	3.4198
	1	−1	−940.6	3.4784
	−1	1	−1086.6	3.4572
	1	1	−973.0	3.5139
Low rail	−1	−1	−1252.1	3.0577
	1	−1	−1113.3	3.0128
	−1	1	−1288.3	3.0824
	1	1	−1130.4	3.0230

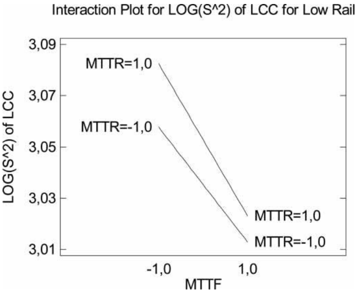


Fig. 6 Interaction plot showing variability in LCC for low rail

on the LCC of high and low rails are shown in Figs 7 and 8, respectively. The figures show the effect on LCC with increase of MTTF and MTTR values from low to high levels.

As shown in Figs 7 and 8, MTTF has a positive effect on LCC and MTTR has a negative effect. The magnitudes of the effects imply that the uncertainty in MTTF has more impact on the change in LCC than the uncertainty in MTTR. Two possible reasons for these differences in magnitudes are uncertainty levels in the parameters and given importance levels in the LCC formulation. The interaction between MTTF and MTTR is not significant in any of the cases.

Monte Carlo simulation was used to determine the probability distribution of LCC and estimate the associated variability cost. A two-sided 90 per cent confidence level was considered for this distribution. LCC figures were generated by combinations of upper, mean, and lower limits of MTTF and MTTR that were generated by Monte Carlo simulation. As shown in Table 7, the difference between upper and lower limits can provide the variability cost associated with LCC. The negative sign on the costs indicate that they were calculated as total present value.

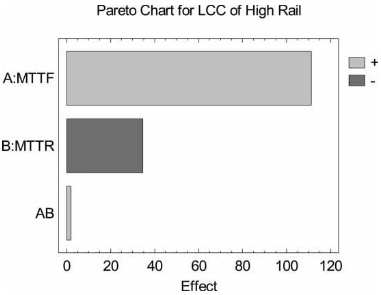


Fig. 7 Effect of MTTF and MTTR on uncertainty of LCC for high rail

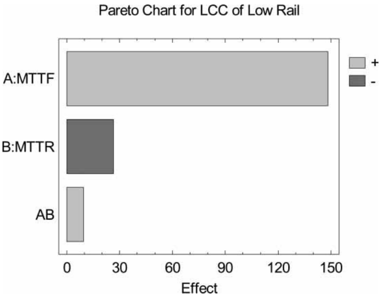


Fig. 8 Effect of MTTF and MTTR on uncertainty of LCC for low rail

Table 7 Simulated probability distribution of LCC

		High rail	Low rail
LCC (SEK)	Probability distribution	Weibull- 3 parameter Probability ($\eta = 219.6026$, $\beta = 3.0731$, $\gamma = 815.1878$)	Weibull- 3 parameter ($\eta = 170.7607$, $\beta = 2.2115$, $\gamma = 1049.3146$)
	Upper limit	−1024.9	−1214.7
	Mean	−1011.5	−1200.5
	Lower limit	−999.0	−1187.6

5 DISCUSSION AND CONCLUSIONS

LCC is being used as a tool to help in making effective maintenance decisions. However, there are various uncertainties associated with estimation of LCC. This paper presents level I and II uncertainties, out of which level II uncertainty has been dealt with. Level II uncertainty can be due to economic parameters, e.g. discounting rate, which has not been explored in this paper. The paper investigates more the uncertainties caused by technical parameters, i.e. R&M parameters. The uncertainty in R&M parameters exist because of their probabilistic nature, which contributes to the uncertainty in LCC estimation. For better estimation of uncertainty in LCC, this paper outlines a methodology based on a combination of Monte Carlo simulation and DoE. This combination gives a possibility to identify parameters that are influential on the LCC estimation and its variability. The proposed methodology can be used to estimate the uncertainty in LCC by considering uncertainties in all parameters simultaneously, in contrast to sensitivity analysis, where the parameters are considered one by one. Hence, the methodology can contribute to other research efforts, where traditional sensitivity analyses have been performed. The simulations are used to make the deterministic LCC equations probabilistic. DoE is applied to guide how the R&M parameters should be varied in a systematic way. The paper also illustrates cost models for different maintenance and renewal actions carried out on track. The uncertainty in LCC is presented as variable costs with associated distributions. When the variable costs are added to the LCC it becomes more robust. Hence, it helps the decision-makers to make more effective decisions about maintenance policy by considering LCC.

For further research, all the developed cost models for railway track can be combined into one model. The proposed methodology can then be applied to this new cost model. However, one major challenge will be to get relevant data to use as input to this sensitivity analysis. Another challenge is to deal with the large number of runs that will result by a full two-level factorial design, since the number of runs will double with each added parameter. However, this is not any major problem since DoE principles can be applied to reduce the number of runs by using fractional factorial designs that still will give valuable information.

REFERENCES

1 **Zhao, I., Chan, A. H. C., Roberts, C.,** and **Stirling, A. B.** Assessing the economic life of rail using a stochastic analysis of failures. *Proc. IMechE, Part F: J. Rail and Rapid Transit*, 2006, **220**(F2), 103–110.

2 **Karlsson, V.** An overall view of maintenance. *Eur. Railw. Rev.*, 2005, **11**(3), 11–17.

3 **Coleman, D. E.** and **Montgomery, D. C.** A systematic approach to planning for a designed industrial experiment. *Techometrics*, 1993, **35**, 1–27.

4 **Kumar, S.** *A study of the rail degradation process to predict rail breaks*. Licentiate Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden, vol. 73, 2006.

5 **Ascher, H.** and **Feingold, H.** *Repairable system reliability: modelling, inference, misconceptions and their causes*, 1984 (Marcel Dekker, New York).

6 **Kumar, U.** and **Klefsjö, B.** Reliability analysis of hydraulic system of LHD machines using power law process model. *Reliab. Eng. Syst. Saf.*, 1992, **35**, 217–224.

7 **Bisgaard, S.** and **Fuller, H. T.** Reducing variation with two-level factorial experiments. *Qual. Eng.*, 1995, **8**, 373–377.

APPENDIX

Notation

C_{b}	cost of ballast in SEK/km
C_{eb}	equipment cost for ballast cleaning in SEK/h
C_{ebr}	equipment cost for ballast renewal in SEK/h
C_{efr}	equipment cost for fastener renewal in SEK/h
C_{eg}	equipment cost for grinding in SEK/h
C_{er}	equipment cost to repair rail breaks in SEK/h
C_{err}	equipment cost for rail renewal in SEK/h
C_{esr}	equipment cost for sleeper renewal in SEK/h
C_{et}	equipment cost for track inspection in SEK/h
C_{eta}	equipment cost for tamping in SEK/h
C_{f}	cost of fasteners in SEK/km
C_{L}	average labour cost in Swedish Kroner (SEK)/h
C_{lu}	cost of lubrication material for each lubricator per year in SEK

C_r	cost of rail in SEK/km
C_s	cost of sleeper in SEK/km
f_{rb_i}	failure rate of rail (breaks) in the ith curve
K	class of curve radii
L	total length of track section in km
L_i	length of ith curve in km
L_r	average length of rail replacement due to rail break
m_{b_i}	interval for ballast cleaning for ith curve in MGT
m_{br_i}	interval for ballast renewal for ith curve in MGT
m_{fr_i}	interval for fastener renewal for ith curve in MGT
m_{g_i}	interval for grinding for ith curve in MGT
m_{rb_i}	mean time to rail breaks in ith curve in MGT
m_{rr_i}	interval for rail renewal for ith curve in MGT
m_{sr_i}	interval for sleeper renewal for ith curve in MGT
m_t	interval for track inspection in MGT
m_{ta_i}	interval for tamping for ith curve in MGT
M	gross tonnage per year in MGT
M	life period of track in MGT
n_{g_i}	number of grinding passes on ith curve
n_{l_i}	number of wayside lubricators in ith curve
N	life period of track (equivalent to M) in years
r	discount rate
T_{b_i}	mean time to clean ballast for ith curve in h/km
T_{br_i}	mean time for ballast renewal for ith curve in h/km
T_{fr_i}	mean time for fastener renewal for ith curve in h/km
T_{g_i}	mean time to grind for ith curve in h/km
T_lu	mean time to refill lubrication material for each lubricator in hour
T_{rb_i}	mean TTR rail break in ith curve in hour
T_{rr_i}	mean time for rail renewal for ith curve in h/km
T_{sr_i}	mean time for sleeper renewal for ith curve in h/km
T_t	mean time to inspect track in h/km
T_{ta_i}	mean time to tamp for ith curve in h/km