Department of Civil, Environmental and Natural Resources Engineering
Division of Operation, Maintenance and Acoustics

ISSN 1402-1544
ISBN 978-91-7583-320-0 (print)
Luleå University of Technology 2015

Stephen Mayowa Famurewa

Maintenance Analysis and Modelling for Enhanced Railway Infrastructure Capacity

Stephen Mayowa Famurewa
Maintenance analysis and modelling for enhanced railway infrastructure capacity

Stephen Mayowa Famurewa

Division of Operation and Maintenance Engineering
Luleå University of Technology
Luleå, Sweden
ABSTRACT

Railway transportation is a sustainable mode of transportation for reasons of safety, cost, carbon emission and energy requirements. It has a notable role in economic expansion in terms of passenger and freight services. In recent years, there has been a continuous demand to increase the competitiveness of railway transport via quantity and quality of service delivered. For instance there is a growing need to shift a substantial volume of freight and passenger traffic to rail.

To meet the demand for enhanced railway infrastructure capacity, large modification of the infrastructure, improvement of traffic planning process and improvement of maintenance and renewal process are required. The obvious solution would be capital expansion of infrastructure but this is a long-term cost-intensive approach for improving railway transport performance. This, therefore makes successive improvement of maintenance and renewal (M&R) process an ideal and feasible way of improving availability, capacity and service quality of existing railway infrastructure. This thesis addresses improvements in maintenance to enhance capacity and service quality through systematic maintenance analysis for effective planning and maintenance optimisation for efficient scheduling.

This thesis is divided into two parts: the first part deals with maintenance analysis and the second addresses maintenance optimisation. Both parts are aimed at enhancing maintenance effectiveness by improving track possession utilisation and infrastructure integrity. The first part suggests assessment and analysis methods to support continuous improvement of railway infrastructure performance. It entails the use of historical operation and maintenance data to identify, improve and eliminate weak links and bottlenecks. The second part deals with planning and scheduling of maintenance tasks from condition deterioration viewpoint. This part uses infrastructure condition data with model driven approaches to schedule maintenance tasks with the aim of ensuring efficient use of track possession time and maximisation of availability and capacity.

First, a fuzzy inference system is developed for computing the integrity index or composite indicator to relate maintenance functions to capacity situation. This is a good measure of the M&R need on a line as imposed by operational profile, capacity consumption and adopted maintenance strategy. It provides additional information that can be used to support high level M&R decisions for enhanced capacity. Second, risk matrix and an adapted criticality analysis method are proposed for identifying weak links and critical assemblies/items that are bottlenecks limiting operational capacity and service
quality. The focus is to address the problem of train mission interruption and reduced operational capacity. A pertinent result is classification of railway zones into different risk categories and a hierarchical list of improvement for the lower-level systems.

Third, a methodology is developed and demonstrated to quantify maintenance needs through deterioration modelling and to optimally allocate possession time for remedial actions on track. A case study of geometry maintenance is used to demonstrate the approach. The approach suggests a practical tamping plan with optimum allocation of track possession time, while track geometry quality is retained within specified limits. The methodology is extended to stochastic simulation of track geometry quality and integrated into a possession scheduling routine. The outcome of the proposed approach demonstrates that optimisation of tamping cycle length and shift duration, as well as tamping process improvement present opportunities for improved utilisation of possession time. Fourth, a short term maintenance scheduling model is developed to efficiently use available train-free periods for repair of inspection remarks such that availability and capacity are optimised. This model supports efficient scheduling of maintenance works that are not accommodated in the long-term maintenance plan. The outcome shows that an effective inspection plan and efficient scheduling model can be integrated to reduce capacity loss due to infrastructure condition.

Finally, the maintenance analysis methods and decision support models presented in this thesis are practical and feasible short-term plans for making maintenance more effective to enhance railway infrastructure availability, capacity and service quality.

KEYWORDS: maintenance improvement, railway infrastructure, availability, capacity, quality of service, bottlenecks, track possession time, tamping, optimisation, maintenance performance indicators, planning and scheduling, asset integrity, inspection
ACKNOWLEDGMENT

The research work presented herein was carried out between Aug. 2010 and Feb. 2015 at the division of Operation and Maintenance Engineering, Luleå University of Technology and Luleå Railway Research Centre (JVTC). The interest and financial support of the Swedish Transport Administration (Trafikverket) has been the means to this end.

Foremost, I would like to express my profound gratitude to my supervisor Prof. Uday Kumar who placed great confidence in my capability to undergo this research work. His mentorship and guidance have contributed greatly to this accomplishment. I would also like to appreciate my assistant supervisor Dr. Matti Rantatalo for sharing his time and experience with me. I am thankful to Dr. Arne Nissen, Prof P-O Larsson Kräik, Lars Wikberg, Matthias Asplund, Anders Backman and other personnel of Trafikverket for providing the required data and information, and for sharing their experiences with me.

I wish to appreciate Prof. Jan Lundberg, Dr. Behzad Ghodrati, Dr. Alireza Ahmadi, Dr. Aditya Parida, Dr Phillip Tretten and other faculty members for their support during this research work. I would like to say thank you to Christer Stenström, Amparo Morant, Hussan Hamodi and all other colleagues at the division of operation and maintenance engineering. I would like to thank Dr. Xin Tao and Dr. Musa Idris for their fruitful discussions. I would like to acknowledge Selin Jönsson in place of her late father Jens, who was a close friend.

Great thanks are due to my wife Abiola, and our children Jeremiah and Joanna for their understanding and unflinching support. I am very grateful to my parents Mr and Mrs Famurewa, and my brothers Sunday and Ayodele and their families for their invaluable roles in this achievement. I am using this opportunity to appreciate the wonderful brethren who have supported us with prayers and encouraging words, Mr. Obudulu Ogbonna, Dr. Awe Samuel, Dr. Leif Berglund, Mr. Fadiola Yinka, James Akinola, Andrews Omari, Esi Nunoo, Gbenga Omoniyi, Niyi Abiri and their families.

Finally, all of my help comes from God, the eternal creator and giver of wisdom and grace for this time and in the eternal life.

Stephen Mayowa Famurewa
February 2014
Luleå, Sweden
LIST OF APPENDED PAPERS

PAPER 1

PAPER 2

PAPER 3

PAPER 4

PAPER 5

LIST OF OTHER PAPERS


DISTRIBUTION OF WORKS

The works carried out in the appended papers have been contributed by the thesis author as well as the other co-authors. The contributions of the author and the co-authors in the papers are highlighted in the table below:

1. Idea conception
2. Method and technique selection
3. Data compilation and processing
4. Model building
5. Results and discussions
6. Article writing
7. Review

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephen M. Famurewa</td>
<td>1–6</td>
<td>1–6</td>
<td>1–6</td>
<td>1–6</td>
<td>1–6</td>
</tr>
<tr>
<td>Matti Rantatalo</td>
<td>5, 7</td>
<td>5, 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uday Kumar</td>
<td>7</td>
<td>7</td>
<td>1,7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Arne Nissen</td>
<td></td>
<td></td>
<td>3, 7</td>
<td>3, 7</td>
<td></td>
</tr>
<tr>
<td>Xin Tao</td>
<td></td>
<td></td>
<td>4, 5, 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matthias Asplund</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulla Juntti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aditya Parida</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christer Stenström</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diego Galar</td>
<td>2, 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PART I

CHAPTER 1 – INTRODUCTION

1.1 Background .................................... 3
1.2 Problem Statement ................................ 5
1.3 Research Purpose and Objectives .................. 6
1.4 Research Questions ............................ 6
1.5 Research Scope and Limitation .................... 8
1.6 Thesis Outline .................................. 9

CHAPTER 2 – RESEARCH METHODOLOGY

2.1 Research approach ............................ 11
2.2 Research process .............................. 12
2.3 Research method ............................... 14
   2.3.1 Exploratory methods ....................... 14
   2.3.2 Data Collection method ..................... 16
   2.3.3 Data analysis and modelling techniques .... 17

CHAPTER 3 – RAILWAY INFRASTRUCTURE MAINTENANCE

3.1 Railway infrastructure capacity & service quality .... 21
   3.1.1 Capacity .................................. 21
   3.1.2 Quality of service .......................... 23
   3.1.3 Capacity and QoS enhancement plans ............ 23
3.2 Maintenance improvement ........................ 25
   3.2.1 Efficient maintenance execution process .......... 25
   3.2.2 Effective maintenance analysis .................. 28
   3.2.3 Optimum maintenance possession scheduling .... 29
3.3 Maintenance optimisation theory ................... 30
3.4 Review of maintenance schedule models ............. 32

CHAPTER 4 – DATA ANALYSIS AND MODEL FORMULATION

4.1 Development of aggregated health index ............ 35
4.2 Constraint identification ........................ 39
4.3 Deterioration based maintenance scheduling ........ 42
4.4 Possession optimisation for deterioration based maintenance .... 44
4.5 Potential failure based maintenance scheduling .... 49
Chapter 5 – RESULTS AND DISCUSSION

5.1 Maintenance analysis ........................................ 55
  5.1.1 Development of aggregated health index .............. 55
  5.1.2 Constraint identification with adapted criticality analysis .... 58

5.2 Maintenance optimisation .................................... 62
  5.2.1 Deterioration based maintenance scheduling ............ 63
  5.2.2 Potential failure based maintenance scheduling .......... 71

Chapter 6 – CONCLUSIONS

6.1 Key findings ............................................. 75
6.2 Contributions ............................................ 77
6.3 Suggested future work .................................... 77

References ..................................................... 79

PART II ......................................................... 87
PART I
THESIS SUMMARY
1.1 Background

Railway transportation is an important mode of transportation for reasons of safety, cost, carbon emissions and energy requirements. It is a sustainable mode of transportation that can support the expansion of industrial activities and people’s mobility through freight and passenger services (Nyström, 2008, Patra et al., 2010). The present state of existing railway infrastructure and the need to shift a substantial volume of freight and passenger traffic to railway are issues that require attention in the transportation industry (European Commission, 2011).

The operational capacity of a given railway infrastructure depends on its technical state or quality and the way it is utilised (UIC, 2004, Patra et al., 2010, Tzanakakis, 2013). Capacity utilisation is largely influenced by market requirements, traffic planning, regulations and other operational requirements. An important aspect in railway infrastructure maintenance is the dependence between operational capacity (with associated Quality of Service QoS) and infrastructure condition. High operational capacity and expected QoS is guaranteed when railway infrastructure is in a good state with high quality. Conversely, increase in capacity or traffic loads leads to rapid quality deterioration of infrastructure and deformation of its components. This consequently leads to higher M&R needs and more requests for track possession that eventually reduces the operational capacity (Tzanakakis, 2013).

A holistic view of railway infrastructure capacity and its influencing factors is presented in Figure 1.1 with emphasis on infrastructure quality. The upper part of the figure gives a quick background into the study, it shows the dependence of infrastructure quality on installed quality, operating conditions, environmental conditions and M&R conditions.

Historical record shows that there has been 15% and 28% increases in rail freight tonnage-
kilometre and passenger-kilometre, respectively over 1990—2007 in EU15 countries (Menaz and Whiteing, 2010). In Sweden, the average annual growth of traffic on the railway network from 1960 to 2010 was 1.1%, a figure above the corresponding increase in road and water-borne traffic (Trafikverket, 2012b). In addition, a minimum annual increase of 1% in traffic tonnage is anticipated up to 2050 (Trafikverket, 2012b). The increased traffic volume is transported on upgraded or same infrastructure with few instances of new track constructions. For example, the length of operated track in Sweden has remained significantly unchanged over the past decade (Trafikanalys, 2011), and the traffic volume in terms of freight tonnage-kilometre and passenger-kilometre have increased by 17% and 28% respectively over the same period (Trafikanalys, 2011).

However, the rises in traffic volume have affected to a great extent the dependability performance of the infrastructure and the delivered operational performance or service quality (Nyström, 2008, Åhrén, 2008). Previous study has shown that infrastructure failure contributes significantly to reduced achievable capacity and service quality on the Swedish railway network (Nyström, 2008). For instance, the annual frequency of traffic interrupting failure is within the range of 143—195 and 74—131 per billion tonnage kilometre between 2010 and 2014 on the entire Swedish network and iron ore line, respectively. The reported urgent potential failure frequency or inspection remarks is significant as well. Based on historical records, the total reported delay hour has dropped, but the contribution of infrastructure failure is substantial and increasing in terms of both absolute values and proportions as shown in Figure 1.2. This can be connected with high
traffic volume, design infrastructure conditions and inadequate maintenance.

![Graph showing contribution of infrastructure-related failure to total delay on Swedish network and iron ore line]

Figure 1.2: Contribution of infrastructure-related failure to total delay on Swedish network and iron ore line

To achieve the designed traffic quantity and quality with the existing railway infrastructure, large modification of (new investment) the infrastructure or improvement of relevant processes such as maintenance and renewal process is required. In recent years, several studies have been conducted for improving the competitiveness of railway transport through capacity and service quality improvement. These extend from improvement of rail services, rail management systems and rail technology (Menaz and Whiteing, 2010). An obvious solution to the transport quality and quantity challenge would be capital expansion of infrastructure, but this is a long-term cost-intensive solution for improving railway transport (Abril et al., 2008). Therefore successive improvement of maintenance and renewal (M&R) process becomes a cost effective and feasible way of improving capacity and delivered service quality with existing railway infrastructure.

1.2 Problem Statement

The behaviour of tracks and other railway systems under increased loading and inadequate maintenance is different from that of other engineering assets. Uniquely, the effect of inadequate maintenance on railway system performance takes a relatively long time before it is apparent. The effects of such extended maintenance inadequacy include high infrastructure unreliability, irreversible and rapid loss of quality and frequent interruption of train mission. In the absence of appropriate infrastructure modification or M&R improvements, the hampered performance will result into significant reduction of the achievable capacity and service quality. Therefore, a major concern in railway mainte-
nance engineering is reviewing current maintenance practices for enhancing capacity and service quality. Some of the problems identified in an initial exploratory study include the following:

- Lack of data-driven tools to support key maintenance decisions for improving infrastructure performance based on their health condition.
- Lack of model-based planning and optimisation tools to support decisions on maintenance resource allocation and utilisation of given track possession.
- Lack of quantitative methods to prioritise maintenance tasks based on their impact on achievable capacity and service quality.

Maintenance improvement is a cost effective and feasible way of addressing the above mentioned issues to enhance the performance of existing railway infrastructure.

1.3 Research Purpose and Objectives

The purpose of this study is to develop decision support models to enhance railway infrastructure capacity and service quality by improving maintenance performance and utilisation of maintenance possession time. This study is aimed at using data-driven methods to support maintenance decisions at both the tactical and the operational levels within an infrastructure manager’s organisation. The objectives of the research work in precise terms are listed below:

- Map current railway maintenance practices to identify deficiencies in relation to expected performance levels and suggest improvement potentials.
- Develop a method for aggregating information about infrastructure condition to support maintenance decisions for enhanced capacity.
- Study maintenance analysis methods used within railway and other related industries for continuous improvement of maintenance process.
- Develop a decision support tool for augmented utilisation of track possession time and efficient maintenance.

1.4 Research Questions

The following research questions are formulated to achieve the purpose of this study, as well as to serve as cardinal points around which the research is centred.

*RQ1* How can information about infrastructure condition be aggregated to support maintenance decisions for enhanced capacity?
1.4. Research Questions

RQ2 Which maintenance analysis method is suitable to identify the "weakest links" on a railway section from viewpoints of capacity and punctuality?

RQ3 How can track possession time be optimised using the prognostic maintenance approach and deterioration based scheduling models?

RQ4 How can a data-driven approach be used for efficient scheduling of maintenance tasks into available train-free windows?

A tabular presentation showing the connection between the appended papers and the research questions is given in Table 1.1, and thereafter, a brief summary of the appended papers is given.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQ 2</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQ 3</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RQ 4</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Paper 1 suggests a method of computing a composite performance indicator for infrastructure management. The main issue addressed in the paper is the quantification of the integrity of railway infrastructure under certain traffic profiles. Such an integrity index provides an additional perspective of capacity limitation on a track section, helps to relate M&R to the capacity condition of a network and facilitates effective maintenance decision making.

Paper 2 presents the application of risk matrix as a maintenance analysis method for identifying of track zones that are bottlenecks limiting operational capacity and quality. It also presents a criticality analysis method to create a hierarchical improvement list for addressing these constraints. Doing so will facilitate maintenance decisions and continuous improvement.

Paper 3 presents a methodology to quantify maintenance need, optimally allocate track possession time and effectively use the allocated time. A case study of tamping action is used to demonstrate the approach.

Paper 4 presents a data-driven scheduling model for effective track possession management and availability maximisation. The stochastic degradation model and the formulated schedule optimisation problem are integrated to support planning and scheduling of track geometry maintenance to reduce track possession time. The three main objectives of the study are: determination of optimum shift duration, optimisation of tamping cycle length, improvement of tamping process for augmented utilisation of track possession
Paper 5 presents a data-driven scheduling method to efficiently use available train-free periods for restoring potential failure remarks such that availability and capacity are optimised. A short term maintenance scheduling problem was formulated to support the effective and efficient scheduling of maintenance works that are not accommodated in the long-term maintenance plan. The formulated problem focused on reducing the sum of maintenance cost, possession cost, window start-up cost and penalty cost resulting from delayed restoration.

Figure 1.3 shows the contributions of the appended papers to specific aspects of the overall research purpose. Simply put, the research has been divided into two parts, namely, maintenance analysis and maintenance optimisation. The first part deals with performance monitoring and evaluation to identify and categorise capacity constraints as well as to suggest action list for enhancing service quality and quantity. The second part focuses on maintenance optimisation for efficient use of track possession time and employs the prognostic deterioration model and data-driven scheduling models.

![Figure 1.3: Research structure showing contribution of appended papers to the overall research purpose](image)

1.5 Research Scope and Limitation

This study covers different aspects of the railway infrastructure maintenance process, especially planning and scheduling, analysis and assessment and improvement as related to track possession time and infrastructure performance. The research covers literature
review, survey, exploratory data analysis, model development and case studies for improvement of maintenance function with focus on infrastructure availability, capacity and service quality.

The iron ore line ("Malmbanan") of the Swedish Transport Administration is used for demonstration of the models and methods proposed in this study owing to data availability, stakeholder’s interest, technical support, and other considerations. However, the proposed maintenance principles can be extended to metropolitan regions and other routes of interest. Instances of train cancellation due to infrastructure failure could not be used in the maintenance task analysis owing to unavailability of reliable information about cancelled trains. The analysis methods and prediction models have only considered the scenario of nominal usage of the infrastructure and not extreme traffic characteristics such as speed and axle loads. Extreme weather conditions and other boundary conditions are not modelled explicitly in the study. New innovations for high speed maintenance and inspections are not considered in this study since it is limited to existing infrastructure.

1.6 Thesis Outline

This thesis consists of the research summary, and five appended journal papers. The thesis summary consists of six chapters that describe the relevant theoretical background to this research work, methodology, literature review, analysis and modelling techniques, results and discussions, and conclusions of the work.

The first chapter herein introduces the research with the problem statement and pedagogic descriptions of the research purpose and questions. This provides background information for understanding the relevance of this research and its contextual perspective. The second chapter describes the scientific and systematic approach followed in this study. It explains the various stages in the research process and provides the rationale for selecting the method used herein. The third chapter presents a literature review on railway infrastructure maintenance with a focus on maintenance improvement for capacity and service quality enhancement. The fourth chapter presents the frameworks of each of the papers appended to the thesis, including the data analysis and modelling techniques. The fifth chapter presents the results and discussions of the study under two broad divisions: maintenance analysis for effective planning and maintenance optimisation for efficient scheduling. Finally, the findings, contributions of the research work and suggestions for future work are given in the sixth chapter.
Chapter 2

RESEARCH METHODOLOGY

An investigation into a particular field of knowledge requires a systematic and scientific approach to establish a fact or principle. Such systematised effort in a scientific way is called research and can be described using the methodology and the method deployed in the research process. The research methodology is the science of how research is done scientifically and it emphasises the various steps considered in a research process to obtain insights or solutions to a set problem along with the logic behind said steps (Kothari, 2009). Engineering research requires implementation of appropriate methodologies, methods and procedures to solve engineering problems (Thiel, 2014). For instance, the selection and use of certain methods for experimentations, testing, observation, data recording, data analysis, as instruments for performing research operations requires methodological motivation (Kothari, 2009).

2.1 Research approach

This work is an applied research within railway engineering with focus on infrastructure maintenance and has utilised relevant statistical data analysis, operations research techniques and mathematical models. Quantitative research is based on the measurement of quantities and is applicable to phenomena that can be expressed quantitatively (Kothari, 2009). Quantitative approach was chosen in this study considering the problem statement, possibility of repeatable evaluation of the result, peculiarity of railway operations, availability of data, and interest of stakeholders. However, a qualitative approach was employed at the initial research stage to obtain direction and relevant problem description from experts and interested stakeholders given the applied nature of this work.

The research methodology employed is a combination of exploratory and descriptive research types. An exploratory research approach was used in the initial stage to create opportunities for considering different aspects of the problem. The aim was to cover different interesting aspects of data-driven decision management from the viewpoints of improving railway service quality and capacity. This initial stage helped to identify
problems related to maintenance and time on track, and support the formulation of the research questions within the scope of the available resources and stakeholder interest. In addition, it helped to identify the types of data available and the possibility of collecting new data.

In the second stage, quantitative descriptive research was employed using field-setting data collection with statistical analysis and other advanced analytical methods. The choice of a descriptive approach was based on the fact that historical failure and maintenance data are readily available and there is a need to depict accurately the state and characteristics of the infrastructure in the past and suggest improvements. The quantitative analysis aspect has two dimensions: diagnostic - to provide a comprehensive picture of historical trends with existing information and prognostics - to project into the future for better maintenance planning and decision making. This research approach and methods are tailored towards the purpose of the research work: to enhance railway infrastructure service quality and capacity by improving the utilisation of track possession time and resource allocation.

Some limitations of the selected research methodology and methods were mentioned in the introduction. Furthermore, some of the results are specific to the case study and cannot be generalised for the entire network. The unavailability of some other essential data led to the use of expert judgements and opinions, which are subjective and limited to the experts’ experience and exposure. Overall, the methods and approaches employed herein are unique and can be repeated or adapted for implementation.

2.2 Research process

Engineering research requires several actions and efforts that are expected to be coordinated systematically and logically in a process. In essence, a research process presents a series of essential action steps, along with the interconnections and sequencing of these steps, for effectively achieving the aims of a research project (Kothari, 2009). Figure 2.1 presents a brief overview of the research process used in this study. The connections between the different methods and their inclusive stages and techniques are shown in the figure.

The flow shows the links among closely related activities carried out over the course of this research work. There are eight distinct activities from the initial stage to the final stage. These activities are connected in one way with a kind of overlap in time sequence. For scientific presentation of the research process, the eight distinct activities were clustered into six stages.

The first stage involved a survey and literature review to capture the big picture and insight into subject and problem. This led to the second stage wherein the research objective and question were formulated with clearer perspective. The third stage involved
2.2. Research process

Figure 2.1: Research process

case study selection and data collection. Using a deliberate or purposive selection approach, the iron ore line in the northern part of Sweden was selected as the case study in this research. The fourth stage comprised data cleaning and processing using standards, expert opinions and some preliminary data analysis procedures. Data analysis and model development using statistical techniques and operations research methods constituted the fifth stage.
Research methodology

For each research question, the technique used and the model development procedure differed, as explained in subsequent sections. Techniques were selected based on their appropriateness, available data, literature survey and the author’s viewpoint. The final stage involved reporting and presentation of the results and output as journal articles and thesis.

2.3 Research method

This research study is applied and decision oriented with focus on pertinent problems within the railway industry. To arrive at practical results and solutions to the problem, some research methods and techniques have been used. These methods can be divided into three categories: exploratory, data collection, and analysis and modelling methods.

2.3.1 Exploratory methods

Literature study and survey were employed to explore the research subject in order to gain familiarity of current practices and the state of the art as relevant to the research subject and the study environment.

Literature review

Literature survey of previous works related to this study was conducted. Among the surveyed literature were conference papers, journal publications, PhD thesis, technical reports and EU projects related to railway infrastructure maintenance. The most remarkable search results were obtained with the following keywords:

− Railway infrastructure maintenance
− Track possession assignment
− Track degradation
− Maintenance optimization
− Maintenance planning and scheduling
− Capacity enhancement plans
− Railway infrastructure management
− Maintenance performance measurement in the railway industries
− Continuous improvement.
2.3. Research method

Survey

This formed the qualitative aspect of this research and fulfilled the need to gather vital information for mapping and describing the maintenance practices and principles followed in Trafikverket. This helped to identify factors that influence the technical performance of the railway network. A questionnaire designed to consider the three main categories of parameters influencing the technical performance of railway systems was used as the survey instrument. The categories are detailed in Figure 2.2, which was taken from the standard — RAMS specification for railway infrastructure (CENELEC EN 50126, 1999). Attention was on maintenance conditions because the research is focused on maintenance improvement through systematic maintenance analysis for effective planning and maintenance optimisation with the aim of achieving efficient scheduling.

In short, the different interests addressed in the questionnaire are maintenance planning, scheduling and execution, logistic support, condition monitoring, maintenance contracting, maintenance management systems, conventional practices and other external factors. The effects of these factors and other maintenance process steps on network capacity were investigated and potential solutions were gathered from the interviewees. The targeted questionnaire respondents are personnel of the maintenance contractors, Trafikverket and train operators with relevant experience and job responsibilities.

![Figure 2.2: Factors influencing RAMS parameter (CENELEC EN 50126, 1999)](image-url)
2.3.2 Data Collection method

Data can be defined as fact that can be communicated and stored (Spender, 1996). They are collected carefully by following acceptable procedures and in specific environments, such as libraries, laboratoritories or fields. An essential activity in any research process is the gathering of relevant data which give raw information about the process or phenomena under study. There are two categories of data collection: primary data collection through surveys, experiments, etc., and secondary data collection through compilation, querying and organisation of primary data (Kothari, 2009). The author basically employed secondary data collection in this research and also complemented the collected data with expert opinion and cost data provided by Swedish Transport Administration personnel. The primary data stored in the asset information databases of the IM were originally collected using feedback, field reports (work orders) and observations and then stored in an asset management system. In some instances, data have been collected using mechanical and electronic devices. The device ranged from simple devices to complex machines, such as accelerometer, hand-held rail profile measurement device, trolley based eddy current measurement and track recording car. The author was involved in secondary data compilation and querying from existing databases. The data from railway asset information databases used in this research can be grouped into two categories: (1) operation and maintenance data recorded in different databases, and (2) condition monitoring data (inspection remarks and track geometry data) recorded in separate data management systems.

Operation and maintenance data

Reports, records, observations and relevant incidences from the operation and maintenance of railways are collected daily by the infrastructure manager. The majority of incidents and events stored are collected on the field by train operators, maintenance personnel, individuals and other concerned stakeholders. These data are basically operation and maintenance data recorded in different data management systems with little or no intelligent integration. The data collected as maintenance data are recorded in the form of maintenance work orders, while those collected as operational data are handled as train movement and operational incidents. The data reported in disparate databases include the following: delay or punctuality data, failure data, train position data, maintenance report, inspection remarks, inventory and other relevant asset information data.

The data used in the study have been primarily collected between the years 2007 and 2013. The peculiarities and quality requirements of each analysis or model were considered in the eventual compilation and processing of the collected data. Further explanation of the data and other details can be found in each paper. Importantly national standards and handbooks (Trafikverket, 2007a, 2011a,b, 2012a) were used during data compilation for gaining contextual understanding and deducing relationships. Figure 2.3 shows the names of the data sources and their respective data types as used in this research work. It should be noted that some other quantitative data were obtained from the literature.
2.3. Research method

Inspection records and track geometry data

Examination of a system by observing, testing or measuring its characteristic condition parameters at predetermined intervals is an essential aspect of operation and maintenance. For instance, inspection could be visual or non-destructive testing such as ultrasonic inspection, eddy current check, track geometry measurement and laser inspections. Generally, inspection and condition monitoring of railways are based on the traffic volume and the line speed. For inspection (visual or mechanised), usually reports are generated as inspection remarks after the completion of inspection procedures. These remarks are classified into priority levels based on the seriousness of the observation. For example, within Swedish Transport Administration, the remarks associated with safety- or maintenance-based inspections are grouped into different priority class for action plans, these include: acute, week, month and next inspection priority class (Trafikverket, 2005a, b). These types of remarks and reports were used as potential failure data in this thesis and the appended papers.

In addition, track geometry data is another condition monitoring data that was used in the study. Track geometry monitoring is an important element of any effective preventive maintenance programme. It is needed for planning track geometry intervention strategy (e.g. tamping) that is optimum in the allocation and utilisation of track possession time. In addition it provides useful information to avoid early or too frequent tamping, which degrades the ballast condition, and simultaneously guide against late intervention which can result in a temporary speed restriction or safety issues. For the case study, track geometry monitoring is done three to six times a year, generally between April and October using STRIX or IMV100 measurement trains. Several geometry parameters are recorded by the measurement trains, but only the standard deviation of the longitudinal level over each 200 m of track is used for the geometry quality prognosis and maintenance optimisation. The selection of short-wave longitudinal level data for modelling track geometry condition was based on reviewed academic works, standards and common practices among infrastructure managers (Andrade and Teixeira, 2011, Andrews et al., 2014, CEN EN 13848-1, 2008, Lichtberger, 2005, CEN prEN-13848-6, 2012, UIC, 2008, Vale et al., 2012). The geometry data used in the study were collected between the years 2007 and 2013. For appropriate interpretation and utilisation of the data, some handbooks, standards and expert clarification were employed during data compilation and processing. The resources include: (CEN EN 13848-1, 2008, CEN EN 13306, 2010, Trafikverket, 2005a, b, 2012a, 2014).

2.3.3 Data analysis and modelling techniques

The methods for and process of transforming data into useful information for decision support constitute an essential aspect of scientific research. These methods are used for
the discovery of knowledge from data and for objectively explaining phenomena with patterns considered to be valid, useful, novel, or understandable. Maintenance data, traffic data and other facts collected are analysed further to establish relationships and draw useful information that would serve as a knowledge base for decision making. In general, three main activities were performed in relation to analysis and model development in this research:

- Data checking and cleaning: As mentioned earlier, the author was involved actively in secondary data collection; thus, there is need for adequate scrutiny. This is to confirm the suitability of the data, as well as the reliability, adequacy and source of the data in the context of the problem at hand. Thereafter, the data are cleaned by deleting data that were evidently incorrect and checking the reason for outliers to avoid missing data.

- Preliminary data analysis: A preliminary analysis was performed to carefully check the appropriateness of the data for analysis and modelling in the context of the objectives of the study. The aims of preliminary data analysis included: description of the key features of the data, providing an overview of the information content of the data, preparation of the data in a format useful for further analysis.

*Figure 2.3: Data collection procedure for the study*
2.3. Research method

- Detailed analysis and model development: This is the core of the project, and it was tailored carefully and logically to solve the research problem. Relevant data analysis and modelling methods in the context of this research include statistical techniques, operations research techniques and mathematical models. These were used for establishing relationships between the data and for simplified representation of reality to solve the problem at hand. Different analysis and model development methods were used in each appended paper based on their peculiar requirements and focus. The methods and the adopted techniques are further described in chapter 4.
RAILWAY INFRASTRUCTURE MAINTENANCE

Railway infrastructure management includes the following major responsibilities: management of infrastructure capacity, management of train traffic control on the infrastructure and management of maintenance and renewal functions (Alexandersson and Hultén, 2008). The latter responsibility is very pivotal because it ensures the quality, safety, reliability, maintainability and availability of the infrastructure, which is prerequisite for capacity allocation and train traffic management. In other words, maintenance of existing railway infrastructure affects the achievable capacity and the delivered quality of service on a network. Therefore, one of the main objectives of railway infrastructure maintenance is to increase the achievable capacity (or support the designed capacity) and service quality with the given resources.

This section presents a review of literature and the theoretical background for the work done in this thesis on effective maintenance analysis and optimisation as it is relevant to capacity and service quality enhancement. The topics included in theoretical reviews are as follows: capacity concept, quality of service, capacity enhancement plans, maintenance process, maintenance improvement, maintenance analysis and maintenance optimisation.

3.1 Railway infrastructure capacity and QoS

3.1.1 Capacity

The capacity of railway infrastructure is the total number of possible paths in a defined time window and with given resources, considering the actual path mix, infrastructure manager’s assumption in some nodes and quality demand from the market (Abril et al., 2008, Krueger, 1999, Landex et al., 2006, Patra, 2009, UIC, 2004). On a given railway infrastructure, capacity is a measure of the balance mix of number of trains, average speed, heterogeneity and stability (Landex et al., 2006). A specified mix of these four
capacity elements describes the consumption and utilisation of railway capacity. A typical mix of these elements used to describe capacity balance for metro and mixed traffic is shown in Figure 3.1.

In railway transport, infrastructure capacity could be defined based on inherent, practical, or operational considerations. Assessment of the different types of capacity is necessary to prompt augmentation of infrastructure utilisation and improvement of service quality of railway operations. The different types of capacity mentioned in the literature (Abril et al., 2008, Krueger, 1999, UIC, 2004) are described in Figure 3.2.

Another important aspect of Figure 3.2 is available capacity, an indication of additional
3.1. Railway infrastructure capacity & service quality

capacity that can be managed by the network or route if best practices and improvements are both identified and implemented. Exploring available capacity is the core of capacity enhancement plans and studies.

3.1.2 Quality of service

Quality of service (QoS) is an important indicator in railway transportation. It describes the collective effect of service performance, which determines the degree of satisfaction of a user with the service (IEC, 2014). An interesting aspect of QoS from maintenance viewpoint is the influence of availability on its accessibility and retainability characteristics (IEC, 2014). In railway infrastructure, QoS covers the following transport performance measures: punctuality regularity, reliability robustness, congestion, safety and comfort (IEC, 2014, CENELEC EN 50126, 1999, Nyström, 2008, Söderholm and Norrbin, 2013, Trafikverket, 2011b). This sub-section is not aimed at providing exhaustive information on the service quality of railway transport and its parameters; the above mentioned papers can be referred to for details. The characteristic measure that is directly emphasised in this thesis is punctuality, and other service quality measures such as safety and comfort are dealt with indirectly.

Punctuality is a function of the expected travel and transport times as related to the inherent capability of the system with planned stops and unplanned disturbances (Trafikverket, 2011b). Punctuality is measured by comparing the eventual train arrival times at different stations with the planned arrival times scheduled in the time table (Nyström, 2008, Söderholm and Norrbin, 2013). Various time tolerances are adapted by different IMs based on their business strategies, stakeholder’s expectations and national regulations. In most instances as in this thesis work, the time tolerance for punctual trains is positive 5 min. In addition, an unpunctual train is said to be delayed and the delay time is the additional time at which the arrival time plus 5 min is exceeded.

QoS is considered a function of practical or achievable capacity (Abril et al., 2008, Krueger, 1999), and a change in the capacity of a railway network might affect the expected QoS on the network. Because most railways define capacity at a specific QoS, it is then technically right to express capacity of a railway network as a function of QoS as shown in the hypothetical presentation in Figure 3.3.

3.1.3 Capacity and QoS enhancement plans

In railway transport, capacity situation and QoS can be viewed from market, infrastructure planning, traffic scheduling and operations perspectives (UIC, 2004). The capacity of a railway network is determined by infrastructure design and other parameters such as traffic conditions, maintenance conditions and operational incidents (Abril et al., 2008, Krueger, 1999). Several strategies and plans are commonly deployed to enhance the capacity situation of an infrastructure network. These span from short-term and
inexpensive measures to long-term and expensive measures. Basically, plans for the enhancement of capacity and QoS of railways can be grouped into the three categories: infrastructure modification, traffic planning improvement and M&R improvement. In summary, Figure 3.4 presents the enhancement plans commonly adopted by IMs to support or increase designed capacity. For further reading on capacity enhancement plans, refer to Abril et al. (2008), Boysen (2013), Cambridge Systematics (2007), Famurewa et al. (2013), Ferreira (1997), Gibson (2003), Higgins et al. (1996) and Khadem-Sameni et al. (2010).

![Figure 3.4: Capacity enhancement plans with emphasis on maintenance improvement](image-url)
3.2 Maintenance improvement

The scope of this study is limited to the maintenance of existing infrastructure; thus, maintenance improvement is of interest. Further, improvement aspects such as maintenance task analysis, planning and scheduling are covered.

3.2.1 Efficient maintenance execution process

The execution of maintenance tasks involves a series of subtasks that are carried out serially or concurrently. In general maintenance engineering terminology, maintenance execution can be described based on the time requirement of the broad division of the task as shown in Figure 3.5.

In railways, execution of maintenance tasks are further broken down into the seven subtasks listed below (CENELEC EN 60300-3-14, 2004, Smith and Mignott, 2012):

- Transportation: This involves moving maintenance and other support equipment to the task location.
- Confirmation: Ascertain if possession has been granted either at the beginning of the shift or during the shift when moving from one section to another.
- Waiting: Waiting for personnel, equipment, traffic, or other logistic purposes.
### Table 3.1: Maintenance process from different perspectives


<table>
<thead>
<tr>
<th>Generic maintenance process</th>
<th>Maintenance process in railway industry</th>
<th>Maintenance process in Trafikverket</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance budgeting</td>
<td>Budget determination</td>
<td>Budget allocation</td>
</tr>
<tr>
<td>Setting maintenance objectives</td>
<td>Identifying objectives from regulation &amp; white paper</td>
<td></td>
</tr>
<tr>
<td>Formulating Strategy</td>
<td>Establishing strategy from existing handbook</td>
<td></td>
</tr>
<tr>
<td>Establishing Responsibilities</td>
<td>Long-term quality prediction &amp; Diagnosis</td>
<td>Contract procurement</td>
</tr>
<tr>
<td>Planning</td>
<td>Project prioritisation &amp; selection</td>
<td>Condition assessment</td>
</tr>
<tr>
<td></td>
<td>Project identification &amp; definition</td>
<td>Maintenance need analysis</td>
</tr>
<tr>
<td>Scheduling</td>
<td>Possession allocation and timetabling of track possession</td>
<td>Track possession schedule (BAP and BUP)</td>
</tr>
<tr>
<td>Execution</td>
<td>Implementation</td>
<td>Execution</td>
</tr>
<tr>
<td>Assessment</td>
<td>Work Evaluation</td>
<td>Assessment &amp; verification</td>
</tr>
<tr>
<td>Improvement</td>
<td>Feedback loop</td>
<td>Follow up of contract</td>
</tr>
</tbody>
</table>

- Communication: Conversation on phone to obtain information relevant for maintenance commencement and documentation. With an effective maintenance process and system, this can be eliminated.
- Preparation: Setting up and dismantling of heavy-duty equipment takes considerable time. In addition, it includes track safety and clearance measures.
- Active repair or preventive maintenance time: This is the value-adding subtask that involves actual restoration of function. It covers the technical aspect of isolation, disassembling, cleaning, repairing, refurbishing, replacing, reassembling and testing equipment and components.
- Pre and post measurements: These are carried out to check how much work or restoration is needed in a particular maintenance task.

Based on a study carried out in the EU project AUTOMAIN (Smith and Mignott, 2012), there is great potential for improving infrastructure availability and capacity through efficient task execution by engaging relevant continuous improvement methodologies. Con-
3.2. Maintenance improvement

Continuous improvement methodologies typically use feedback from a process or customers to identify, reduce and eliminate suboptimal processes and initiate small and continual strides rather than giant leaps (ASQ, 2014, Dale et al., 2007, Masaaki, 1997). Some of the methodologies or techniques that are practical from the viewpoint of improving the task execution process in railways are six sigma, lean concept, theory of constraints and other PM frameworks (Dale et al., 2007, Dettmer, 1997, Klefsjö et al., 2001, Masaaki, 1997, Rahman, 1998).

**Six sigma** is a data driven improvement methodology which effectively utilises statistical tools to pinpoint sources of variation and ways of eliminating it. It is assumed that the outcome of the entire process will be improved by reducing the variation of multiple elements (Klefsjö et al., 2001, Nave, 2002).

**Lean** utilises the involvement of people in a value stream to identify and remove waste, which is defined as anything not necessary to produce the product or service. It assumes that waste is the main restriction to profitability and that many small improvements in rapid succession are of great benefit (Nave, 2002, Smith and Hawkins, 2004).

**Theory of constraints** assumes that every system has at least one thing that limits it from achieving higher performance versus its goal— weakest link. It assumes that constraints are opportunities for improvement and thus viewed as positive and that gradual elevation of the system’s constraints will improve its perfor-

*Figure 3.5: Maintenance times (CEN EN 13306, 2010)*
PM frameworks are strategic management and improvement systems employed to align business activities with the vision and strategy of an organisation, monitor organisation performance and facilitate decision making toward achieving strategic goals (BSI, 2014, Bititci, 1997, Neely, 2005, Parida and Chattopadhyay, 2007, Åhrén, 2008).

3.2.2 Effective maintenance analysis

For sustainable performance of railway infrastructure over its entire life span, functional need analysis at the design stage and maintenance need analysis at the operation stage are required (Trafikverket, 2007a). These analyses are vital to meet the infrastructure requirements in terms of quality, safety, reliability, maintainability, availability, capacity and QoS. Maintenance analysis in the context of this work is defined as the procedure for quantifying maintenance needs or identifying maintenance tasks and determining the specific information and resources required by these tasks (CENELEC EN 60300-3-14, 2004, Márquez, 2007). The tasks can be reviewed and adjusted later based on practical constraints such as available outage windows and need for availability maximisation or resource optimisation. For quantification and analysis of maintenance needs, the following approaches are used commonly in physical asset management (CENELEC EN 60300-3-14, 2004, Márquez, 2007, CENELEC EN 60300-3-1, 2005).

- Implementing manufacturers’ recommendations provided in the maintenance and operation manual or in similar documents.
- Adapting personal or organisational experience with the asset or similar assets.
- Studying and analysing technical documentation of the asset, such as drawings, diagrams and technical procedures.
- Considering regulatory and/or mandatory requirements, such as safety conditions of item operation and environmental regulations for item.
- Using inspection reports
- Using degradation models
- Using maintenance engineering techniques such as FMECA.

The last three approaches are promising for infrastructure performance improvement based on the fact that these techniques are data-driven, reliable and can be used to address different objectives.
Failure Mode, Effects and Criticality Analysis (FMECA) is commonly used for maintenance analysis. The requirements and procedures for performing FMECA were established and presented by the Department of Defense, USA (MIL-STD-1629A, 1980). Multi-criteria criticality analysis for effective maintenance priority ranking of engineering assets is another maintenance analysis technique, and some aspects of this technique were presented by Braglia (2000) and Márquez (2007). A review of some critical aspects of risk analysis important for the successful implementation in maintenance engineering as well as the use of risk analysis for the selection and prioritisation of maintenance activities was presented by Aven (2008). Another useful resource regarding the application of risk assessment techniques to maintenance analysis is the international standard on risk management (IEC 31010, 2009). For details on dependability analysis methods (such as fault tree analysis, event tree analysis, failure rate analysis etc.) that can be adapted for maintenance analysis, refer to the dependability management standard (CENELEC EN 60300-3-1, 2005).

In the railway industry, an analysis method to prioritise maintenance actions for railway infrastructure has been presented (Nyström and Söderholm, 2010). A risk evaluation technique developed for the specification and demonstration of reliability, availability, maintainability and safety (RAMS) of railway systems is another useful method (CENELEC EN 50126, 1999). The approach of implementing relevant RCM analysis as presented by Carretero et al. (2003) is also relevant for efficient and effective maintenance analysis. In essence, a well-tailored and data-driven maintenance analysis will help create a hierarchical list of items and assemblies for improvement and modification, in order to ensure the required capacity and QoS.

### 3.2.3 Optimum maintenance possession scheduling

Maintenance scheduling is a very essential aspect in maintenance management that is useful for identifying and assigning the needed support in an efficient way (CENELEC EN 60300-3-14, 2004). Optimum possession scheduling is a prerequisite for effective and efficient track possession management. It requires enough time to request for possession, identify personnel, acquire materials and spare parts from external inventory, ensure that transportation and support equipment are available, prepare work plans, provide necessary training, etc. The scheduling of (M&R) works is based on a priority system, to ensure that the most urgent and important tasks are carried out first and resources are used efficiently (CENELEC EN 60300-3-14, 2004). For instance, the requirements and nature of (M&R) works should be considered for effective possession scheduling. For improvement of capacity and QoS, the possession requirements described in Figure 3.6 need a decision support model or tool for optimum scheduling and utilisation of track possession time.

A study across railway IMs in Europe confirms that long possession windows for maintenance are planned 18—24 months in advance to ensure minimal traffic disruption (Para-
green, 2011). However, short possessions are requested within short time-scales to restore conditional failures reported during inspection and condition monitoring. Such inspections could be visual, or non-destructive such as ultrasonic inspection, eddy current checks, track geometry measurement, laser inspection and other techniques (Stenström et al., 2014, Trafikverket, 2005b, 2007b). Additional information on the general structuring of the railway infrastructure maintenance scheduling process was presented by Dekker and Budai (2002). They presented a state-of-the-art view on the two vital aspects of degradation modelling and scheduling of works for track possession. A short review of maintenance scheduling models is presented later in this chapter.

### 3.3 Maintenance optimisation theory

Maintaining reliable railway infrastructure in terms of technical performance and meeting the designed capacity of a network requires the use of decision support tool to optimise the maintenance plan and schedule. This is of greater importance in large scale maintenance activities such as grinding, tamping, or switches and crossing maintenance. Maintenance of systems or units after failure may be costly, and sometimes, requires extended track possession time to restore the system to a working state. This gives rise to important questions of why, when and how to carry out maintenance. In maintenance engineering, optimisation procedures are used to answer these questions by seeking optimal solutions and making compromises to achieve what is the most important. A balanced mix of cost, risk and performance is the sought optimum condition where the anticipated outcome is feasible with the minimum cost and an acceptable limit of risk (Kumar, 2008).
3.3. Maintenance optimisation theory

Maintenance optimisation models basically aim to find either the optimum balance between costs and benefits of maintenance or the most appropriate moment to execute maintenance (Dekker and Scarf, 1998). The common optimization criteria adopted in maintenance models are as follows:

- Minimisation of system maintenance cost rate.
- Optimisation of system reliability measures.
- Minimisation of system maintenance cost while satisfying reliability requirements.
- Optimisation of system reliability measures while satisfying the system maintenance cost requirement.
- Minimisation of track downtime and outages.

In railway infrastructure maintenance optimisation for capacity increase or track possession reduction, the following aspects must be considered in the modelling procedure: system layout, applicable maintenance policy, possible maintenance level, desirable optimisation criteria, planning horizon, amount of system information and appropriate model tool. Figure 3.7 shows other essential aspects that should be accommodated in a robust maintenance optimization model.

![Figure 3.7](adapted from (Wang and Pham, 2006))
3.4 Review of maintenance schedule models

Maintenance scheduling of railway infrastructure provides a short, medium or long term plan, of how preventive maintenance works will be performed on different segments within a definite horizon. Models are required in a model-driven maintenance schedule. Several researches within railways have addressed different aspects of data-driven maintenance scheduling models for effective track possession management. An aspect of the possession problem was addressed by Higgins (1998) for determining the best allocation of railway maintenance activities and crew to minimise train disruption. A methodology for dividing a railway network into working zones that will be taken out of service to carry out maintenance activities was presented by den Hertog et al. (2005).

Cheung et al. (1999) developed a track possession assignment program for assigning railway tracks to a given set of scheduled maintenance tasks according to defined constraints. A time-space network model was presented by Peng et al. (2011) to solve the track maintenance scheduling problem by minimising the total travel costs of maintenance teams and the impact of maintenance projects on railroad operation. Miwa (2002), Andrade and Teixeira (2011) and Quiroga and Schnieder (2010) addressed preventive maintenance scheduling program related to track geometry quality and tamping operation by using different approaches.

Furthermore, a preventive maintenance scheduling program was presented by Budai et al. (2006) to merge routine tasks and projects on a link over a certain period such that the sum of possession costs and maintenance costs is minimised. A mixed-integer programming model that optimises a production plan and suggests the best possible traffic flow given a fixed set of planned maintenance activities was developed by Forsgren et al. (2013). An optimisation-based possession assessment and capacity evaluation decision support tool was designed by Savelsbergh et al. (2014) to evaluate schedules of planned maintenance and renewal work for rail infrastructure. Finally, reviews on planning and scheduling techniques for preventive maintenance activities on railway were presented by Lidén (2014) and Soh et al. (2012).

A list of relevant research work on the optimisation of maintenance plan and schedule is presented in Table 3.2.
### Table 3.2: Short review on models for railway infrastructure maintenance scheduling

<table>
<thead>
<tr>
<th>S/N</th>
<th>Problem</th>
<th>Objective</th>
<th>Remark</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allocation of maintenance activities to time windows and crews to activities.</td>
<td>Minimise traffic disruption and completion time</td>
<td>Integer Programming model using tabu search algorithm for the solution</td>
<td>Higgins (1998)</td>
</tr>
<tr>
<td>2</td>
<td>Track possession assignment problem - Assign railway tracks to a given set of scheduled maintenance tasks according to a set of constraints</td>
<td>Maximises the assignment of job requests based on priorities and satisfies all constraints</td>
<td>Constraint-satisfaction techniques with heuristics was used.</td>
<td>Cheung et al. (1999)</td>
</tr>
<tr>
<td>3</td>
<td>Optimal maintenance schedule for track irregularities</td>
<td>Minimise tamping cost and maximize improvement of tamping operation</td>
<td>Integer programming model</td>
<td>Miwa (2002)</td>
</tr>
<tr>
<td>4</td>
<td>Preventive maintenance scheduling problem (PMSP)</td>
<td>Minimise possession cost and maintenance cost</td>
<td>Mathematical programming which is NP-hard. Heuristics are used in the solution</td>
<td>Budai et al. (2006)</td>
</tr>
<tr>
<td>5</td>
<td>Track maintenance schedules to improve track workers’ safety</td>
<td>Maximise manageable work load per night for rail track workers</td>
<td>Developed a two-step method of constructing a four week schedule with each working zone of the main lines closed to trains at night exactly once.</td>
<td>van Zante-de Fokkert and Fokkert (2007)</td>
</tr>
<tr>
<td>6</td>
<td>Bi-objective optimization problem for maintenance and renewal decisions related to rail track geometry.</td>
<td>Minimise total cost of planned maintenance and total number of train delays caused by speed restrictions</td>
<td>Simulated annealing technique was used to solve the problem</td>
<td>Andrade and Teixeira (2011)</td>
</tr>
</tbody>
</table>

*Continued on next page*
<table>
<thead>
<tr>
<th>S/N</th>
<th>Problem</th>
<th>Objective</th>
<th>Remark</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Optimising tamping operations in ballasted track</td>
<td>Minimise total number of maintenance actions on described track segment</td>
<td>Mixed binary linear programming</td>
<td>Vale et al. (2012)</td>
</tr>
<tr>
<td>8</td>
<td>Monthly maintenance schedule of track on a network</td>
<td>Minimise total maintenance cost and travel cost</td>
<td>Genetic algorithm optimisation technique</td>
<td>Zhang et al. (2013)</td>
</tr>
<tr>
<td>9</td>
<td>Production team scheduling problem as time-space network model and side constraints</td>
<td>Satisfying selected practical aspects of railway infrastructure including cost.</td>
<td>Simplified scheduling model was used to obtain an initial solution and then a multiple neighbourhood search algorithm was applied to improve the solution</td>
<td>Peng et al. (2011)</td>
</tr>
<tr>
<td>10</td>
<td>Optimising tamping schedule</td>
<td>Maximise an objective function, which is a quantitative expression of the maintenance objectives defined by the railway company</td>
<td>Heuristics algorithm</td>
<td>Quiroga and Schnieder (2010)</td>
</tr>
</tbody>
</table>
Chapter 4

DATA ANALYSIS AND MODEL FORMULATION

The data analysis and modelling techniques used in the study include statistical methods, operations research methods and mathematical models. Different analyses and modelling techniques were used for each appended paper considering the aim, underlying question, relevance of method, available data and other technical aspects. The frameworks and the adopted techniques for each paper are described below.

4.1 Development of aggregated health index

This work is presented in paper 1 and it demonstrates the use of a fuzzy inference system for aggregating selected railway infrastructure performance indicators to relate maintenance function with the capacity situation. The selected indicators consider the safety, comfort, punctuality and reliability aspects of railway infrastructure performance. The resulting composite indicator gives a reliable quantification of the health condition or integrity of railway lines. A case study of the assessment of overall infrastructure performance, which is an indication of capacity limitation, is presented using indicator data of five track sections on the network of Swedish Transport Administration for 2010—2012. The results are presented using a customised performance dashboard for enhanced visualisation, quick understanding and relevant comparison of infrastructure conditions for high level maintenance management.

Figure 4.1 shows a framework for computing the composite indicator as required for the management of physical assets such as railway infrastructure. The main steps in the framework are as follows:

- Selection of indicators
- Selection of aggregation technique
Selection of normalisation method

Aggregation process

![Figure 4.1: Framework for composite indicator computation](image)

The indicators have been selected such that all adequate information necessary for computing a reliable integrity index would be included. In the case study, the selected indicators cover the following:

- Indication of both functional failure and reliability of infrastructure (failure frequency);
- Indication of service performance in terms of QoS, which is a measure of customer satisfaction (punctuality or delay);
- Indication of safety performance (inspection remarks); and
- Indication of functional degradation and durability of infrastructure (TQI).

For weighting and aggregation of the indicators, a number of techniques have been considered: linear, geometric, multi-criteria approach and soft-computing. However fuzzy
logic approach is preferable in the context of this article because the problem being addressed here relates more to assessing the overall integrity of track sections over time for decision management than ranking track sections based on their integrity. Fuzzy logic is used in a process of formulating the mapping from given input parameters to an output using natural language. This method of indicator aggregation is called a fuzzy inference system (FIS) and is summarised as follows:

1. Selection of linguistic quantifier and development of membership function to describe indicators in fuzzy sets. Five linguistic terms (very high, high, average, low and very low) were considered adequate for obtaining manageable and distinct consequent mapping of the fuzzy composite indicator in the FIS. Furthermore, three linguistic terms (High, Average and Low) were used for fuzzification of the input parameters; this is based on the existing goals set by the infrastructure manager. The trapezoidal membership function given by equation 4.1 was used for representing the three fuzzy sets pertaining to the input parameters and the five fuzzy sets pertaining to the composite indicator. The selection of this function was based on its wide use for purposes related to indicator development. The constant terms a, b, c, and d, are parameters describing the trapezoidal membership function used in the development of the fuzzy sets.

\[
\mu_A(x; a, b, c, d) = \max\left(\min\left(\frac{x - a}{b - a}, 1, \frac{d - x}{d - c}\right), 0\right)
\] (4.1)

where

\[ A = \text{fuzzy set with } \begin{cases} \text{Output parameter} = \text{Very high, High, Average, Low, Very low} \\ \text{Input parameter} = \text{High, Average, Low} \end{cases} \]

2. Conversion of a crisp indicator into a fuzzy element by using the fuzzification method to obtain the membership values of each linguistic quantifier.

3. Aggregation of membership values on the antecedent (IF) parts to obtain the firing strength (weight) of each rule. Usually this is done via a fuzzy intersection operation using an AND operator or the minimum implication, as given by equation 4.3.

\[
Fuzzy set A = (x; \mu_A(x)) \quad x \in X \\
Fuzzy set B = (x; \mu_B(x)) \quad x \in X
\]

and operation \( \mu_{A \cap B}(x) = \min\left(\mu_A(x), \mu_B(x)\right) \) (4.3)

or operation \( \mu_{A \cup B}(x) = \max\left(\mu_A(x), \mu_B(x)\right) \) (4.4)
4. Generation of consequents from different combinations of antecedents using the established fuzzy inference rules.

5. Aggregation of the obtained consequents (fuzzy set) from each rule to obtain a single output fuzzy set using an OR operator or the maximum method for union of fuzzy sets. See equation 4.4.

6. Defuzzification of the output fuzzy set using the centre of mass method, where the centre of the area under the curve of the output fuzzy set is obtained using equation 4.5.

\[
Z^* = \frac{\int \mu(z)zdz}{\int \mu(z)dz}
\]  

(4.5)

The method described in the six steps above is implemented in MATLAB using inference rules collected from the experts. For example, computation of the fuzzy composite indicator (FCI) for a line section with FF = 2, Delay = 13.1, TQI = 97 and IR = 12.8 is shown in Figure 4.2. Only two of the 81 rules are applicable, and the crisp output of the Mamdani FIS is equal to 0.80.

Figure 4.2: Computation of FCI using Mamdani FIS
4.2 Constraint identification

This study is reported in paper 2 and it presents the application of graphical methods and the risk matrix as a maintenance analysis method for identifying track zones that are bottlenecks limiting operational capacity and service quality. Furthermore, an adapted criticality analysis method is proposed to create a hierarchical improvement list for addressing the problem of train mission interruption and reduced operational capacity. The underlining theory in this paper is the application of criticality analysis and the theory of constraints for bottleneck identification and improvement. The framework of the study and the analysis carried out are shown in Figure 4.3. The first stage involves simplification and segmentation of the selected line section into 39 traffic zones or segments, each of which represent technical divisions referred to as traffic zones by the Swedish Transport Administration. The 39 zones consist of operational areas and lines joining two operational areas. The second stage involves the compilation and processing of data collected and stored in different databases using relevant standards. The compiled data contain the records of all infrastructure failures and operational consequences in terms of train delays between 2010 and 2012.

Figure 4.3: Outline for analysis done in paper 2

In the third stage, a graphical data analysis method was used to explore and describe the basic features in a way that is relevant to maintenance analysis. Basically, the two indicators (failure frequency and train delay hour) were used to describe failure charac-
Table 4.1: Risk assessment matrix (Carretero et al., 2003, CENELEC EN 50126, 1999)

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Insignificant</th>
<th>Marginal</th>
<th>Critical</th>
<th>Catastrophic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent</td>
<td>Undesirable</td>
<td>Intolerable</td>
<td>Intolerable</td>
<td>Intolerable</td>
</tr>
<tr>
<td>Probable</td>
<td>Undesirable</td>
<td>Intolerable</td>
<td>Intolerable</td>
<td>Intolerable</td>
</tr>
<tr>
<td>Occasional</td>
<td>Tolerable</td>
<td>Undesirable</td>
<td>Undesirable</td>
<td>Intolerable</td>
</tr>
<tr>
<td>Remote</td>
<td>Negligible</td>
<td>Tolerable</td>
<td>Undesirable</td>
<td>Intolerable</td>
</tr>
<tr>
<td>Improbable</td>
<td>Negligible</td>
<td>Tolerable</td>
<td>Undesirable</td>
<td>Undesirable</td>
</tr>
<tr>
<td>Incredible</td>
<td>Negligible</td>
<td>Negligible</td>
<td>Tolerable</td>
<td>Undesirable</td>
</tr>
</tbody>
</table>

In the matrix, the frequency of infrastructure failure and the operational consequences of failure in terms of train delay were considered. It is remarkable to mention that the punctuality or delay characteristic of each traffic zone was modelled using lognormal distribution. A hypothetical grading of the failure frequency and consequential delay was done with input from experts to demonstrate the mentioned methodology.

In the fifth stage, an adapted criticality analysis method is proposed to identify restraining items/assemblies and present a hierarchical list for improvement. The three parameters used in the adapted criticality analysis are as follows:

i. Probability that assembly failure would result in operational consequence POC

\[ POC_a = \frac{f_a}{F_a} \quad (4.6) \]

ii. Total delay consequence of a system failure TD

\[ TD_{az} = \sum_{i=1}^{n} d_i \quad (4.7) \]

iii. Average risk number of parent higher-level system RF
4.2. Constraint identification

\[
RF_z = \left( \prod_{t=1}^{T} (RF_t)^t \right)^{1/T} \tag{4.8}
\]

Furthermore, the three approaches given by equation 4.10—4.12 are used to aggregate the parameters, with the min-max normalisation technique (equation 4.9) used for linear aggregation and the conventional geometric normalisation approach. Equation 4.13 is thereafter used to aggregate the scores of the three aggregation methods and give an overall improvement score.

Min-max normalisation technique for linear aggregation

\[
N'_{jaz} = \frac{C_{jaz} - \min(C_j)}{\max(C_j) - \min(C_j)} \tag{4.9}
\]

Geometric method

\[
S_{az-GM} = \prod_{j=1}^{3} c_{jaz}^{w_j} \tag{4.10}
\]

Normalised linear method

\[
S_{az-LM} = \sum_{j=1}^{3} w_j N'_{jaz} \tag{4.11}
\]

Normalised geometric method

\[
S_{az-NGM} = \prod_{j=1}^{3} N'_{jaz}^{w_j} \tag{4.12}
\]

Overall improvement scores

\[
OIS_{az} = \sqrt{\sum_{m=1}^{m_T} \left( \left( \frac{S_{az-m}}{\sum_{Z=1}^{Z_A} \sum_{A=1}^{A_z} S_{AZ-m}} \right)^2 \times w_m \right)} \quad a \in A \ z \in Z \tag{4.13}
\]

The list of notations used in equations 4.6—4.13 is given in Table 4.2. Detailed explanation of some of the parameters and the methods can be read in the second paper appended to this thesis.
Table 4.2: Table of notation for paper 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$POC_a$</td>
<td>Probability of operation consequence for assembly a</td>
</tr>
<tr>
<td>$f_a$</td>
<td>Frequency of mission interruption failure for assembly a</td>
</tr>
<tr>
<td>$F_a$</td>
<td>Count of reports of failure symptom and events for assembly a</td>
</tr>
<tr>
<td>$TD_{az}$</td>
<td>Total delay owning to the failure of assembly a in zone z</td>
</tr>
<tr>
<td>$RF_z$</td>
<td>Average risk number for zone Z over T years</td>
</tr>
<tr>
<td>$RF_t$</td>
<td>Risk number for zone z during year t</td>
</tr>
<tr>
<td>$C_{jaz}$</td>
<td>Value of parameter j for assembly a in zone z</td>
</tr>
<tr>
<td>$N_{jaz}$</td>
<td>Normalised value of parameter j for assembly a in zone z</td>
</tr>
<tr>
<td>$N'_{jaz}$</td>
<td>Min-max normalised value of $C_{jaz}$</td>
</tr>
<tr>
<td>$w_j$ and $w_m$</td>
<td>weight of parameter j and weight of method m</td>
</tr>
<tr>
<td>$S_{az=m}$</td>
<td>Score of assembly a in zone z using method m</td>
</tr>
<tr>
<td>$OIS_{az}$</td>
<td>Overall improvement score for assembly a in zone z</td>
</tr>
</tbody>
</table>

4.3 Deterioration based maintenance scheduling

Improvement in planning and scheduling of large-scale maintenance activities known for high possession requirements is an essential part of railway infrastructure capacity enhancement. The optimum allocation and efficient utilisation of track possession time plays a key role within this improvement potential. Therefore this study proposes a method to support maintenance decisions for track geometry maintenance, for example, tamping task. A mathematical model was developed to facilitate short-term optimum allocation and utilisation of track possession time for tamping operation. To this end, two key aspects of infrastructure maintenance planning, deterioration modelling for condition prediction and schedule optimisation have been presented in paper 3 as an approach for reducing time on track. Descriptions of the approach and the modelling procedure are given below and presented in Figures 4.4 and 4.4.

First, the track section is divided into 200-m segments to model track geometry degradation with time. Using the standard deviation of the longitudinal level, an exponential model has been used to describe the deterioration of each 200-m segment on the track section with the inspection data collected between 2007 and 2012. The parameters of the model were estimated using a non-linear regression method and the collected data. Second, using the deterioration model, an intervention decision is made with the set threshold such that the time on track for intervention is limited. An empirical recovery model is used to estimate the condition of the tamped segment after the tamping intervention. In addition, a schedule optimisation problem is formulated to support the intervention decision. The problem is in a form similar to mixed-integer linear programming problem with some defined heuristics to arrive at a realistic solution without unnecessary computational complexities. The technique used to solve the problem can be classified as exhaustive search, where the search is limited by the given heuristics.
4.3. Deterioration based maintenance scheduling

The algorithm was written in both FORTRAN and MATLAB. This methodology was extended to estimate the total cost of intervention over a finite horizon. The total cost of intervention included direct cost of maintenance and the cost of track possession.

This procedure was adapted to support a short term planning and scheduling program for a case study of 130-km-long single track from Kiruna to Riksgränsen by varying the intervention strategy (number of early interventions and number of late interventions). The optimum strategy was estimated using the graphical approach to locate the point with the minimum total cost. The result of this paper suggests a tamping plan that will lead to optimum allocation of track possession time while maintaining track geometry quality within the specified limits. Preventive maintenance in the context of this paper is meant to be early intervention with high-performance machine while corrective maintenance refers to late intervention with low-performance machine.

Figure 4.4: Flow chart for the optimisation of track possession and cost of tamping
4.4 Possession optimisation for deterioration based maintenance

This study is reported in paper 4 and is an extension of paper 3, it presents an analysis of track geometry maintenance with the aim of reducing the possession time requirement. The paper demonstrates augmented utilisation of possession time for tamping by using a simulation approach to estimate the optimum tamping cycle length, shift duration and assess improvement potentials of subtasks typical of large-scale maintenance. The approach employed in this article for augmentation of possession time utilisation involves statistical data analysis, simulation and mixed integer linear programming optimisation. The tamping policies considered are as follows: corrective policy (restoration of isolated defects), predetermined policy (line restoration at predetermined intervals based on usage) and prognostic policy (using degradation model to predict geometry quality before intervention). An optimum strategy seeks the ideal combination of the above policies such that both over-maintenance in the form of early intervention and under-maintenance in the form of late intervention are avoided.

The modelling and optimisation procedure to improve the utilisation of possession time for geometry maintenance is simplified in the flow chart shown in Figure 4.6 and explained thereafter.

i. Track simplification and characterisation: This involves mapping the track section with respect to its length, components, maintenance stabilising points as well as estimating the degradation rate of each 200 m-long segment using regression analysis.

ii. Track degradation: The exponential model described by Veit (2007) and modified
4.4. Possession optimisation for deterioration based maintenance

Figure 4.6: Flow chart for geometry quality simulation and schedule optimisation

by Quiroga and Schnieder (2012) is used to model the degradation of each segment. The initial standard deviation of vertical irregularity $\sigma(s,0)$ and the degradation rate $b(s)$ estimated from historical data were inserted into the formula given by equation 4.14 for prediction. Here, $T$ is time in days, $\varepsilon$ is the prediction error, a random variable modelled using truncated normal distribution with mean $\mu_{\varepsilon} = 0$, standard deviation $\sigma_{\varepsilon}$ and the two boundary values $a_{\varepsilon}, b_{\varepsilon}$ were estimated from the data.

$$\sigma(s, T) = \sigma(s, 0)e^{b(s)T} + \varepsilon \quad (4.14)$$

iii. Geometry restoration: For the first restoration of each segment, the first part of equation 4.15 is adopted and an assumption of imperfection in quality restoration is assumed for subsequent restoration attempts. This means that the quality after
Data analysis and model formulation

tamping $\sigma(s, T)_{\text{cum}}$ depends on the quality before tamping $\sigma(s, T)$, cumulative number of tamping operations (cum), quality achieved in last restoration $\sigma(s, T)_{\text{cum}-1}$ and the quality loss QL factor (taken to be 10% in the case study and based on expert opinion).

$$\sigma(s, T)_{\text{cum}} = \begin{cases} 
0.4002\sigma(s, T) - 0.0307 & \text{cum} = 1 \\
\sigma(s, T)_{\text{cum}-1} + QL(\sigma(s, T) - \sigma(s, T)_{\text{cum}-1}) & \text{cum} > 1 
\end{cases}$$ (4.15)

iv. Predetermined intervention decision: The decision regarding when to tamp is based on the cycle length. The exact day in the year for operation commencement can be selected freely based on other maintenance constraints and boundary conditions. Equation 4.16 and the associated formulations given in equation 4.17-equation 4.19 are used for estimating the possession time required for a maintenance cycle.

$$TPC = \sum_{T=1}^{365C} t_w(T)$$ (4.16)

where

Possession time during a shift

$$t_w(T) = t_v'(T) + t_v(T) + t_{\text{proc}}(T)$$

$$t_w(T) < t_w^*$$

Tamping time

$$t_v(T) = \frac{L_s(T) - L_a(T)}{v} + \sum_{i=1}^{N_a} g(ii) \cdot t_{\text{sc}(ii)}(T)$$ (4.17)

where

$$g(ii) = \begin{cases} 
1, & L_s(T) < L_{\text{sc}(ii)} < L_s(T) \\
0, & \text{else}
\end{cases}$$

$L_s(T) < L_a(T) < L_T$

Travelling time

$$t_v'(T) = \min \left| L_s(T) - L_{p(i)} \right| + \min \left| L_s(T) - L_{p(j)} \right|$$ (4.18)
4.4. Possession optimisation for deterioration based maintenance

Process time

\[ t_{\text{proc}}(T) = t_{\text{cm}} + t_{cf} + (t_{\text{wt}} + t_{\text{pre}}) \cdot np(T) \]  

Estimated time required for predetermined maintenance over the planning horizon

\[ TPC_{PH} = \left\lfloor \frac{PH - 1}{C} \right\rfloor TPC \]  

v. Corrective intervention decision: This is made using the prediction model and intervention threshold from the handbook (Trafikverket, 2014). For improved utilisation of possession time, locations with quality close to the threshold or surpassing it are merged in the same window such that the overall corrective maintenance time over the planning horizon is minimized. The objective function defined in equation 4.21 and the associated constraints are used for the optimisation.

\[ TPS_{PH} = \min \left( \sum_{T=1}^{PH} t_w(T) \right) \]  

s.t.

\[ t_w(N(T)) < t_w^* \]

\[ N(T) \leq N'(T) \]

\[ \sigma(s, T) < \sigma^* + 0.3 \]

Possession time during a corrective maintenance shift

\[ t_w(T) = t_w^*(T) + t_v(T) + t_{\text{proc}}(T) \]

\[ t_w(T) < t_w^* \]

\[ N'(T) = \sum_{s=1}^{S} f \left( \sigma(s, T) - \sigma^* \right) \]  

\[ f(x) = \begin{cases} 1 & x \geq 0 \\ 0, & \text{else} \end{cases} \]

Active repair time

\[ t_v(T) = \frac{N(T) \cdot d}{v} \sum_{i=1}^{S_{sc}} f(i_i) \cdot t_{sc(i_i)} \]
Data analysis and model formulation

\[ f(ii) = \begin{cases} 
1 & s(ii) = s(i), \quad i = 1, 2, \ldots, N \\
0, & \text{else}
\end{cases} \]

Travelling time

\[
 t'_v(T) = \min \left( \frac{s(1) \cdot d - L_{P(i)}}{v'}, \frac{s(N) \cdot d - L_{P(j)}}{v'} \right) + \left( s(N) - s(1) - (N - 1) \right) \cdot d 
\]

(4.24)

Process time

\[
 t_{\text{proc}}(T) = t_{\text{cm}} + t_{\text{cf}} + t_{\text{wt}} + (N - n_a)t_{\text{pre}} 
\]

(4.25)

vi. The total track possession time required over the planning horizon is estimated using equation 4.26. To reduce the possession time, shift length, cycle length and other process parameters are varied and the optimum conditions are obtained using the graphical approach.

\[
 TPT = TPC_{PH} + TPS_{PH} 
\]

(4.26)

This procedure was demonstrated with a case study to suggest a long-term maintenance plan for 130-km single track from Kiruna to Riksgränsen. The geometry data used was collected between 2007 and 2012. Measurement data are used to estimate the deterioration constant of each 200-m segment and the error distribution parameters as well, which are then used for geometry prognosis of each track segment. The simulation input data accounted for practical and realistic descriptions of the maintenance breakdown structure and machine parameters, which are listed in Table 4.3.

---

Table 4.3: Tamping machine parameters and subtask duration from maintenance process observation (Smith and Mignott, 2012) and expert discussion

<table>
<thead>
<tr>
<th>Machine/process parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travelling speed (v')</td>
<td>80 km/h</td>
</tr>
<tr>
<td>Tamping speed (v)</td>
<td>0.8 km/h</td>
</tr>
<tr>
<td>Switch tamping (t_{sw})</td>
<td>30-70min</td>
</tr>
<tr>
<td>Preparation time (t_{pre})</td>
<td>10 min</td>
</tr>
<tr>
<td>Confirmation time (t_{cf})</td>
<td>30 min</td>
</tr>
<tr>
<td>Communication time (t_{com})</td>
<td>20 min</td>
</tr>
<tr>
<td>Waiting time (t_{wt})</td>
<td>10 min</td>
</tr>
</tbody>
</table>
Table 4.4: Table of notation for paper 4

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TPT$</td>
<td>Total track possession time required over the planning horizon</td>
</tr>
<tr>
<td>$TPC_{PH}, TPS_{PH}$</td>
<td>Total possession time for cycle maintenance, and total possession time for spot maintenance</td>
</tr>
<tr>
<td>$N'(T)$</td>
<td>Number of segments above or close to intervention threshold $\sigma^*(2.6 \text{mm})$ on day $T$</td>
</tr>
<tr>
<td>$N(T)$</td>
<td>Number of segments tamped during corrective intervention</td>
</tr>
<tr>
<td>$t_{\text{prep}}, t_{w}$</td>
<td>Preparation time for setup, dismantling and safety measures, and total travelling time during a shift</td>
</tr>
<tr>
<td>$t_v, t_w$</td>
<td>Total tamping time during a shift, and shift duration ($\text{Max. duration } t_w^* = 6h$)</td>
</tr>
<tr>
<td>$d, n_a$</td>
<td>Length of each segment (200 m), and number of adjacent segments among tamped segments</td>
</tr>
<tr>
<td>$L_T, L_{SC}$</td>
<td>Total length of the track section, and location of turnouts on the track section</td>
</tr>
<tr>
<td>$L_s(T), L_e(T)$</td>
<td>Locations of start point $L_s$ and endpoint $L_e$ on the track section during a tamping shift on day $T$</td>
</tr>
<tr>
<td>$L_p$</td>
<td>Locations on the track section designated as temporary machine pick up points $L_p(i)$ and park at the end of a shift $L_p(j)$ ($i, j \in 1, ..., 4$)</td>
</tr>
<tr>
<td>$v_c$</td>
<td>Working speed during corrective interventions ($0.5 \text{ km/h}$)</td>
</tr>
<tr>
<td>$s(i)$</td>
<td>Index of segments with quality above threshold, ($i = 1, ..., 4$)</td>
</tr>
<tr>
<td>$s(ii)$</td>
<td>Index of segments with turnouts</td>
</tr>
<tr>
<td>$n_p$</td>
<td>Number of train passages during a shift</td>
</tr>
<tr>
<td>$PH, Y$</td>
<td>Planning horizon ($PH = 14 \text{ years in the case study}$), ($Y = 1, ..., PH$)</td>
</tr>
<tr>
<td>$t_{\text{sc}(ii)}, N_{sc}$</td>
<td>Additional working time required for turnouts with index $ii$, and number of turnouts on the section</td>
</tr>
<tr>
<td>$C, t_{\text{proc}}$</td>
<td>Maintenance cycle, and process time</td>
</tr>
</tbody>
</table>

For additional information about the simulation and optimisation method, refer to the fifth paper attached to this thesis.

4.5 Potential failure based maintenance scheduling

This study is reported in paper 5. It presents a data-driven scheduling approach to efficiently use available train-free periods for restoring potential failures such that availability and capacity are maximised. The formulated maintenance scheduling problem is to support effective and efficient scheduling of maintenance works that are not accommodated in the long-term plan. The problem is focused on the reduction of maintenance cost, possession cost and penalty cost. It is modelled as a mixed integer program with quadratic constraints and then solved using branch and cut algorithm. The flowchart of
the model is shown in Figure 4.7. The first stage entails division of the selected track section into maintainable sections. The segmentation is done such that each maintainable section would be manageable during maintenance from the logistics, safety and operational viewpoints. It is assumed that the maintenance service provider would be given permission to occupy only one maintenance section per train free window as other sections could be occupied by trains. The second stage deals with extraction of train-free windows from the existing timetable or train movement data. For the case study on monthly condition based maintenance schedule, 51 windows over a period of one month were considered usable from the traffic, safety and resource-availability perspectives. The durations of the selected windows varied between 1 and 3 h with an average size of 1.5 hours and about 75% of the windows were smaller than this average value. Furthermore, 50 maintenance tasks were selected from the historical records of potential failure; these represent the expected monthly workload for the track section. The tasks included in the monthly workload are S&C, overhead wire, rails, fastener and signal repairs, as well as ballast and sub-ballast spot tamping. An estimation of the possession requirement of each task is done by expert based on experience and this estimate varies between 0.25 and 3 h, depending on the type of work and estimated extent of damage.

The third stage involves formulation of the objective function. The objective function minimises the total cost, which is the sum of the direct and indirect maintenance costs. The direct maintenance cost is simply referred to as maintenance cost that comprises of fixed cost per task and a variable cost that depends on the estimated possession time required by each task. The indirect cost consists of variable possession cost, fixed window start-up cost and penalty cost. This objective function is given in equation 4.27.
4.5. Potential failure based maintenance scheduling

\[ \min \sum_{m \in M} \sum_{w \in W} f_{mw} X_{mw} \quad (4.27) \]

where \( f_{mw} \) is the aggregated cost for using window \( w \) on day \( d_w \) for task \( m \) with deadline on day \( D_m \). \( X_{mw} \) is the decision variable for carrying out task \( m \) during window \( w \). \( X_{mw} \) is a binary variable where 1 implies that task \( m \) is implemented in window \( w \) and 0 means otherwise. The aggregated cost of using window \( w \) for task \( m \) depends on the size of the task, deadline of the task and the cost parameters given in equation 4.28. In explicit terms, \( f_{mw} \) depends on the time \( t_m \) required for the implementation of task \( m \), maintenance cost per hour \( C_m \), fixed cost per task task \( C_{mst} \), hourly cost for window possession \( C_w \), start up cost (fixed charge) for using a window \( C_{ws} \) and the daily penalty cost \( C_p \) for exceeding the task deadline \( D_m \).

\[ f_{mw} = C_m t_m + C_{mst} + C_w t_m + F(X_{mw}) (C_{ws}) + F(d_w - D_m) (C_p) \quad (4.28) \]

where

\[ F(d_w - D_m) = \begin{cases} (d_w - D_m), & (d_w - D_m) > 0 \\ 0, & (d_w - D_m) \leq 0 \end{cases} \quad (4.29) \]

\[ F(X_{mw}) \in \{0, 1\} \quad (4.30) \]

s.t.

\[ \sum_{m \in M} F(X_{mw}) = \begin{cases} 1, & \sum_{m \in M} X_{mw} \geq 1 \\ 0, & \sum_{m \in M} X_{mw} < 1 \end{cases} \quad w \in W \quad (4.31) \]

The fourth stage involves formulation of the constraints, which are explained below:

Constraint 1: Implementation of maintenance tasks within a window should not exceed the window duration. This is presented by equation 4.32, where \( t_m \) is the time required to fix task \( m \), \( t_w \) is the duration of window \( w \) and \( Tr_{mw} \) is the time required to travel between the locations of two tasks \( m \) and \( m' \) on the same segment. An average travelling time of 10 min is used in the case study.

\[ \sum_{m, m' \in M} \left( t_m X_{mw} + Tr_{mwm'} \right) \leq t_w \quad m \neq m' \quad w \in W \quad (4.32) \]
Constraint 2: All tasks must be completed, that is, a task expected to take $t_m$ hours should have a total sum of $t_m$ hours. This constraint is defined in equation 4.33.

$$\sum_{w \in W} t_m X_{mw} = t_m \quad m \in M$$  \hspace{1cm} (4.33)

Constraint 3: This constraint is introduced to reduce travelling during a possession. It ensures that only repair tasks that are close to each other and on the same segment can be merged in a window. Equation 4.34 describes this constraint, where $m$ and $m'$ are two different tasks on segments $s_m$ and $s_{m'}$, respectively. $N_s$ is the total number of segments. This constraint is handled as a quadratic constraint.

$$X_{mw} X_{m'w} = \begin{cases} 0, & s_m \neq s_{m'} \\ X_{mw} - X_{m'w} & \end{cases} \quad m, m' \in M, m \neq m', w \in W, s_m, s_{m'} \in S, S(1,2,3,\ldots,N_s)$$  \hspace{1cm} (4.34)

However, in an alternative approach the problem is solved as a simple mixed integer linear program, and the quadratic constraint in equation 4.34 is replaced with a new linear constraint presented in equation 4.35. The new constraint ensures that only one task is carried out in a window. This alternative approach provides only the baseline solution for comparison.

$$\sum_{m \in M} X_{mw} \leq 1 \quad w \in W$$  \hspace{1cm} (4.35)

Boundary condition for variables is defined in equation 4.36 below.

$$X_{mw} \in \{0,1\} \quad m \in M, w \in W$$  \hspace{1cm} (4.36)

The model described above has a linear objective function and a combination of linear and quadratic constraints. Because the variables are binary, the model is treated as a mixed integer quadratic constraint program (MIQCP), a special case of the mixed integer program MIP. The MIQCP models, hereafter referred to as model 1 and 2, are solved using a branch and cut algorithm that combines the advantages of a pure branch and bound scheme and cutting planes scheme.

In model 1, the continuous QCP relaxation approach is used at the root and other nodes of the tree, and the sub-problem at the nodes are solved using the barrier algorithm. In model 2, the operation of the optimisation engine is modified for reliable improvement of the solution by pre-linearizing all quadratic terms in the model. This is achieved by introducing new variables to replace the quadratic terms and new constraints such
that the original problem remains unchanged. The sub-problems at the tree nodes are then solved using continuous LP relaxation with simplex algorithm. Furthermore, an alternative approach named model 3 is used to model the problem using simple mixed integer linear program (MILP). This is done by removing the quadratic constraints given by equation 4.34 and introducing the linear constraint given by equation 4.35 to obtain a baseline solution for comparison.
The results of this thesis are presented to reflect two vital aspects of maintenance improvement required for capacity and QoS enhancement: maintenance analysis for effective planning and maintenance optimisation for efficient scheduling. The results are presented such that the research questions are addressed under these two broad divisions. The results presented and discussed include: development of overall health index to support capacity investigation and maintenance decisions, development of suitable maintenance analysis method for performance constraint identification, deterioration based scheduling and possession scheduling for potential failure.

5.1 Maintenance analysis

This section presents the results related to performance analysis of networks, track sections, assemblies and items. This entails performance monitoring and evaluation using historical data for identifying, reducing and eliminating suboptimal processes or bottlenecks for enhancing railway transport service quality. The result presented here are useful to support maintenance decisions both at the tactical and the operational levels.

5.1.1 Development of aggregated health index

Research question 1: How can information about infrastructure condition be aggregated to support maintenance decisions for enhanced capacity?

The methodology for assessment of the health condition of railway infrastructure presented in section 4.1 is demonstrated with five track sections in the years 2010, 2011 and 2012. The information contained in each indicator is integrated into the fuzzy value to provide an overall picture of the line condition that complements the result of capacity analysis and simulation.
Figure 5.1 shows the computed health value of line 4 in a performance dashboard, giving information on the integrity of the line for the years 2010, 2011, and 2012. The performance dashboard gives the FCI value, which is an indication of the status of the lines and overall integrity. Additional information, which can be obtained from the fuzzy composite indicator presented in a simplified performance dashboard, is the trend of the indicator. An improving trend is shown by an upward arrow in the dashboard, while a deteriorating trend is shown by a downward arrow. The FCI value is graduated from 0 to 1 to reflect possible variation in the overall state of the line. The value of the FCI is, however, not meant to provide detailed information about the physical state of individual components but to check whether there is significant improvement or deterioration in the integrity of the infrastructure.

Figure 5.1: Performance dashboard for line 4

Figure 5.2 presents the fuzzy indicator value for the five lines considered in this article for the year 2012. This simplified presentation of the composite indicator provides a quick insight into the need for maintenance improvement, renewal, or investment on different lines. Adding this information to the capacity statement gives a new dimension from infrastructure viewpoint and helps maintenance service providers to easily convey the need for improvement or modification to decision makers.

Figure 5.2: Presentation of fuzzy composite indicators of the five lines for the year 2012

Figure 5.3 shows the computed FCI for the five lines over a period of three years. The health values or conditions of the lines presented as FCIs are connected to the following
5.1. Maintenance analysis

factors: inherent system condition, operating conditions, age, environmental conditions and maintenance conditions. The low FCI of line 1 is not only obvious in comparison with other lines, but it is also pronounced in its low value over the three investigated years. A reason for this is the heavy haul traffic operated on it and the high capacity utilisation of the line. The integrity of line 1 is influenced by its high operation profile; an axle load of 30 tonnes and an average daily traffic volume of 90,000 gross tonnes. The line condition is traffic induced because it is clear that there exists a non-linear relationship between the infrastructure condition and the traffic volume (Lyngby et al., 2008). Another factor that is common to both lines 1 and 2 is the influence of environmental conditions on the state of the lines; these lines are located in a region with harsh winter conditions.

The condition of line 3 apparently got better in 2011 but eventually deteriorated in 2012. Given that maintenance and renewal (M&R) efforts are often focused on lines with high class and capacity consumption, the conditions of line 2 and line 3 are suspected to be low owing to their low capacity consumption. Line 4 is a double line with mixed traffic on the western region, and it has maintained a health value greater than 0.6 over the three years under consideration. Even though the total length of the track is long, the reported failure frequency has been consistently low and the track quality index has been high. These make the integrity of the line to be considerably good in relation to the average capacity utilisation. However if the operational capacity is to be further increased, additional M&R measures would be required. The condition of line 5 is quite good owing to its high health value, which was above 0.8 during 2010, 2011 and 2012. It is a line with more than 200 trains per day and high gross tonnage-kilometre, and yet

![Composite indicators for five lines](image-url)

*Figure 5.3: Composite indicators for five lines*
the performance or condition of the infrastructure is remarkable. An apparent inference is that the M&R practice on this line is effective in relation to the capacity condition and could be extended to other lines. Moreover, the state of the line is an indication that it is ready to accommodate additional traffic as long as possible conflicts can be resolved during timetable simulation.

The use of the suggested line integrity index will complement conventional capacity analysis methods. For example, high level maintenance decision would be based on both overall capacity status and infrastructure health status. However, the quality of FCI depends on data quality and experience of the experts engaged in the development of the inference rules. Furthermore, there is need to extend the punctuality indicator to cover incidences of cancelled trains, as well as standardise the inspection strategy in terms of frequency, details, and priority classification remarks on all lines for consistency.

5.1.2 Constraint identification with adapted criticality analysis

Research question 2: Which maintenance analysis method is suitable to identify the weakest links and critical items on a track section from capacity and punctuality viewpoints?

Basic graphical analysis using bar and Pareto charts, risk matrix and adapted criticality analysis methods have been used to explore and describe the failure characteristics and loss of punctuality on the selected line section. Historical data between 2011 and 2013 were used and the aim was to identify the weakest links for maintenance improvement.

Figure 5.3 shows train mission interruption due to infrastructure failures and the contribution of each assembly/system. Switches and crossings (S&C), track circuit and track contributed the most to the number of mission interruptions or failures on the line section in 2012. On the other hand, overhead cable, track circuit and track had the most operational consequences in terms of the delay time in the same year. However, the location of the actual weak spot is unknown, and there is need for additional analysis to provide spatial information for effective planning and action steps.

Pareto charts of the frequency of traffic interrupting failure and train delays as a result of the failures on the 39 zones of the line described in chapter 4 are shown in Figures 5.5 and 5.6 respectively. These are useful for identifying critical locations that contribute significantly to the capacity and punctuality problems on the line. Figure 5.6 shows that approximately 30% of the traffic impacting failure occurred in zones 13, 3 and 10, while about half of the zones were responsible for 80% of train mission interruptions on the line in the year 2012. Figure 5.6 also shows that about 40% of the delays caused by infrastructure failure could be traced to zones 18-20, 8—9, 12—13 and 2—3. In addition, a third of the zones are critical based on the Pareto 80% estimation of significance. Comparing Figures 5.5 and 5.6, it can be seen that the hierarchical listing of the zones based on their states differs from the failure frequency and delay consequence perspectives.
5.1. Maintenance analysis

Figure 5.4: Contribution of each infrastructure type to train mission interruption

Figure 5.5: Pareto chart of higher-level system failure frequency in 2012

Furthermore, the risk assessment matrix shown in Table 4.1 was used to combine the information in the individual indicators. The procedure for risk assessment described in sub-section 4.2 was used to group each traffic zone into risk categories. This is an indication of their respective contributions to service quality and capacity reduction on the line section. In 2012, zones 13 and 18—20 fell in the intolerable risk categories and required attention in the form of maintenance or renewal.
Figure 5.6: Pareto chart of delay consequence for higher-level system failure in 2012

Figure 5.7: Categorisation of the traffic zones based on the risk of limiting service quality and capacity in 2012
5.1. Maintenance analysis

In continuation of the analysis, two actions can be taken to support planning for improvement: cause and effect analysis for the intolerable and undesirable risk zones or additional criticality analysis to identify capacity- and punctuality-critical items that should be considered for maintenance. Depending on the strategy or available resources, a cause and effect analysis can be carried out at this stage using Figure 5.8. The cause and effect can be used in the discussion to analyse the identified constraints influencing the functional performance of the line section.

Finally, the information about the risk categories of the zones is used for the adapted criticality analysis described earlier. Criticality scores of the various systems and assemblies in the different critical zones were estimated and used for priority ranking for maintenance intervention. The criticality scores estimated using the three aggregation techniques and the overall improvement score for planning and improvement are listed in Table 5.1. It can be seen that the resulting priority ranking of the aggregation methods differs in some cases.

The overall improvement scores combine the information from the aggregation techniques and provide final ranks that can be interpreted as a measure of maintenance needs for the year 2013. The outcome of the analysis is a hierarchical list of the lower-level system for effective maintenance planning to enhance operational capacity and punctuality of the line. The systems on top of the priority list (overhead cable on zone 18–20, track circuit on zone 15 and alternative power line on zone 12–13) are considered to be the weakest links or restraining assemblies or systems. These should be prioritised in maintenance planning to improve the operational capacity and quality on the line section.

---

Figure 5.8: Root causes of infrastructure bottleneck
5.2 Maintenance optimisation

The management of maintenance process requires appropriate philosophy and optimized policy to achieve maintenance excellence. For instance, considering the problem described in the introduction of this thesis, maintenance optimisation is a practical way of enhancing railway capacity and quality of service. This is achievable through effective possession management at the tactical and operational levels. Two models have been proposed to
support (i) long term possession scheduling for track geometry maintenance (ii) efficient scheduling of conditional failure or other routine track works into available train-free windows in the short term.

5.2.1 Deterioration based maintenance scheduling

Research question 3: How can track possession time be optimised using prognostic maintenance approach and deterioration based scheduling models?

Developments in railway management have led to increasing need for optimum planning and scheduling of maintenance activities to support the demanding target for improved capacity, safety, cost effectiveness and other service quality aspects of railway transport. To this end a methodology for optimum scheduling of tamping is proposed to minimise the direct cost of intervention and the cost of track possession while the geometry quality is within the desired level. In order to optimise track possession for geometry maintenance, two approaches have been presented: deterministic degradation approach for short term planning reported in paper 3, and the probabilistic approach for long term planning reported in paper 4.

Deterministic approach

The focus of the study was to suggest a geometry maintenance plan for a short term of about 2 years. Two key aspects of infrastructure maintenance planning are considered in this paper: deterioration modelling and scheduling optimisation. An emphatic consideration in this approach was the use of two different tamping machines with different quality outputs and accessibility. The high-quality machine is used for early intervention (preventive tamping) but has limited accessibility while the second machine is used for late intervention (preventive tamping) and has higher accessibility.

The degradation rate of the longitudinal level for each 200-m segment was estimated using the exponential model explained in the model procedure and the collected inspection data. The distribution of degradation constant of all 592 segments is shown in Figure 5.9. It is heavily skewed toward the right, indicating the existence of critical spots with rapid degradation in their geometric quality. More than 50% of the track sections has an exponential degradation rate between 0.00024 and 0.00060. The differential degradation rate along the track sections reflects non-homogeneity and variation in the track components along the track length. In fact, the 200-m track segments can be regarded as non-identical units in terms of quality deterioration. An essential maintenance requirement revealed by the figure is the balance of preventive and corrective tamping because a higher exponential rate will require a greater number of interventions than a lower exponential rate, as also noted by Arasteh Khouy et al. (2013).

Using the model procedure described earlier, the number of corrective interventions that will be required in 2 years for different numbers of preventive tamping shifts was es-
Results and discussion

Figure 5.9: Distribution of longitudinal defect growth rate

Estimated and the result is shown in Figure 5.10. The sensitivity of the result is also shown using the 90% prediction limit of the recovery model fit. Increasing the number of allocated shifts for preventive tamping decreases the consequent number of corrective tamping shifts up to a point where there is no need for corrective tamping after the initial round at the beginning when a few sections were above both intervention thresholds.

Figure 5.10: Corrective interventions and preventive interventions
5.2. Maintenance optimisation

Using the proposed optimisation procedure shown with different cost ratios for the two tamping policies, the direct cost of tamping interventions over a short period of two years was estimated and is shown in Figure 5.11. A high value of the \( \frac{C_p}{C_c} \) ratio results in higher cost of intervention when the number of preventive maintenance shifts increases. Considering a cost ratio of \( \frac{C_p}{C_c} = 1 \), the economic optimum policy is to have only a few preventive intervention shifts. However, track possession time and track quality are other parameters that need to be considered.

![Figure 5.11: Total maintenance costs with different cost ratios](image)

Figure 5.12 shows the total track possession time over a period of 2 years for different numbers of preventive intervention shifts. The strategy with only corrective intervention \( (N_{p,\text{shift}} = 0) \) promises lower track possession time in comparison with the present maintenance strategy that is not supported by any scheduling model. Strategies with an \( N_{p,\text{shift}} \) value between five and eight are efficient because of the low track possession time within this relatively short planning horizon. Furthermore, Figure 5.13 shows the total cost of intervention (using \( C_{DT} = 2C_c \)) with different numbers of shifts allocated for preventive intervention. According to the results shown in Figure 5.13, selecting strategies with more than eight preventive maintenance shifts will result in additional cost owing to overly frequent track possession. In this study \( N_{p,\text{shift}} = 8 \) is suggested because it is cost-efficient and produces better quality than other strategies with fewer preventive maintenance shifts. This study demonstrates the use of deterioration based scheduling models for possession management of large scale tasks. However, additional considerations such as an adequately long planning horizon are required for further conclusion on the case study.
Probabilistic approach

The deterministic approach presented earlier was extended to accommodate more practical aspects of geometry maintenance, to optimise possession and to study additional improvement possibilities. In contrast to the deterministic approach, the Monte Carlo simulation approach was used to model the evolution of track geometry quality. In addition, the tamping policies considered are as follows: corrective policy (restoration of isolated defects), predetermined policy (line restoration at predetermined intervals based on usage) and prognostic policy (using degradation model to predict geometry quality.
The focus of the study was to suggest optimum tamping cycle length and shift duration so that the possession time requirement over a long horizon is reduced. It further evaluates the improvement potentials of subtasks typical of large scale maintenance works.

The time required for predetermined tamping during a cycle with different shift lengths is estimated using the approach described in chapter 4 with the specified process and machine parameters. The result presented in Figure 5.14 indicates that the possession requirement decreases exponentially with an increase in the maintenance window duration when the section is completely closed to traffic. A short maintenance window requires several shifts and overly long track possession time to complete a tamping cycle on a track section. This suggests that maintenance shifts must be long enough to reduce the impact of non-value added tasks such as travelling time and other process subtasks. However, the capacity utilisation of an important line would not permit such a long, uninterrupted white period. The number of train passages to be allowed within the white period is also shown in Figure 5.14, and this traffic requirement is often a major cause of extended time on the track. Figure 5.15 shows the mean possession time for each window size considering the triangular distribution of train passages for each window. For augmented track possession time utilisation, it is obvious from Figure 5.15 that the optimum possession in a shift should be between 5.5 and 6 h. These white periods are efficient because of the reduced impact of non-value-added tasks.

![Figure 5.14: Maintenance window and track possession time for a tamping cycle and train passages that would be cancelled](image-url)
For reducing time on track for geometry maintenance, there is a need to optimise the predetermined cycle length. The geometry quality degradation of each segment is estimated using an exponential function with error term that is normally distributed. The Monte Carlo simulation technique was used and 10 000 simulation cycles were carried out. Figure 5.16 shows the estimated total possession requirement for geometry maintenance of the track section with 95% confidence interval over the planning horizon. The required possession duration is very high with too short a maintenance cycle, but as the interval duration increases, possession reduces to a point where it starts to increase again until it is more or less constant. Placing the maintenance cycle length between years 3 and 6 appears to be somewhat optimum from possession duration viewpoint. However, a cycle length of 3 years seems to be relatively small and could present a risk of reducing the lifespan of ballast and other track components. Choosing a cycle length of 4 years is optimum from the possession view point, but extending the cycle length to 6 years can be an optimum solution from both the possession and the ballast life span viewpoint. This is consistent with the submission of the best practice guide for optimum track geometry durability that repeated tamping actions themselves can cause additional ballast damage and will therefore reduce its service life (UIC, 2008).

In addition, the achievable benefits of improving the tamping process are analysed herein. The improvements include machine improvement, operators’ skill improvement, traffic management improvement, and resource scheduling and planning improvement. The action plans suggested for the improvement of each subtask are given in Table 5.2. Figure 5.17 shows that improving tamping speed creates the largest potential for reducing track possession time. However, improvement of tamping speed is limited by design if durable track quality is to be ensured. This, therefore, makes small improvements in other subtasks important and appreciable. In addition, reducing the number of train
interruptions during a maintenance shift has good potential for reducing the track possession time because set up and waiting times are reduced considerably.

\[ \text{Figure 5.16: Total track possession time to suggest optimum maintenance cycle length} \]

<table>
<thead>
<tr>
<th>Machine/process parameter</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travelling speed</td>
<td>Better scheduling tool, permission to travel at line speed</td>
</tr>
<tr>
<td>Tamping speed</td>
<td>Improvement in operator skill, automatic setting of tamping parameters, and selection of suitable tamper</td>
</tr>
<tr>
<td>Preparation time</td>
<td>Reducing traffic interruption which increase the frequency of set up and dismantling, and standard operations procedure.</td>
</tr>
<tr>
<td>Confirmation time</td>
<td>Automatic track occupation control system, standard operations procedure</td>
</tr>
<tr>
<td>Communication time</td>
<td>Improved management system and communication process</td>
</tr>
<tr>
<td>Waiting time</td>
<td>Standard planning procedure, reduction of traffic interruptions, which leads to additional waiting time. Lean traffic management</td>
</tr>
</tbody>
</table>
Furthermore, the tasks are grouped into value-added (active tamping), necessary non-value added (preparation and travelling) and non-value added tasks (waiting, confirmation, communication). Figure 5.18 shows that the improvement of value-added tasks with the existing tamping machine can lead to a 15% reduction in track possession time over a tamping cycle.
Elimination of the non-value added tasks as possession requiring tasks reduced possession time by 28%. The necessary non-value added tasks can be improved to reduce the possession time by 13% without any machine change. A total reduction of up to 45% in possession time can be achieved using the existing tamping machine. In addition, if the tamping machine is replaced with a high-speed dynamic tamper, the potential reduction owing to the modified tamping speed alone would be 36%.

5.2.2 Potential failure based maintenance scheduling

*Research question 4: How can a data-driven approach be used to efficiently schedule maintenance tasks into available train-free windows?*

The result of using the proposed model and the alternative models for efficient possession management of potential failure and deferrable repairs is presented in this subsection. Model 1 uses the continuous QCP relaxation approach and barrier algorithm to solve the sub-problem at the root and other nodes of the tree. In model 2, the optimisation operation is modified by pre-linearizing all quadratic terms in the model before solving the sub-problems at the tree nodes using continuous LP relaxation with simplex algorithm. For the purpose of comparison, the performance of an alternative MILP scheduling model is presented as model 3.

The overall performances of the models are summarised in Table 5.3 and elaborated thereafter. Models 2 and 3 generated the optimal solution in less than 1 min while model 1 yielded the best feasible solution in approximately 8 min even though it was allowed to run until the set time limit of 3 h. Model 3 is an expensive approach, in the sense that it does not permit combinations of tasks on the same segment into one window. Therefore, the total maintenance cost associated with model 3 is the highest among the three models. Model 2 is associated with the minimum total maintenance cost while model 1 is very close to model 2 and far better than model 3 in terms of cost.

Looking further into the schedule generated by each model and comparing it with their respective deadlines, model 2 has the best performance with all works scheduled and no task delayed. In model 1, only one task would be implemented after the deadline, while in model 3, 4 tasks would be delayed. In terms of the number of days for which capacity would be affected owing to infrastructure conditions, model 3 has the worst performance while model 2 has the best performance, i.e. no reduction in capacity. The average window utilisation is the highest for model 2 owing to the possibility of merging maintenance tasks in a single window. Considering the number of windows utilised, models 1 and 2 utilise fewer windows to complete all the tasks, leaving behind four unused windows that can be used for other purposes.

In addition to the overall performance evaluation of the models given in Table 5.3, a
Table 5.3: Performance evaluation of the models

<table>
<thead>
<tr>
<th>Method</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIQCP</td>
<td>MIQCP</td>
<td>MILP</td>
<td></td>
</tr>
<tr>
<td>Algorithm</td>
<td>QCP relaxation (Barrier)</td>
<td>Prelinearisation, LP relaxation (Simplex)</td>
<td>LP relaxation (Simplex)</td>
</tr>
<tr>
<td>Solution</td>
<td>Feasible</td>
<td>Optimal</td>
<td>Optimal</td>
</tr>
<tr>
<td>Constraint violation</td>
<td>Nil</td>
<td>Nil</td>
<td>Nil</td>
</tr>
<tr>
<td>Optimum value (€)</td>
<td>34420</td>
<td>33627</td>
<td>40168</td>
</tr>
<tr>
<td>Number of delayed works</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Number of affected days</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Average window utilisation</td>
<td>85</td>
<td>87</td>
<td>82</td>
</tr>
<tr>
<td>Number of windows used</td>
<td>47</td>
<td>47</td>
<td>50</td>
</tr>
</tbody>
</table>

breakdown of the total maintenance cost for the optimal task schedules generated by the three models is given in Figure 5.19. The total direct maintenance cost $C_{\text{maint}}$ and possession cost $C_{\text{poss}}$ are similar for all models because these cost elements are functions of estimated repair time, and all tasks are expected to be completed in a window. However, the distinct differences between the optimality of the models are the total penalty cost $C_{\text{pen}}$ and the window start-up costs $C_{\text{wind-st}}$. The schedule generated by model 2 has no penalty cost because no task is delayed and its window start-up cost is small because the schedule minimises the number of used windows.

![Figure 5.19: Breakdown of the total maintenance cost for the proposed model and alternative models](image)
For the suggested model (model 2), the initial window duration and left-over time for each window are presented in Figure 5.20 for visual assessment of the possession allocation efficiency. The obviously high left-over time represent the unused windows. Approximately 80% of the remaining windows can be considered practically unusable for new repair works because they are too small to travel and start a new task. Furthermore, the left-over window duration can be analysed and classified, as shown in Figure 5.21 for unplanned opportunity based track works such as manual inspection and routine checks.

![Figure 5.20: Initial window duration and left-over time for the proposed model](image)

The class A windows are efficiently used and are perhaps not usable for other works if encroachment into the maintenance withdrawal time before the next train is to be avoided. The class B windows can still be used for opportunity based maintenance such as small-scale track works, routine checks or inspection on the same segment where the window time was originally used. The class C windows are unused and can thus be used for any type of work on any segment provided other utilisation constraints are not violated.

An important aspect of the proposed approach for possession management is the analysis of the optimal schedule or reason for infeasibility. In instances where all tasks cannot be scheduled in available windows owing to the number or the size of the windows, a review of the task can be conducted. For instance, the review could allow the possible break-up of some tasks into smaller chunks or remove less significant works that will be spread later over the left-over usable windows. The model can be adapted with little improvement to support other scheduling cases, including night possession with long
duration, where merging of tasks in different segments is allowed within the same window.

\[ 0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0 \quad 1.2 \quad 1.4 \]

\[ 0 \quad 2 \quad 4 \quad 6 \quad 8 \quad 10 \quad 12 \quad 14 \quad 16 \quad 18 \]

**Figure 5.21:** Histogram plot of the left-over window durations
Chapter 6

Conclusions

To further increase the competitiveness of railway transport via quantity and quality of delivered service, improvements to the maintenance and renewal process are required. This thesis has addressed vital aspects of maintenance improvement for enhancing railway infrastructure capacity and service quality. The work herein covers maintenance analysis for identifying weak links and capacity critical items to support continuous improvement using historical maintenance data. The work also extends to planning and scheduling of maintenance tasks from the condition viewpoints. This involved the use of infrastructure condition data with a model-driven approach to suggest when maintenance tasks should be carried out to facilitate efficient utilisation of possession times and maximise availability. The key findings of the case studies and the contributions of the research are given in subsequent subsections.

6.1 Key findings

- The proposed Fuzzy composite indicator is a better indication of railway line integrity than failure frequency, which is used widely in railway maintenance management. FCI aggregates information about potential failure, failure and operational consequence of failure in a single integrity index. It can be used to support maintenance decisions for enhanced capacity.

- Fuzzy composite indicator is a good measure of the M&R need on a line as imposed by operational profile, capacity consumption and adopted maintenance strategy. For instance, lines with consistently low FCI in the case study are those with heavy haul operation, harsh weather condition or inadequate M&R owing to lack of possession time.
Conclusions

− The presented risk assessment matrix is useful for localising bottlenecks and constraints that limit service quality and capacity of a line. For instance, in the case study, traffic zones that fell into the intolerable risk category are “capacity weak links” that require improvement actions. The proposed cause and effect diagram can be used for initiating investigation into the causes of these bottlenecks at the operational level. Furthermore, the weakest items/assembles on the weak links are identified using the adapted criticality analysis for continuous improvement. These items should be prioritised during maintenance planning to improve the operational capacity and service quality on the studied track section.

− There was a varying degradation of the geometry quality along the studied 130-km track section and about half of the segments had an exponential degradation rate of between 0.00024 and 0.0006. This indicated the need for a deterioration-based scheduling model to ensure effective possession management.

− The track quality degradation model supports the quantification of tamping needs, while the scheduling model supports decisions on how to merge restoration of spot failures and use the possession time optimally for geometry maintenance.

− If train passages are to be accommodated during extended maintenance possession time, a maintenance window of 5.5—6 h is optimum for large scale maintenance in terms of total possession requirement. Also choosing a cycle length of 4 years is optimum from the possession view point, but extending the cycle length to 6 years can be an optimum solution from the possession, cost and ballast life span viewpoints.

− Improvement of tamping speed creates the largest potential for reducing track possession time. However, the increase in tamping speed is limited by design if durable track durability is to be ensured. A total reduction of up to 45% in track possession time per tamping cycle can be achieved using the existing tamping machine.

− The use of maintenance windows for routine works or condition based maintenance is a promising approach for possession management especially in corridors where complete night dedication for maintenance is not practical. The proposed MIQCP model is a good approach to efficiently schedule maintenance tasks into available train-free windows. For instance, the model generates a schedule with minimum cost and zero capacity loss due to infrastructure condition in the case study.
6.2 Contributions

The contributions of this research are highlighted below:

− Suggestion of an approach to compute a composite indicator or integrity index of railway lines for maintenance decision making. This gives complementary information that is useful during capacity investigation and can be included in the network statement.

− Proposal of adapted criticality analysis method to create a hierarchical improvement list for addressing the problems of train mission interruption and reduced operational capacity. This will facilitate maintenance decisions and continuous improvement at both the operational and the tactical levels.

− Development of degradation-based scheduling model to suggest practical tamping plan and optimum allocation of track possession time while retaining track geometry within specified limits. The model combines stochastic simulation of track geometry quality and maintenance scheduling routine.

− Development of a methodology to estimate the optimum maintenance interval and shift duration for large-scale maintenance works such as geometry maintenance. This facilitates improved usage of track possession time.

− Development of a short-term maintenance scheduling model to use efficiently available train-free periods for restoration of inspection remarks such that availability and capacity are optimised. This will help support the restoration of potential failures and other works not accommodated in the long-term maintenance plan.

6.3 Suggested future work

In continuation of the work done so far in this thesis and for practical consideration of other essential maintenance aspects, the following works are suggested for future research:

− Addition of other relevant performance indicators to the improvement analysis so that safety and economic aspects are covered in addition to the operational aspects.

− Inclusion of boundary condition in degradation based scheduling by introducing operational hindrance costs when maintenance cannot be carried out owing to seasonal limitations. Addition of other constraints in the planning and scheduling model to represent railway maintenance resource state.

− Development of long-term maintenance plan for other maintenance tasks such as grinding, which are considered to have a large impact on network capacity consumption.
Conclusions

- Extension of the short term possession scheduling model to long term planning horizon. In addition, consider other scheduling scenarios such as multiple track scenarios, task implementation order and other technical or logistic prerequisite for the tasks.

- Development of robust planning and scheduling tools that supports the coordination of maintenance and renewal activities at the national level for effective possession management.
References


References


PART II
APPENDED PAPERS
Composite indicator for railway infrastructure management

Abstract The assessment and analysis of railway infrastructure capacity is an essential task in railway infrastructure management carried out to meet the required quality and capacity demand of railway transport. For sustainable and dependable infrastructure management, it is important to assess railway capacity limitation from the point of view of infrastructure performance. However, the existence of numerous performance indicators often leads to diffused information that is not in a format suitable to support decision making. In this paper, we demonstrated the use of fuzzy inference system for aggregating selected railway infrastructure performance indicators to relate maintenance function to capacity situation. The selected indicators consider the safety, comfort, punctuality and reliability aspects of railway infrastructure performance. The resulting composite indicator gives a reliable quantification of the health condition or integrity of railway lines. A case study of the assessment of overall infrastructure performance which is an indication of capacity limitation is presented using indicator data between 2010 and 2012 for five lines on the network of Trafikverket (Swedish Transport Administration). The results are presented using customised performance dashboard for enhanced visualisation, quick understanding and relevant comparison of infrastructure conditions for strategic management. This gives additional information on capacity status and limitation from maintenance management perspective.

Keywords Composite indicator · Infrastructure capacity · Fuzzy logic · Performance dashboard · Strategic decisions · Line integrity

1 Introduction An essential task in railway infrastructure management is the evaluation of the network capacity. The standard method for the calculation of railway capacity follows criteria and methodologies from international perspective [1]. The use of simulation tools and techniques has enhanced the analysis of railway capacity for improvement for infrastructure managers [2–5]. These tools have not only supported the estimation of capacity consumed but also have helped in evaluating how it has been utilised and how it can be better utilised. An efficient management of infrastructure capacity should accommodate different views and requirements relating to customer need, infrastructure condition, timetable planning and actual operating conditions [1]. Generally, some factors have been identified as constraints to achievable capacity since they apparently limit capacity enhancement attempts in traffic management. These limitations include priority regulations, timetable structure, design rules, environmental, safety and technical constraints [1]. In Sweden, the infrastructure manager makes an annual evaluation of the infrastructure capacity situation and utilisation. This evaluation gives the track occupation time on all the line sections and
also capacity limitation due to additional train path demands that cannot be met because of excessively high track occupation time [6]. For example, in 2012 about 7% of all the line sections in the Swedish railway network had an average daily consumption greater than 80%, 15% line sections between 60% and 80%, and 77% of the line sections with less than 60% track occupation time [6]. Furthermore, capacity limitation is based on the level of capacity consumption in relation to additional request for traffic volume, weight per metre, axle load and train paths.

Addressing railway capacity from the point of view of infrastructure integrity assurance is not well addressed by the present capacity assessment procedures. It is, therefore, a subject of interest in maintenance research. An issue that is addressed in this study is the extension of capacity analysis to quantification of health condition of railway infrastructure under certain traffic profile. Such integrity indicator or measure of infrastructure performance gives an additional measure of capacity limitation on a line. Using infrastructure performance indicators in capacity analysis help to relate maintenance and renewal functions to the capacity condition of a network and also facilitate effective maintenance decision making.

Conventionally, the assessment and analysis of infrastructure performance is carried out using individual indicators such as punctuality, frequency of failure, track quality index, etc., separately. Extensive studies on the identification and management of performance indicators which are related to railway infrastructure have been studied by Stenström et al. [7] and Åhreń and Parida [8]. However, such indicators should be aggregated to present the condition or integrity of infrastructure in a holistic way such that it can be related to the capacity condition of the infrastructure. To this end, railway infrastructure performance indicators and the process of aggregating them as a composite indicator are studied in this paper.

The argument surrounding the use of composite indicator has been addressed by Galar et al. [9], where the strengths and weakness of composite indices are highlighted. Composite indicator has been proven to be a tool for benchmarking and strategic decision making [9–12], and can be used for monitoring maintenance and renewal in a capacity enhancement programme. A detailed technical guideline for the construction of high-quality composite indices was given by Nardo et al. [10]. In addition to this, the framework to guide the development of composite indices in the field of asset management has been presented by Galar et al. [9]. The contribution of this paper is the development of composite performance indicator for infrastructure management, useful for relating maintenance functions to the capacity condition of a network and facilitating effective maintenance decision making.

The rest of the paper is organized as follows: Sect. 2 presents the framework for computing composite indices, and Sect. 3 describes a fuzzy logic approach for the development of fuzzy composite indicator (FCI). The details of the case study are presented in Sect. 4, and the results and discussion are presented in Sect. 5. The final section presents the concluding remark of this paper.

2 Framework for computing composite indicator

The integrity and usefulness of composite indices depend largely on the framework which guides the computation process. To develop a composite indicator with acceptable quality and approximate characterisation of the state of a physical asset, it is essential to deploy a well-structured guideline that addresses the core issues. This will prevent both overestimation and underestimation of the overall state of the asset. Figure 1 provides a framework for the computation of composite indicator as required for the management of physical assets such as railway infrastructure. The core issues of the framework are as follows:

- Selection of indicators
- Selection of aggregation technique
- Selection of weighing method
- Aggregation process

![Fig. 1 Framework for composite indicator computation](image)
2.1 Selection of indicators

Systems of performance indicators for general physical asset management and precisely for railway infrastructure management have been presented in different literatures [7, 8, 13–15]. These indicators are used for the assessment of maintenance contracts, infrastructure integrity and service quality, and also prompt alert for quick intervention. All indicators are, however, not required in the development of a composite indicator, and there is need to use appropriate criteria in the selection of most relevant indicators. The indicators selected should present adequate information necessary for the computation of a reliable integrity index. In the case study, the selected indicators cover the following:

- indication of both functional failure and reliability of the infrastructure (failure frequency);
- indication of service performance in terms of quality of service which is a measure of the customer satisfaction (punctuality or delay);
- indication of safety performance (inspection remarks); and
- indication of functional degradation and durability of the infrastructure (Track quality index).

2.1.1 Failure frequency

This is the count of the number of times a component or system on a line is not able to perform the required function. Failure categories suitable for use in railway applications have been classified into three classes: immobilising failure, service failure and minor failure. In this study, the count of failure is limited to functional failure that interrupts the traffic flow leading to significant and major consequences on either economy or operation. Minor failures that do not prevent a system or line from achieving its specified performance or cause train delay are not considered in the failure count because of the extensive and complex nature or railway systems.

2.1.2 Punctuality

This is an aspect of operational consequence arising from interruption in the planned travel times of trains due to the reduction or termination of the functional performance of the infrastructure. It is measured either in terms of minutes of delay or the number of trains that arrived earlier or later than schedule. Further, the philosophy of punctuality differs from one infrastructure manager to another; hence, it is common to use non-negative arrival delay which is estimated after 5 min post the scheduled arrival time.

2.1.3 Track quality index (TQI)

This is a value that characterizes the track geometry quality of a track section based on the parameters and measuring methods that are compliant with the standard. Since there are different kinds of analyses and uses of track quality geometry data, therefore the aggregation and computation method for track quality index could be on detailed, intermediate and overview levels [25]. This study utilized an overview TQI which summarizes a large amount of data for strategic decisions or for long-term network management by infrastructure managers. The track quality index used in this study was selected for the following reason: to reflect the integrated track quality view by combining standard geometry quality parameters, to identify with the standard quality index used by the infrastructure manager (Trafikverket), and to provide for easy fuzzy description by experts using linguistic term. Equation 1 shows the formula used for the evaluation of TQI, and Fig. 2 gives a hypothetical illustration and description of TQI values. A track with a perfect geometry quality has a TQI equal to 150 but it degrades over time based on traffic loading, formation condition, track layout and other factors.

\[
TQI = 150 - \frac{100}{3} \left( \frac{\sigma_{TH}\_LL}{\sigma_{TH}\_LL} + 2 \frac{\sigma_{TH}\_AC}{\sigma_{TH}\_AC} \right),
\]

where \( \sigma_{TH}\_LL \) and \( \sigma_{TH}\_AC \) denote the standard deviations of the longitudinal level, and of the combined alignment and cross level; \( \sigma_{TH}\_LL \) and \( \sigma_{TH}\_AC \) represent the comfort threshold of the parameters.

2.1.4 Inspection remarks

Examination of a system by observing, testing or measuring its characteristic condition parameter at predetermined intervals is an essential aspect of operation and maintenance. Such an inspection could be a visual inspection or non-destructive testing such as ultrasonic inspections, eddy current checks, track geometry measurement, laser inspections and other dedicated techniques. For the railway infrastructure, inspection is based on the traffic volume and the line speed. It is a usual practice that reports are generated as inspection remarks after inspection. The remarks are classified into priority levels on the basis of the seriousness of the observation. The priorities of the remarks considered in the case study are acute and weekly categories [16].

In the selection process, it is important to carefully address likely correlations between the indicators, especially if a linear or geometric aggregation method is used. Table 1 shows that there is a significant correlation between failure frequency and delay time (using the
Spearman’s rho for monotonic relationship and $p$ value for statistical significance), whereas other indicators have neither a linear nor a non-linear correlation. However, approximately 20% of the variation in the delay time is not explained by the failure frequency, showing that operational consequence in terms of delay is not fully explained by the failure frequency. In addition, in the field of traffic management, the total delay caused by infrastructure integrity is a function of the traffic volume and homogeneity, downtime (summation of active maintenance time and waiting time) and frequency of failure. Thus punctuality although correlated with the failure frequency is considered in the construction of the FCI, since it gives additional information on the consequence of failure on customer, which is not explained by the frequency of failure.

2.2 Aggregation of indicators

Considering the need to integrate different variables and indicators in a single indicator, several methodologies/techniques have been developed and deployed to aggregate such indicators. The available techniques and methods for the aggregation of indicators include the following:

- **Linear aggregation or Simple Additive Weighting (SAW) method**—It is useful when all indicators have comparable measurement units and all physical theories are respected or when they can be normalised. Weak indicators can be masked or compensated by other strong indicators; thus this method requires careful implementation.
- **Geometric method**—It can be used for indicators with non-comparable ratio scale where reduced measure of compensation is required in the aggregation of the constituent indicators.
- **Multi-criteria approach with specific rules**—This is basically used when a number of criteria/indicators are involved in the computation and when highly different dimensions are aggregated in a composite indicator. Basically it entails an evaluation of N alternatives using C criteria, and then aggregating the result using special rules and theories. Examples include: Analytic hierarchy process (AHP), ELECTRE, TOPSIS, VI-KOR, etc.
- **Soft computing approach**—This approach is used when the constituent indicators can be expressed in linguistic terms and then aggregated using computing with words (such as fuzzy logic). The advantages of this approach include the following: modelling of non-linear behaviour; accommodation of imprecision in the normalisation of the data; aggregation without subjective allocation of weights to the indicators; ranking of alternatives in such a way that the output value can be treated as the health value or integrity index. On the other hand, the reliability of the composite indicator depends on the experience of the expert group; it requires additional information to explain the underlying physical phenomenon responsible for the variation of its value.

The fuzzy logic approach is preferable in the context of this article, since the problem being addressed relates more to assessing the overall integrity of line over time for strategic purpose than ranking them based on their integrity.

### Table 1: Spearman’s rho and $p$ value for statistical correlation between the indicators

<table>
<thead>
<tr>
<th></th>
<th>FF</th>
<th>Delay</th>
<th>IR</th>
<th>TQI</th>
<th>$p$ value</th>
<th>FF</th>
<th>Delay</th>
<th>IR</th>
<th>TQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>1.00</td>
<td>0.79</td>
<td>−0.09</td>
<td>0.21</td>
<td>FF</td>
<td>1.00</td>
<td>0.00</td>
<td>0.69</td>
<td>0.36</td>
</tr>
<tr>
<td>Delay</td>
<td>0.79</td>
<td>1.00</td>
<td>−0.15</td>
<td>0.20</td>
<td>Delay</td>
<td>0.00</td>
<td>1.00</td>
<td>0.51</td>
<td>0.38</td>
</tr>
<tr>
<td>IR</td>
<td>−0.09</td>
<td>−0.15</td>
<td>1.00</td>
<td>−0.42</td>
<td>IR</td>
<td>0.69</td>
<td>0.51</td>
<td>1.00</td>
<td>0.06</td>
</tr>
<tr>
<td>TQI</td>
<td>0.21</td>
<td>0.20</td>
<td>−0.42</td>
<td>1.00</td>
<td>TQI</td>
<td>0.36</td>
<td>0.38</td>
<td>0.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>

FF failure frequency; TQI track quality index, IR inspection remarks

Fig. 2: Description of track quality index
3 Fuzzy logic method

Fuzzy logic is based on imprecise human reasoning and exploits the tolerance for imprecision to solve complex problems and support decision making on complex systems [17–19, 20]. The underlying technique in fuzzy logic is computing with words or linguistic variables. The concept of linguistic variables creates the possibility for an approximate characterisation of processes which are too complex or too imprecise, by conventional quantitative analysis. It is a logical way to map an input space to an output space using a fuzzy set [21]. The capability of a fuzzy system for making implications between antecedents and consequents makes it appropriate for complex system analysis [18, 19]. This explains the application of fuzzy logic in the aggregation of indicators for the computation of a composite indicator suitable for strategic purposes.

3.1 Fuzzy inference system

The fuzzy inference system (FIS) is a process of formulating the mapping from given input parameters to an output using a natural language technique known as fuzzy logic [11]. Basically, the input parameters into FIS can either be fuzzy or crisp inputs, and the outputs are mostly fuzzy sets, but can be transformed to crisp outputs, since this is preferable for easy decision making. An FIS can be decomposed into three phases—input phase, aggregation phase and output phase as shown in Fig. 3.

The input phase involves a linguistic description of the parameters and fuzzification to obtain a fuzzy set of each input parameter. The aggregation phase has two steps that facilitate the mapping of the input parameters to output, i.e. inference rules and fuzzy set operation. The output phase defines the fuzzy set of the output parameter and also presents the final indicator in either fuzzy or non-fuzzy value [18].

3.2 Membership function

The membership of an element from the universe in a fuzzy set is measured by a function that attempts to describe vagueness and ambiguity due to the nature of the boundaries of the fuzzy sets. Elements of a fuzzy set are mapped to a space of membership values using a function-theoretic form [18]. This function associates all elements of a fuzzy set to a real value within the interval 0–1.

3.3 Aggregation process

The aggregation process involves two operations known as inference rules and fuzzy set operations. Fuzzy inference rule is a collection of linguistic statements that describe how the FIS should make a decision regarding the integration of the input into an output [18]. These rules form the basis for the FIS to obtain the fuzzy output that can be transformed into a non-fuzzy numerical value which are required in a no-fuzzy context. This is mainly based on the concepts of the fuzzy set theory and relations; it uses linguistic variables as its antecedents and consequents. The antecedents are the IF expressions which should be satisfied. The consequents are the THEN statements which are inferred as output, when the IF antecedents are satisfied [22]. The common inference rules are formed by general statements such assignment, conditional or unconditional statements [22]. The connectors used in the fuzzy rule-based system are ‘OR’ and ‘AND’ and their operations are described as follows:

Fuzzy set $A = \{ x, \mu_A(x) \}$, $x \in X$,

Fuzzy set $B = \{ x, \mu_B(x) \}$, $x \in X$,

AND operation $\mu_{A \cup B}(x) = \min(\mu_A(x), \mu_B(x))$. (2)

Fig. 3 Fuzzy inference system for computation of composite indices

\[ \text{INPUT} \rightarrow \text{AGGREGATION PROCESS} \rightarrow \text{OUTPUT} \]

\begin{itemize}
  \item Indicator selection
  \item Linguistic description
  \item Membership Function
  \item Fuzzy set
  \item Linguistic description
  \item Membership Function
  \item Fuzzy set
  \item Composite indicator
\end{itemize}
3.4 FIS approach

The most common approaches used in fuzzy inference systems are the Mamdani and Takani Sugeno approaches [22]. Basically, the working principle of Mamdani FIS can be explained as follows [11, 18]:

1. Selection of linguistic quantifier and development of membership function to describe the indicators in fuzzy sets.
2. Conversion of the crisp indicator into a fuzzy element using fuzzification method to obtain the membership values of each linguistic quantifier.
3. Aggregation of the membership values on the antecedent (IF parts) to parts the firing strength (weight) of each rule. Usually this is done in a fuzzy intersection operation using an AND operator or the minimum implication as shown in Eq. 2.
4. Generation of the consequents from the different combinations of antecedents using the established fuzzy inference rules.
5. Aggregation of the obtained consequents (fuzzy set) from each rule to obtain a single output fuzzy set using an OR operator or the maximum method for union of fuzzy sets. See Eq. 3.
6. Defuzzification of the output fuzzy set using the centre of mass method or the centre of gravity under the curve of the output fuzzy set. Z* is the defuzzified value or centre of mass obtained from the algebraic integration of the membership grade of element Z in the output fuzzy set C using Eq. 4.

\[ Z^* = \int \frac{\mu_c(z)}{\sum \mu_c(z)} z \, dz \]  \hspace{1cm} (4)

3.5 Composite indicator for railway management

There is a need to combine the information provided by simple output indicators to facilitate strategic decision making. Thus four indicators have been selected to develop a composite indicator for the assessment of the integrity of railway infrastructure. The selected indicators are hereafter referred to as the input parameters of a FIS, which are aggregated to obtain an indicator known as FCI. The FCI is graduated from 0 to 1 to indicate the integrity of the infrastructure, which is afterwards described by five linguistic terms or fuzzy sets. The selected linguistic terms are considered adequate for a simplified scaling of the FCI and for obtaining distinct consequent which can be easily managed in the FIS. A trapezoidal membership function has been used for developing the fuzzy sets for the

<table>
<thead>
<tr>
<th>Output Parameter</th>
<th>Low (a, b, c, d)</th>
<th>Average (a, b, c, d)</th>
<th>High (a, b, c, d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>(0, 0, 3, 6)</td>
<td>(3, 6, 7, 10)</td>
<td>(7, 10, 30, 30)</td>
</tr>
<tr>
<td>TQI</td>
<td>(0, 0, 70, 80)</td>
<td>(70, 80, 85, 95)</td>
<td>(85, 95, 150, 150)</td>
</tr>
<tr>
<td>IR</td>
<td>(0, 0, 5, 10)</td>
<td>(5, 10, 13, 18)</td>
<td>(13, 18, 50, 50)</td>
</tr>
</tbody>
</table>

FF failure frequency (failure/10^5 train km), TQI track quality index, IR inspection remarks (remarks/10^4 tonnage km)

composite indices, i.e. very high, high, average, low and very low. The selection of this function is based on its wide use for purposes related to indicator development. It is described by the expression given in Eq. 5. Further, three linguistic terms or fuzzy sets (high, average and low) have been used in the fuzzification of the input parameters based on the existing goal levels set by the infrastructure manager. The trapezoidal membership function in Eq. 5 was used for representing the three fuzzy sets, i.e. high, average and low.

\[ \mu_a(x; a, b, c, d) = \max \left( \frac{x - a}{b - a}, \frac{1 - x - c}{d - c} \right), \]  \hspace{1cm} (5)

where \( A = \) fuzzy set

\[ A = \{ \text{Output Parameter} \rightarrow \text{Very High, High, Average, Low, Very Low} \} \]

The constant terms \( a, b, c \) and \( d \) are parameters describing the trapezoidal membership function used in the development of the fuzzy sets. Table 2 shows the parameters of the membership functions used for the input parameters, while Fig. 4 shows the membership function of the FCI. These parameters cover the possible range of value of the indicators and are obtained on the basis of statistics, existing goals and expert opinion.

4 Case study

An assessment of the integrity of selected lines on the Swedish transport administration network is carried out using composite performance indicator. The approach described in the previous section is applied to compute the FCI. Some lines are selected to cover the different maintenance regions of the railway administration. The traffic characteristics on the lines differ, as well as boundary conditions such as the weather and local conditions. A brief description of the lines is provided in Table 3. In addition, the capacity situation on the five lines in 2011 as carried out by Wahlborg and Grimm [6] using the conventional view of time table planning and UIC 406 capacity method is presented in Fig. 5.
Results and discussions

The result of the assessment of health condition of the selected lines using the FIS is presented and discussed below. Figure 6 shows the procedure for computing FCI for one of the lines for the year 2010. Considering FF = 2, Delay = 13.1, TQI = 97 and IR = 12.8, only two of the 81 rules are applicable, and the crisp output of the Mamdani FIS is equal to 0.80.

5.1 Fuzzy composite indicator

The procedure shown in Fig. 6 is followed to compute the aggregated non-fuzzy value which is the health value for each of the five lines in the years 2010, 2011 and 2012. The information contained in each indicator is integrated into the fuzzy value to provide an overall picture of the line condition that complements the result of capacity analysis and simulation. The FCI value is graduated from 0 to 1 to reflect the possible variation in the overall state of the line. The value of the FCI is, however, not meant to give detailed information about the physical state of individual components, but rather to check whether there is significant improvement or deterioration in the integrity of the infrastructure. For enhanced visualisation and understanding of the result of the FIS, a customised performance dashboard tool is used for presenting the performance information. These images act as a gateway to scorecards, help in quick problem identification, and accentuate the additional value for the time and resources spent on performance management. Figure 7 shows the performance dashboard for line 4, giving information on the integrity of the line for the years 2010, 2011 and 2012. The performance dashboard gives the value of the FCI that is an indication of the status of the lines and a measure of capacity limitation. Additional information which can be obtained from the FCI presented in a simplified performance dashboard is the trend of the indicator. An improving trend is shown by an upward arrow in the dashboard, while a deteriorating trend is shown by a downward arrow. It is worth mentioning that the infrastructure manager does not have targets for the FCI for the different lines class yet; thus the level colouration in the performance dashboard is only used for demonstrating possibilities presented by this approach.

Figure 8 presents the fuzzy indicator value for the five lines considered in this article for the year 2012. This simplified presentation of composite indicator gives quick insight into the need for maintenance, renewal or investment on the different lines and is useful for evaluating the overall performance of the maintenance service providers. Adding this information to capacity statement gives a new dimension from infrastructure point of view and helps maintenance service providers to easily convey the need for improvement to strategic decision makers.

Figure 9 shows the computed FCI for the five lines over a period of 3 years. The health value of line 1 is the least and that of line 5 is the highest; these indicate that the infrastructure on line 5 is in good condition and that the

---

Table 3 Description of selected lines

<table>
<thead>
<tr>
<th>Line</th>
<th>Maintenance region</th>
<th>Type of traffic</th>
<th>Average daily tonnes</th>
<th>Track (km)</th>
<th>Axle load</th>
<th>Line class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>North</td>
<td>Iron ore</td>
<td>90,263</td>
<td>125</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>North</td>
<td>Mixed</td>
<td>32,179</td>
<td>175</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>East</td>
<td>Mixed</td>
<td>74,014</td>
<td>59</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>West</td>
<td>Mixed</td>
<td>73,552</td>
<td>231</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>South</td>
<td>Mixed</td>
<td>121,678</td>
<td>102</td>
<td>22</td>
<td>1</td>
</tr>
</tbody>
</table>

Line class 1 metropolitan areas, 2 large connecting lines and 3 other important goods and passenger lines

---

5 Results and discussions

The result of the assessment of health condition of the selected lines using the FIS is presented and discussed below. Figure 6 shows the procedure for computing FCI
lines with low FCI require improvement. The conditions of the lines are connected to the following factors: inherent system condition, operating conditions, age and maintenance conditions. The low FCI of line 1 is not only obvious in comparison with other lines, but it is also pronounced in its low value over the 3 years investigated. A reason for this is the heavy haul traffic operated on it and the high capacity utilisation of the line. The integrity of line 1 is basically influenced by its high operation profile; an axle load of 30 tonnes and an average daily traffic volume of 90,000 gross tonnes. The line condition is traffic induced as it is clear that there exists a non-linear relationship between the infrastructure condition and the traffic volume [23]. Another factor which is common with both line 1 and line 2 is the influence of the environmental condition on the state of the lines; these lines are located in the region with harsh winter conditions.

The condition of line 3 apparently got better in 2011, but eventually deteriorated in 2012. Since, maintenance and renewal (M&R) efforts are often focused on lines with high class and capacity consumption, the conditions of line 2 and line 3 are, therefore, suspected to be low owing to their low capacity consumption. Line 4 is a mixed and double line on the western region and has maintained a health value greater than 0.6 over the 3 years under consideration. Even though the total length of the track is long, the reported failure frequency has been consistently low and the track quality index is high. These make the integrity of the line to be considerably good in relation to the average capacity utilisation, however, if the operational capacity is to be further increased, additional M&R measures would be required. The condition of line 5 is quite good owing to its high health value that is above 0.8 during 2010, 2011 and 2012. It is a line with more than 200 trains per day and high gross tonnage kilometre, yet the performance or
condition of the infrastructure is remarkable. An apparent inference is that the M&R practice on this line is effective in relation to the capacity condition and could be extended to other lines. Also, the state of the line is an indication that it is ready to accommodate more traffic as long as possible conflicts can be resolved during timetable simulation.

This approach of computing line integrity index complements the conventional capacity analysis methods especially in a format that the maintenance department can appreciate for decision making.

5.2 Comparison between FCI, FF, CISAW and CIAHP

There is a need to compare the crisp output from the FIS with some other standard methodologies for computing composite indicators. Basically there is no well-established technique for aggregating indicators for strategic purposes in the railway industries. However, the result of the fuzzy logic approach is compared with failure frequency and also indices obtained from SAW (CISAW) and AHP (CIAHP) approaches.

5.2.1 FCI and FF

A common practice in railway transport is to use frequency of traffic interrupting failure of a line for characterising the condition of the infrastructure. Figure 10 shows a quick view of the comparison between FCI and failure frequency (FF) for the year 2010. From the perspective of failure frequency, lines 1 and 5 are both on the extreme positions showing a very high and a very low failure frequency, respectively. This is not expected to change even with contribution from other factors in the fuzzy logic approach. This is supported by Fig. 10, as line 1 has the least fuzzy score, while line 5 has the highest. It gives some level of confidence to the soft computing approach of assessing the integrity of railway infrastructure. Furthermore, the additional information from other indicators used in the computation of FCI makes it a better indication of the line integrity. From the perspective of failure frequency, line 4 has very close condition with lines 2 and 3. Upon the addition of information on the operational consequence of failure, track quality and inspection remarks as shown in the FCI values, line 4 can be clearly recognised to have better condition and perhaps better M&R practices.

5.2.2 FCI and CISAW and CIAHP

In order to justify and motivate the use of FCI, its result is compared with composite indices obtained from simple additive weighing (CISAW) and AHP (CIAHP) methods. In the SAW method, the simple indicators are normalised using inverse min–max function shown in Eq. 6. The same experts used in the fuzzy aggregation rules were involved in the derivation of weights for the different indicators.
using pairwise comparison. The final computation of the composite indicator is done using the expression given in Eq. 7.

\[ I_{ij} = \frac{1}{C_0} \frac{X_{ij}}{\min(X_i)} - \frac{\min(X_i)}{\max(X_i) - \min(X_i)} \]  

(6)

\[ C_{SAW} = \sum_{p=1}^{n} w_p I_{ij} \]  

(7)

where \( I_{ij} \) and \( X_{ij} \) represent the normalised value and actual value of indicator \( i \) for line \( j \) and year \( t \), respectively. \( X_i \) represents the actual value of indicator \( i \) for all the lines and for the year \( t \), whereas \( w_p \) is the weight of indicator \( p \) and \( n \) is the total number of indicators.

The AHP combines intuition and logic with data and judgment based on experience. The procedure developed by Saaty [24] is followed and Expert Choice software is used to implement AHP as appropriate in the context of this study. The software is employed to structure the computation process, and measure the importance of constituent indicators using pairwise comparison. It also facilitates the absolute measurement for deriving priorities of the selected lines with respect to the indicators. The objective information from data and the subjective judgment of experts are then synthesized to obtain priorities for the lines, these are then regarded as the composite indices \( C_{IAHP} \) describing the integrity of the lines.

In Fig. 11, the FCI values are compared with the scores of SAW and AHP approaches. The values of the three techniques are quite close especially for line 1 and line 5, where failure frequency, inspection remarks and punctuality show extreme status. It is obvious from Fig. 11 that very similar result will be obtained if the lines were to be ranked based on their integrity using the scores from the three techniques. However, the values obtained using the SAW technique are notably high in some instances due to the problem of compensability (deficit in one dimension is compensated by a surplus in another). Furthermore, the normalisation employed in the SAW approach gives a normalised value of zero to lines with least indicator grade, thus leading to remarkably low values of \( C_{SAW} \). The priority value of the AHP technique is appropriate for ranking the lines, but the computation requires review if the values are to be considered as integrity measure of the lines whose evolution is to be analysed. Considering the purpose of the study, FCI approach gives a reliable integrity measure of the lines, since the integrity measure of any line is not relative to other lines, and thus can be monitored over the years. Also the problem of trade-offs or compensability is reduced.

However, the quality of the FCI depends on the experience of the experts and the quality of the data used. There is room for improvement of the quality of the data used in this study. The indicator for punctuality can be extended to cover incidences of cancelled trains due to infrastructure failures. Another important aspect is the need to standardise the inspection strategy in terms of frequency, details and priority classification of inspection remarks on all the lines for reliability sake.

6 Conclusions

In this study, we have demonstrated the application of FIS in computing a composite indicator to relate maintenance and renewal function to capacity situation and also to enhance decision making. The proposed FCI will facilitate the assessment of M&R in terms of infrastructure and traffic performance. This information will support efficient and effective strategic decision making and a long-term infrastructure management plan to increase the operational capacity and service quality of a network. The concluding remarks on the case study presented are as follows:
• Line 5 has consistently high FCI value that could be considered as effective maintenance and renewal (M&R) and readiness to accommodate more traffic if other conditions are met. The integrity of line 4 is considerably good in relation to the average capacity utilisation, however, if the operational capacity is to be further increased, additional M&R measures could be required.

• Line 1 has an undistiguished low FCI value probably because of its heavy operational profile and perhaps inadequate M&R due to a lack of time to compensate for it. This is an indication for a review of the M&R strategy to meet the demanding heavy haul on the line.

• Lines 2 and 3 exhibit average FCI over the years, most likely due to low M&R efforts owing to the low capacity consumption. Increasing the traffic volume will require a raise in the M&R efforts to maintain a high service quality.

• FCI is a better indication of the line condition than the failure frequency which is the conventional indicator used widely in railway maintenance management.

In future work, the reliability of the proposed indicator would be improved by considering other relevant simple indicators and by applying fuzzy AHP technique for the aggregation.

Acknowledgments We would like to express our sincere gratitude for the financial support of Trafikverket and Luleå Railway Research Centre. Also, we would like to acknowledge Mr Lars Wikberg and Dr Arne Nissen for their support in data collection, as well as Prof. Ashok Deshpande for his technical support.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Maintenance analysis for continuous improvement of railway infrastructure performance

Maintenance analysis for continuous improvement of railway infrastructure performance

Stephen M. Famurewa*1, Matthias Asplund1, Matti Ranta-talo1,2, Aditya Parida3 and Uday Kumar3,4
1Division of Operation & Maintenance Engineering, Luleå University of Technology, 97187 Luleå, Sweden
2Luleå Railway Research Centre, 97187 Luleå, Sweden

(Rceived 20 November 2013; final version received 12 March 2014; accepted 31 March 2014; published online 27 February 2015)

Railway transport system is massive and complex, and as such it requires effective maintenance to achieve the business goal of safe, economic and sustainable transportation of passengers and goods. The growing demand for improved service quality and capacity target by railway infrastructure managers requires appropriate maintenance analysis to facilitate continuous improvement of infrastructure performance. This paper presents the application of risk matrix as a maintenance analysis method for the identification of track zones that are bottlenecks that limit operational capacity and quality. Furthermore, an adapted criticality analysis method is proposed to create a hierarchical improvement list for addressing the problem of train mission interruption and reduced operational capacity. A case study of a line section of the Swedish network is presented. The result classifies the zones on the line section into different risk categories based on their contribution to loss of capacity and punctuality. In addition, an improvement list for the lower-level system is presented to facilitate maintenance decisions and continuous improvement at both operational and strategic levels.

Keywords: railway capacity; quality of service; maintenance performance indicators; maintenance decision support; bottleneck; risk assessment; maintenance analysis; improvement

Introduction

The railway system is a complex system with rolling stock, fixed infrastructure and other assets. The aggregated performance of these systems is directed towards achieving the business goal of transporting passengers and goods in an efficient, safe and environmentally friendly way. Moreover, the management of this system is not a trivial task in the presence of organisational challenges and technical issues in addition to ambitious targets to be met by the operation of the system.

The technical demand on the railway system is increasing; there is a drive for an increase in speed, axle load, volume of traffic and other essential operational requirements (Ekberg & Paulsson, 2010). However, if the existing infrastructure is to meet the capacity demand without compromising the quality of service, the maintenance process should be subjected to continuous improvement. In manufacturing and maintenance processes, generic philosophies, such as the lean concept, total productive maintenance (TPM), total quality management, business process re-engineering, kaizen, Deming cycle (plan-do-study-act) and theory of constraints, have been studied and deployed for continuous improvement (Dale et al., 2007; Dettmer, 1997; McKone, Schroeder, & Cua, 1999; Nakajima, 1988; Smith & Hawkins, 2004). Furthermore, common tools used in TPM and lean maintenance include the 5-S tool, value stream mapping, just-in-time, the Kanban pull system and other quality management tools, such as Pareto chart, ANOVA, control chart, cause and effect diagram (Dale et al., 2007; Nakajima, 1988; Smith & Hawkins, 2004).

An adequately designed performance measurement (PM) system is a management and improvement tool that can be used as a basis for decision-making by the strategic, tactical and operational levels of management (Bititici, 1997). It is not only limited to the measurement of performance but also extends to evaluation and comparison of performance, diagnosing strengths and weaknesses, setting objectives and targets, facilitating improvements, sharing results in order to inform and motivate people and control of progress and changes over time (EN 15341, 2007).

Furthermore, an essential aspect of performance management of railway infrastructure is the systematic maintenance analysis of the network at various levels with relevant performance measures. This is the core of any continuous improvement programme in the railway industries. Failure Mode, Effects and Criticality Analysis (FMECA) is a commonly used technique for maintenance analysis; the requirements and procedures for performing FMECA are established and presented in (MIL-STD-
Multi-criteria criticality analysis for effective maintenance priority ranking of engineering asset has been studied and presented by Braglia (2000) and Marquez (2007). Moreover, the review of some critical aspects of risk analysis important for the successful implementation in maintenance as well as the use of risk analysis for the selection and prioritisation of maintenance activities has been carried out (Aven, 2008; Misra, 2008). In the railway industries, an analysis method for prioritising maintenance actions for railway infrastructure has been presented (Nyström & Söderholm, 2010). Risk evaluation technique has been developed for the specification and demonstration of reliability, availability, maintainability and safety (RAMS) of railway systems (EN 50126, 1999).

The problem of applying reliability-centred maintenance (RCM) to large-scale railway infrastructure networks to achieve an efficient and effective maintenance concept has been addressed, and a toolkit to perform relevant RCM analysis has been implemented (Carretero et al., 2003).

The growing demand for improved service quality and capacity targets by railway infrastructure managers (IMs) is a challenge that requires adapted maintenance analyses for isolating and eliminating the weakest link or bottlenecks. The objective of this paper was to study maintenance analysis methods and performance indicators for different infrastructure indenture levels as required for continuous improvement in the railway industries. The contribution of this paper includes the application of risk matrix for the classification of railway systems into risk categories and the presentation of an adapted criticality analysis method for the generation of an improvement list for assemblies and systems based on the weakest link theory.

The organisation of the paper is as follows: the next section provides a brief description of railway infrastructure management and the following section explains maintenance performance measurement (MPM) with a perspective related to maintenance analysis. The subsequent section presents maintenance analysis methods useful for continuous improvement at different indenture levels of railway infrastructure. The ‘Case study’ and ‘Results’ sections present a case study on the analysis of failure and delay data for a line section. The ‘Conclusion’ section presents the findings, suggestions and concluding remarks of the study.

**Performance measurement system**

An MPM system is a management tool that can be used to identify and facilitate strategic improvements in railway infrastructure management. PM systems have been implemented in railway infrastructure management for the following reasons: to improve maintenance strategy,

- ensuring the quality, safety, reliability, maintainability and availability of the infrastructure.

The visions of IMs are mostly created from the above responsibilities, and their strategic objectives are informed by the visions in addition to other stakeholders’ requirements. Figure 1 connects the overall business plan of infrastructure management with the maintenance function. The maintenance service is provided by either internal or external agents. For example, the budget for operation and maintenance of railway infrastructure by the Swedish Transport Administration is around 15% of the total budget from Swedish national transport strategic plan (Trafikverket, 2012). With such input, the maintenance process is anticipated to reach the set objectives, which will in turn contribute towards the overall IM objectives. This definitely requires a robust MPM system that will assess the outcome of maintenance. The phases of a robust MPM system include design of measures, data collection, analysis, improvement and control.

**Railway infrastructure management**

The management of railway infrastructure by public or private organisations often covers the following responsibilities (Alexandersson & Hultén, 2008):

- developing and allocating the capacity of the infrastructure;
- handling the train traffic control function on the infrastructure and
measure value created by maintenance, justify investment or renewal, support future decision, improve resource allocations, check compliance with safety recommendations and standard benchmarks and follow maintenance contracts (Parida & Chattopadhyay, 2007). Examples of integrated PM frameworks that are in use in different industries and also applicable in railway industries include DuPont pyramid of financial ratios, SMART pyramid, PM matrix, 'tableau de bord', balanced score card and performance prism (Kennerley & Neely, 2002; Parida & Chattopadhyay, 2007).

The frameworks are often structured into categories reflecting distinct perspectives of the overall business plan, which include finance, customer, internal business process, learning and growth, quality, productivity, flexibility, quality of work life, innovation, health safety and environment, employee satisfaction, resource utilisation, delivery and so on (Åhrén & Parida, 2009; Ghaleyini, Noble, & Crowe, 1997; Medori & Steeple, 2000; Neely, 2005; Parida & Chattopadhyay, 2007).

Hierarchical structure of PM system

A PM system should be presented clearly to show the link from the overall business goal to measurable operational activities. Figure 2 shows a simple interconnection of goals from the strategic or top management level to the operational level in an effective PM system. In order to measure the contribution of maintenance to the overall transport objective and identify improvement potential, clear maintenance performance indicators (MPIs) are essential. These indicators could be related directly to the condition of the asset or contribution of maintenance to operational quality and quantity.

In addition, the following maintenance result area or performance criteria have been identified for railway infrastructure maintenance: safety, dependability, sustainability, punctuality, robustness, maintenance cost, possession time, passenger comfort or ride quality and resource utilisation (Åhrén & Parida, 2009; Kumar, Galar, Parida, Stenström, & Berges, 2013; Medori & Steeple, 2000; Parida & Chattopadhyay, 2007; Stenström, Parida, & Galar, 2012; Tsang, 1998).

MPM system

For the purposes of maintenance analysis and continuous improvement of railway infrastructure performance, an MPM system for railway infrastructure is synthesised from the literature and presented in Table 1 (Åhrén & Parida, 2009; Kumar et al., 2013; Medori & Steeple, 2000; Parida & Chattopadhyay, 2007; Stenström et al., 2012; Tsang, 1998). Relevant strategic perspectives, maintenance result areas and performance indicators are presented and related in the table.

However, for an MPM system to be a useful tool to facilitate improvement, it is important to deploy relevant data-driven analysis and interpretation method using historical indicator data. Such analysis is the link required for the transformation of collected data into information and knowledge base to support decision-making and continuous improvement.

Theoretical background on maintenance analysis for continuous improvement

The demand on present infrastructure management requires that the maintenance function should adequately support the operational capacity and performance of a network, the constituting routes, line sections, traffic zones, systems and so on. Therefore, this makes systematic maintenance analysis of a railway network at the relevant level the core of any continuous improvement programme for railway infrastructure performance. Such maintenance analysis emphasises the risk of failure to be the main driver of maintenance decisions, rather than the failure itself.
Thus, failures with critical consequence on operation, economic or safety are given attention. Figure 3 shows the system breakdown structure of a corridor and the relevant maintenance analyses required to develop an effective maintenance programme that will support the design performance of a railway network.

Maintenance analysis required for systematic decision-making and continuous improvement can be carried out on a higher indenture level of railway infrastructure using quantitative indicators of its degree of credibility from the market perspective. It can also be performed from a technical point of view using multi-criteria criticality analysis of routes and line sections. This analysis often involves the aggregation of different indicators, using techniques such as the linear aggregation method, geometric method, analytical hierarchy process (AHP), fuzzy logic and other multi-criteria aggregation techniques. The maintenance analysis of higher-level systems to identify infrastructure bottlenecks in terms of train mission interruption and capacity limitation can also be performed using a suitable risk assessment technique. Further analyses down the indenture level include fault tree analysis, failure mode and effects analysis, root cause analysis, adapted analysis method, reliability and maintainability analyses among others. These maintenance analysis methods have been addressed in various literatures (Braglia, 2000; Carretero et al., 2003; EN 50126, 1999; IEC 60300-3-1, 2003; IEC 60300-3-14, 2004; IEC 31010, 2009; Marquez, 2007; MIL-STD-1629A, 1980; Misra, 2008; Nyström & Söderholm, 2010). The selection of a suitable analysis method should be based on the purpose of analysis, available data, dependability and system engineering requirements. The maintenance analysis presented in this paper is aimed at isolating and eliminating the weakest link or bottleneck, which are constraints to both capacity and quality of service on the studied line section. It will support homogenisation of technical performance along the section and thereby improve the operational capacity.

Identification of infrastructure bottlenecks using risk matrix

The case study presented in the subsequent section adapted the risk assessment procedure for the specification and demonstration of RAMS (EN 50126, 1999) and the cause and effect diagram for the analysis and improvement of the performance of a higher-level system. The risk assessment matrix given in Table 2 is used to develop a risk matrix to categorise each traffic zone in the case study into one of four possible risk categories. This provides resourceful information for the prioritisation of maintenance activities during planning and scheduling.
Adapted criticality analysis method for improvement of line capacity

Maintenance analysis can be extended to lower-level systems to support operational level decision-making. To address the problem of train mission interruption and reduced operational capacity caused by poor infrastructure integrity, a data-driven approach should be adopted. The analysis method proposed for homogenising the railway line condition and reducing traffic interruption is adapted from the theory of constraints (Dettmer, 1997) and existing criticality analysis methods (Aven, 2008; Carretero et al., 2003; Márquez, 2007). The objective is to identify restraining systems and present a hierarchical list for improvement because the capacity and integrity of a line are not stronger than the weakest section.

For this purpose, three parameters are selected: the probability that system failure would result in an operational consequence (POC), total delay consequence of a system failure in a year (TD), and average risk number of the parent higher-level system (see Equations (1)–(3)). Furthermore, three approaches: a geometric method, normalised linear method, and normalised geometric method, which are expressed in Equations (5)–(7), are used to aggregate the parameters. Equation (8) is then used to integrate the scores of the three aggregation methods to give an overall improvement score. The list of notations used in Equations (1)–(8) is given in Table 3.

\[
P_{\text{failure}}^{\text{operational consequence}} = \frac{F_a}{F_a} ,
\]

\[
TD_a = \sum_{i=1}^{m} d_i .
\]

The characteristic risk number of zone \( z \) over a period of \( T \) years is estimated as the average risk number using the geometric mean, given by

\[
RF_z = \left( \prod_{t=1}^{T} RF_t \right) \left( \sum_{t=1}^{T} r_t \right) .
\]

For the linear aggregation, the min–max normalisation method shown in Equation (4) is used to normalise the parameters before aggregation using Equation (6). On the other hand, the geometric method uses the actual value of the selected parameter for aggregation according to Equation (5), whereas the normalised value from Table 4 is used for the normalised geometric method according to Equation (7) as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>POC &lt;sub&gt;a&lt;/sub&gt;</td>
<td>Probability of operation consequence for assembly ( a )</td>
</tr>
<tr>
<td>( f_a )</td>
<td>Frequency of mission interruption failure for assembly ( a )</td>
</tr>
<tr>
<td>( C_{\text{ave}} )</td>
<td>Total delay owning to the failure of assembly ( a ) in zone ( z )</td>
</tr>
<tr>
<td>( RF_z )</td>
<td>Average risk number for zone ( z ) over ( T ) years</td>
</tr>
<tr>
<td>( C_{\text{ave}} )</td>
<td>Risk number for zone ( z ) during year ( t )</td>
</tr>
<tr>
<td>( N_{\text{ave}} )</td>
<td>Value of parameter ( j ) for assembly ( a ) in zone ( z )</td>
</tr>
<tr>
<td>( N_{\text{ave}}^{\text{max}} )</td>
<td>Normalised value of parameter ( j ) for assembly ( a ) in zone ( z )</td>
</tr>
<tr>
<td>( w_j )</td>
<td>Weight of parameter ( j ) and weight of method ( m ) (equal weight is assumed in the case study)</td>
</tr>
<tr>
<td>( S_{\text{ave}} )</td>
<td>Score of assembly ( a ) in zone ( z ) using method ( m )</td>
</tr>
<tr>
<td>OIS &lt;sub&gt;a&lt;/sub&gt;</td>
<td>Overall improvement score for assembly ( a ) in zone ( z )</td>
</tr>
</tbody>
</table>

Table 2. Typical risk evaluation matrices (Carretero et al., 2003; EN 50126, 1999).

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Likely impact or consequence of failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent</td>
<td>Insignificant</td>
</tr>
<tr>
<td>Probable</td>
<td>Undesirable</td>
</tr>
<tr>
<td>Occasional</td>
<td>Tolerable</td>
</tr>
<tr>
<td>Remote</td>
<td>Negligible</td>
</tr>
<tr>
<td>Improvable</td>
<td>Negligible</td>
</tr>
</tbody>
</table>

Table 3. List of notations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>POC &lt;sub&gt;a&lt;/sub&gt;</td>
<td>Probability of operation consequence for assembly ( a )</td>
</tr>
<tr>
<td>( f_a )</td>
<td>Frequency of mission interruption failure for assembly ( a )</td>
</tr>
<tr>
<td>( C_{\text{ave}} )</td>
<td>Count of reports of failure symptom and events for assembly ( a )</td>
</tr>
<tr>
<td>( TD_{\text{ave}} )</td>
<td>Total delay owning to the failure of assembly ( a ) in zone ( z )</td>
</tr>
<tr>
<td>RF &lt;sub&gt;z&lt;/sub&gt;</td>
<td>Average risk number for zone ( z ) over ( T ) years</td>
</tr>
<tr>
<td>( C_{\text{ave}} )</td>
<td>Total delay owning to the failure of assembly ( a ) in zone ( z )</td>
</tr>
<tr>
<td>( N_{\text{ave}} )</td>
<td>Average risk number for zone ( z ) during year ( t )</td>
</tr>
<tr>
<td>( N_{\text{ave}}^{\text{max}} )</td>
<td>Value of parameter ( j ) for assembly ( a ) in zone ( z )</td>
</tr>
<tr>
<td>( w_j )</td>
<td>Weight of parameter ( j ) and weight of method ( m ) (equal weight is assumed in the case study)</td>
</tr>
<tr>
<td>( S_{\text{ave}} )</td>
<td>Overall improvement score for assembly ( a ) in zone ( z )</td>
</tr>
</tbody>
</table>

Table 4. Normalisation of POC, RF and TD.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Category</th>
<th>Value</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>POC &lt;sub&gt;a&lt;/sub&gt;</td>
<td>Intolerable</td>
<td>4</td>
<td>&gt;2000</td>
</tr>
<tr>
<td>Probable</td>
<td>Undesirable</td>
<td>3</td>
<td>1000–1999</td>
</tr>
<tr>
<td>Occasional</td>
<td>Tolerable</td>
<td>2</td>
<td>500–999</td>
</tr>
<tr>
<td>Remote</td>
<td>Negligible</td>
<td>1</td>
<td>200–499</td>
</tr>
<tr>
<td>Improvable</td>
<td>Tolerable</td>
<td>2</td>
<td>100–199</td>
</tr>
<tr>
<td>Incredi</td>
<td>Tolerable</td>
<td>1</td>
<td>0–49</td>
</tr>
</tbody>
</table>

Structure and Infrastructure Engineering
• Min–max normalisation technique:

\[
N_{sc}^{'} = \frac{C_{sc} - \min(C_j)}{\max(C_j) - \min(C_j)}
\]  

(4)

• Geometric method of aggregation:

\[
S_{sc}^{GM} = \prod_{j=1}^{l} C_{scj}^{m}.
\]  

(5)

• Normalised linear method of aggregation:

\[
S_{sc}^{NM} = \sum_{j=1}^{l} w_j N_{scj}^{m}.
\]  

(6)

• Normalised geometric method of aggregation:

\[
S_{sc}^{NGM} = \prod_{j=1}^{l} N_{scj}^{m}.
\]  

(7)

In order to combine the outcome of the aggregation techniques to present a single hierarchical list, the normalised criteria distance approach proposed by Pham (2013) is adapted. The weight of the nth aggregation technique is denoted by \(w_m\), the score of assembly \(a\) in zone \(Z\) using technique \(m\) is denoted by \(S_{sc}^{m,a}\) and the total number of assemblies in zone \(Z\) is \(A\), while the total number of zones is \(Z_T\). The distance of the normalised aggregation scores from the origin for an assembly \(a\) in zone \(Z\) is defined as the overall improvement score, given by

\[
OIS_{sc} = \sum_{m=1}^{m} \left( \frac{S_{sc}^{m,a}}{S_{sc}^{m,a} + \sum_{m=1}^{m} S_{sc}^{m,a}} \right)^2 \times w_m
\]

\(a \in A\) and \(z \in Z\)  

(8)

### Case study

To facilitate improvement, the dependability and punctuality aspects of maintenance performance are analysed in a case study. A line section on the network of the Swedish Transport Administration (Trafikverket) is considered. A brief description of the classification of the railway network belonging to the administration is shown in Table 5. The table gives an overview of the challenges confronting maintenance management in terms of capacity utilisation, difficulty getting track possession time and the requirement for service quality.

The line section studied is 168 km, from Boden to Gallivare, and it is the longest section (about 39% of the track length) on the heavy haul line, ‘Malmbanan’, which belongs to line class 2. To meet the requirements in Table 5, it is essential to analyse the performance of past maintenance actions and decisions using a system of indicators. This will create room for improvement. The line described in the case study is divided into 39 traffic zones or segments representing the technical divisions referred to as traffic zones by the Swedish Transport Administration. The zones are established partitions used in traffic operational management by the administration. The 39 zones consist of: (1) operational areas and (2) lines joining two operational areas. The operational areas include station areas where interlocking/signalling boxes are used to control the movement and path of trains within its areas. It could also be a train stop point where passengers board and alight.

The data analysed are the record of all train delays on the chosen line section between 2010 and 2012. In line with the objective of this paper, the analysis is basically focused on mission or traffic interruptions that are related to infrastructure failure. An overview of the data presented in Figure 4 shows the failure characteristics of the line and the operational consequence in terms of delay minutes for 3 years.

### Results and discussion

**Analysis of mission interruption and operational capacity**

Train delays in for the year 2012 on the line section under consideration with associated causes are presented in Figure 5. Unclassified causes account for the largest proportion of train delay, about 27% in 2012. A significant observation is the impact of infrastructure failure, which is about 21% of the train delay in the same year. It is

### Table 5. Line classes in Trafikverket and their description.

<table>
<thead>
<tr>
<th>Line class</th>
<th>Transport value</th>
<th>Capacity utilisation</th>
<th>Difficulty for track maintenance time</th>
<th>Traffic sensitivity to disturbance</th>
<th>Requirement on punctuality</th>
<th>Requirement on comfort</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Very important</td>
<td>High</td>
<td>Very high</td>
<td>Very high</td>
<td>Very high</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>Very important</td>
<td>Medium to high</td>
<td>High</td>
<td>High</td>
<td>Very high</td>
<td>High</td>
</tr>
<tr>
<td>3</td>
<td>Important</td>
<td>Medium</td>
<td>Average to high</td>
<td>Medium</td>
<td>Basic</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>Less important</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Basic</td>
<td>Basic</td>
</tr>
<tr>
<td>5</td>
<td>Less important</td>
<td>Low</td>
<td>Very low</td>
<td>Basic</td>
<td>Basic</td>
<td>Basic</td>
</tr>
</tbody>
</table>

Notes: Transport value is the significance of the line from socio-economic perspective. Safety requirement for all the line classes is the same, i.e. very high.

*Detailed description of the categories can be found in Söderholm and Norrbin (2013).*
interesting to see the contribution of the categories to monthly mission interruption on the line section. The plot to the right in Figure 5 shows that unclassified factors and secondary causes contribute at least 20% of total monthly delay on the section in non-extreme situations or average occasions and contribute more than 30% in extreme situations. Classifying these two categories into the actual causes of failure will support comprehensive analysis. Furthermore, infrastructure failure has a delay consequence that is widespread from low to high; this typifies the expected randomness of failure event. In addition, it shows the variation in maintenance performance, operation profile and weather conditions over the months of the year.

The monthly performance of the infrastructure in 2012 in terms of daily mission interruption is presented in Figure 6. The figure gives a good overview of how scheduled daily train mission profiles have been implemented and limitations related to infrastructure failure. The worst monthly performance is in July, where 48% of the scheduled daily train mission was successful without infrastructure failure, i.e. there was an infrastructure-related incident that interrupted traffic every other day in the month of July. On the other hand, the best monthly performance of 93% was recorded in September, where daily traffic interruption was reported only in 2 days. This shows the need for improvement in the system, operating or maintenance conditions.

Figure 7 depicts the train mission interruption due to infrastructure failure and the contribution of each assembly/system. Switches and crossings (S&C), track circuit and track contribute the most to the number of mission interruptions or failures on the line section. On the other hand, overhead cable, track circuit and track have the most operational consequence in terms of the delay time. However, this is aggregate information for the infrastructure types which neither gives specific information for decision-making nor facilitates effective action steps for continuous improvement of the line.

Maintenance analysis using Pareto chart
In order to meet the operational requirement of high punctuality and capacity utilisation, further analysis is carried out on the zone level. This analysis is promising because these traffic zones become the system of interest and reasonable improvement potential can be suggested. Among several benefits of such a case study is the opportunity to identify critical areas that are bottlenecks to the fulfilment of availability and operational capacity goals on a line section or network. In addition, it helps to identify sections that are in the worst condition on the line, contributing to a higher risk of quality loss and capacity reduction.
A graphical representation of the frequency of traffic interrupting failure on the 39 zones on the line described in the previous section is shown in Figure 8. This is useful for identifying critical locations that contribute significantly to the capacity and punctuality problems on the line. The Pareto chart shows that approximately 30% of the traffic impacting failure occurred in zones 13, 3 and 10, while about half of the zones were responsible for 80% of the train mission interruption on the line in the year 2012.

However, it is essential to investigate the operational consequence of these failures because there are many subsystems or items that can fail. Figure 9 shows train delays as a result of failures in each of the traffic zones. About 40% of the delays caused by infrastructure failure could be traced to zones 18–20, 8–9, 12–13 and 2–3. In addition, a third of the zones are critical based on the Pareto 80% estimation of significance. Comparing Figures 8 and 9, it could be seen that the hierarchical listing of the state of the zones differs from the failure frequency and delay consequence perspectives.

### Maintenance analysis using risk matrix

The information presented separately in Figures 8 and 9 can be aggregated to extract information about critical spots and bottlenecks that are contributing significantly to the risk of quality loss and capacity reduction. To facilitate effective decision-making, it is necessary to aggregate these two condition indicators in an index or with visual representation. For further analysis and to classify each traffic zone into risk categories, the risk matrix described in previous section is used.

![Figure 7. Contribution of each infrastructure type to train mission interruption.](image)

![Figure 8. Pareto chart of higher-level system failure or traffic zones in 2012.](image)
First, the mean delay time on each zone is estimated using statistical distribution. According to the literature, random events, such as delay times can be modelled using exponential, gamma, lognormal or Weibull distribution (Yuan, 2006). It is assumed that the delay time considered in this study (failure-related delay) can be modelled adequately using lognormal distribution. Using the Anderson Darling’s goodness-of-fit test, lognormal distribution ranks among the best distributions with good fit. The general acceptability and simplicity of lognormal distribution were also considered in its choice.

A hypothetical grading of the frequency of failure and consequential delay was performed to demonstrate the mentioned methodology. This grading could be a subjective judgement of a group of experts within the IM organisation, and it should reflect both the acceptance limit and quality definition of the organisation. The contribution of each traffic zone to the overall risk of quality loss is presented in Figure 10 for the year 2011 and Figure 11 for the year 2012. Figure 10 shows that traffic zones 8, 7, 2–3 and 6–7 have intolerable contributions, whereas 15, 17, 15–16 and 10–11 have tolerable contributions. All other zones on the matrix have undesirable contributions with the exception of zone 4 with negligible impact in the year 2011.

As a result of re-investment and other intervention measures deployed in 2012, the integrity of the line improved and the condition of the zones changed. The risk matrix for

![Figure 9. Pareto chart of delay consequence of higher-level system failure in 2012.](image)

![Figure 10. Categorisation of the traffic zones based on risk of limiting service quality and capacity in 2011.](image)
the year 2012 presented in Figure 11 identified two zones, 13 and 18–20, to be of intolerable risk contribution, the risk category of other zones can be seen in the figure. Zones that have less than three mission-interrupting failures are not classified into risk category because they are not considered characteristic behaviour but rather incidental.

The risk matrix gives information about the risk category of each traffic zone in a certain period and identifies critical spots or bottlenecks along the linear asset where the capacity of the entire line is limited. The intolerable zones in Figures 10 and 11 are the critical spots that restrain the flow of traffic through the line.

Figure 11. Categorisation of the traffic zones based on the risk of limiting service quality and capacity in 2012.

Figure 12. Root cause analysis of infrastructure bottleneck.
Furthermore, to facilitate continuous improvement, it is necessary to develop a practical routine for additional maintenance analysis of the zones in the intolerable and undesirable risk categories. Figure 12 is a cause and effect diagram that could be used to capture the factors contributing to the intolerable state of the zones and their root causes. Zones with an undesirable risk contribution should be maintained to reduce their impact. Zones with a tolerable impact should be controlled with necessary measures, and those with negligible impact should be observed and their maintenance ought to be standardised.

**Adapted analysis method for improvement of line capacity**

The maintenance analysis is extended to address the problem of train mission interruption and reduced operational capacity caused by poor infrastructure integrity. The analysis method described in the theory section and Equations (1)–(8) is used to score the various systems and assemblies in the zones for the purpose of priority ranking for maintenance intervention. The result of the computation for the case study is shown in Table 6. It can be seen that the resulting priority ranking of the aggregation methods differs in some cases. The overall improvement scores combine the information from the aggregation techniques and provide final ranks that can be interpreted as a measure of maintenance need on the line for the year 2013. The outcome of the analysis is a hierarchical list of the lower-level system based on their contribution to the reduction of the operational capacity and punctuality of the line. The systems on top of the priority list (overhead cable on zone 18–20, track circuit on zone 15 and alternative power line on zone 12–13) are

<table>
<thead>
<tr>
<th>Traffic zone</th>
<th>Assembly/system</th>
<th>POC</th>
<th>Total delay</th>
<th>Risk number</th>
<th>$S_{GM}^{(1)}$</th>
<th>$S_{LM}^{(2)}$</th>
<th>$S_{NGM}^{(3)}$</th>
<th>OIS$^{(4)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–20</td>
<td>Overhead cable</td>
<td>0.27</td>
<td>2677</td>
<td>2.52</td>
<td>1840$^{(5)}$</td>
<td>0.57$^{(6)}$</td>
<td>50.41$^{(7)}$</td>
<td>0.1464$^{(8)}$</td>
</tr>
<tr>
<td>17</td>
<td>S&amp;C</td>
<td>0.28</td>
<td>40</td>
<td>1.26</td>
<td>14$^{31}$</td>
<td>0.07$^{30}$</td>
<td>2.5$^{26}$</td>
<td>0.0050$^{27}$</td>
</tr>
<tr>
<td>17–18</td>
<td>Track circuit</td>
<td>0.48</td>
<td>98</td>
<td>1.73</td>
<td>81$^{19}$</td>
<td>0.22$^{24}$</td>
<td>69.5$^{19}$</td>
<td>0.0156$^{24}$</td>
</tr>
<tr>
<td>16–17</td>
<td>Track circuit</td>
<td>0.48</td>
<td>66</td>
<td>1.94</td>
<td>62$^{20}$</td>
<td>0.24$^{23}$</td>
<td>7.8$^{17}$</td>
<td>0.0168$^{23}$</td>
</tr>
<tr>
<td>16</td>
<td>Track</td>
<td>0.46</td>
<td>300</td>
<td>1.94</td>
<td>268$^{20}$</td>
<td>0.27$^{21}$</td>
<td>15.6$^{16}$</td>
<td>0.0307$^{13}$</td>
</tr>
<tr>
<td>2</td>
<td>Cable lines</td>
<td>1.00</td>
<td>96</td>
<td>1.78</td>
<td>171$^{13}$</td>
<td>0.42$^{8}$</td>
<td>14.1$^{14}$</td>
<td>0.0310$^{12}$</td>
</tr>
<tr>
<td>14</td>
<td>No fault found</td>
<td>0.21</td>
<td>24</td>
<td>1.41</td>
<td>3$^{34}$</td>
<td>0.06$^{33}$</td>
<td>1.4$^{34}$</td>
<td>0.0039$^{34}$</td>
</tr>
<tr>
<td>13–14</td>
<td>Track circuit</td>
<td>0.48</td>
<td>336</td>
<td>1.94</td>
<td>314$^{4}$</td>
<td>0.28$^{19}$</td>
<td>15.6$^{16}$</td>
<td>0.0330$^{12}$</td>
</tr>
<tr>
<td>12–13</td>
<td>Alternative power line</td>
<td>0.33</td>
<td>884</td>
<td>3.00</td>
<td>884$^{4}$</td>
<td>0.44$^{6}$</td>
<td>36.0$^{3}$</td>
<td>0.0803$^{3}$</td>
</tr>
<tr>
<td>7</td>
<td>Track circuit</td>
<td>0.48</td>
<td>349</td>
<td>3.00</td>
<td>481$^{6}$</td>
<td>0.42$^{8}$</td>
<td>24.0$^{6}$</td>
<td>0.0505$^{6}$</td>
</tr>
<tr>
<td>3</td>
<td>No fault found</td>
<td>0.21</td>
<td>16</td>
<td>1.94</td>
<td>6$^{5}$</td>
<td>0.14$^{31}$</td>
<td>1.9$^{17}$</td>
<td>0.0077$^{14}$</td>
</tr>
<tr>
<td>10</td>
<td>Interlocking</td>
<td>0.23</td>
<td>55</td>
<td>3.15</td>
<td>39$^{3}$</td>
<td>0.32$^{17}$</td>
<td>6.3$^{2}$</td>
<td>0.0187$^{20}$</td>
</tr>
<tr>
<td>13</td>
<td>No fault found</td>
<td>0.21</td>
<td>8</td>
<td>3.63</td>
<td>6$^{5}$</td>
<td>0.37$^{14}$</td>
<td>3.6$^{7}$</td>
<td>0.0202$^{19}$</td>
</tr>
<tr>
<td>15</td>
<td>Track circuit</td>
<td>0.48</td>
<td>50</td>
<td>3.63</td>
<td>87$^{18}$</td>
<td>0.48$^{2}$</td>
<td>7.3$^{18}$</td>
<td>0.0276$^{14}$</td>
</tr>
<tr>
<td>9</td>
<td>Track circuit</td>
<td>0.48</td>
<td>1066</td>
<td>2.18</td>
<td>1112$^{12}$</td>
<td>0.40$^{10}$</td>
<td>34.9$^{9}$</td>
<td>0.0920$^{7}$</td>
</tr>
</tbody>
</table>

Note: Only traffic zones with failure frequency more than 3 in 2012 are presented.
considered the weakest links or restraining assemblies or systems. These should be prioritised in maintenance planning to improve the operational capacity and quality on the line section.

Conclusions

This paper has addressed systematic maintenance analyses for different infrastructure indenture levels required for developing an effective maintenance programme to support design performance of a railway network. A synthesised system of indicators relevant for maintenance analysis for the continuous improvement of railway infrastructure performance has been presented. The concluding remarks for the maintenance analyses carried out in the case study are as follows.

1. Infrastructure failure has a significantly high impact on the operational capacity and train punctuality on the line.
2. The quantification of maintenance need and hierarchical listing of the traffic zones differ when failure frequency and consequential delay impacts are used for separate analysis. The presented risk matrix is, however, useful for aggregating the two maintenance measures to support decision-making.
3. Traffic zones 8, 7, 2–3 and 6–7 are capacity bottlenecks in 2011, and zones 13 and 18–20 fall into the same category in 2012 due to their intolerable contribution to the reduction in service quality and line capacity. The proposed cause and effect diagram can be of use for initiating investigation into the causes of these bottlenecks at the operational level.
4. The outcome of the adapted criticality analysis method for continuous improvement shows that the overhead cable on zone 18–20, track circuit in zone 15 and alternative power line in zone 12–13 are the weakest links or restraining assemblies. These should be prioritised in maintenance planning for the year 2013 to improve the operational capacity and service quality on the line section.
5. Finally, the decision on the distribution of maintenance and reinvestment budgets can be supported with these analyses.

For future research, a holistic approach that extends the maintenance analysis from the topmost indenture level to the lowest maintainable items on a network will be considered. Other relevant MPIs will be included as criteria in the analysis. This will not limit the perspective and application of the analysis to punctuality and capacity improvement at the operational level but rather extend to the safety and economic aspects.

Acknowledgements

The authors acknowledge the financial support of Trafikverket and Luleå Railway Research Centre (JVTG) and also the technical support of Mr Lars Wikberg of Trafikverket.

Notes

1. Email: mathias.asplund@ltu.se
2. Email: matti.rantatalo@ltu.se
3. Email: aditya.parida@ltu.se
4. Email: uday.kumar@ltu.se

References


Optimisation of maintenance track possession time:
a tamping case study

Optimisation of maintenance track possession time: A tamping case study

Stephen M Famurewa1, Tao Xin1,3, Matti Rantatalo1,2 and Uday Kumar1,2

Abstract
Optimum allocation and efficient utilisation of track possession time are becoming important topics in railway infrastructure management due to increasing capacity demands. This development and other requirements of modern infrastructure management necessitate the improvement of planning and scheduling of large-scale maintenance activities such as tamping. It is therefore necessary to develop short-, medium- and long-term plans for performing tamping on a network or track section within a definite time horizon. To this end, two key aspects of infrastructure maintenance planning are considered in this paper; deterioration modelling and scheduling optimisation. An exponential deterioration function is applied to model the geometry quality of a series of 200 m segments of a 130 km line section, and an empirical model for recovery after tamping intervention is developed. These two models are subsequently used to generate a methodology to optimise a schedule for tamping intervention by minimising the total cost of intervention including the cost of track possession while geometry quality is ascertained to be within a desirable limit. The modelling considers two types of tamping interventions, preventive and corrective, with different intervention limits and tamping machines. The result of this paper suggests a tamping plan which will lead to optimum allocation of track possession time while maintaining the track geometry quality within specified limits.

Keywords
Tamping, track possession time, degradation, longitudinal level, optimization

Introduction
The improvement of the technical performance of track structures is essential to support the design capacity and at the same time improve the service quality of railway transport. Moreover, increasingly stringent safety requirements and demand for capacity for both freight and passenger traffic requires adequately supported intervention measures with optimum allocation and utilisation of track possession time. These intervention measures are categorised into track maintenance and track renewal tasks. Among the most important maintenance concerns for track structures are how to predict and control degradation of track structure and how to maintain the geometrical quality of the track. These factors influence ride quality and passenger comfort during operation, and also make a major contribution to the dynamics of the entire train/track system. Tamping is considered to be a maintenance task that has a large impact on the capacity of a railway network due to its particular requirements such as track possession time, quality demand, heavy machinery involved and scheduling challenges.

Effect of tamping operations on capacity
The time required to restore the geometry characteristics of a track is significant when the capacity of an existing network is considered. Depending on the maintenance philosophies and track management strategies, the track possession demand for tamping can vary for similar track sections. If an effective tamping strategy is not deployed, the track design capacity might not be achieved. Similar to other maintenance activities on railway infrastructures, the parameters that affect the total track possession time are the duration of the white period for each possession, travelling speed of the machine, working speed, preparation time and time for logistics considerations.

1Division of Operation & Maintenance Engineering, Luleå University of Technology, Sweden
2Luleå Railway Research Centre, Luleå, Sweden
3Beijing Jiaotong University, Beijing, People’s Republic of China

Corresponding author:
Stephen M. Famurewa, Division of Operation & Maintenance Engineering, Luleå University of Technology, Luleå 971 87, Sweden. Email: stefam@ltu.se

Date received: 27 January 2013; accepted: 3 June 2013
window and working speed are important, scheduling procedure is of particular interest in this study because redesign and innovation aspects are not within the scope of this work.

In the past, different principles have guided the planning, scheduling and implementation of maintenance actions: these include manufacturers’ recommendations, experience within the railway organisation, assumed deterioration, availability of maintenance equipment and other basic factors. However, these factors are not able to support the growing demand for capacity, safety, cost-effectiveness and other service quality requirements for railway transport. To this end, several techniques and methods have been developed for optimum planning and scheduling of railway infrastructure maintenance.

Track possession scheduling and optimisation

Developments in railway management have led to increasing need for optimum planning and scheduling of maintenance activities. The parameters of interest in several maintenance optimisation tools and techniques include maintenance costs, labour cost, life cycle cost, asset performance, track possession time, punctuality and other service quality parameters. Basically, maintenance optimisation of railway infrastructure gives short-, medium- or long-term plans for how preventive maintenance will be performed, on which segments and within which time horizon. To this end, an overview of railway infrastructure maintenance planning has highlighted two vital aspects of infrastructure maintenance planning: deterioration modelling and maintenance scheduling. On the aspect of track deterioration, important parameters to be taken into consideration for prognostics are initial quality, initial settlement and rate of deterioration. The significance of the initial quality of the track at the time of installation was investigated by Veit, where the life cycle management perspective of track structures was also studied. The rate of deterioration is governed by an integrated process of material degradation, traffic-induced degradation and maintenance. These phenomena are due to the design and layout of the track, rail profile, condition of the ballast, bearing capability of the subgrade, drainage problems, axle loads and traffic volume. On the scheduling aspect, maintenance activities are allocated to available time intervals, or optimum track possession windows are created for maintenance during the timetable schedule. Higgins proposed a model to determine the best allocation of maintenance activities and crews to minimise traffic disruption and completion time. Miwa developed a mathematical programming model for an optimal tamping schedule; it indicated the track division for which tamping must be implemented within a specified horizon. Cheung et al. developed a track possession assignment program to assign railway tracks to a given set of scheduled maintenance tasks according to defined constraints. The objective of the program is to create an assignment plan that maximises the assignment of job requests based on priorities while satisfying all imposed constraints. A preventive maintenance scheduling program was presented by Budai to cluster routine activities and projects for a link over a certain period such that the sum of possession costs and maintenance costs is minimised. They developed some heuristics such as ‘most frequent work first’ and ‘most costly work first’ to solve the formulated preventive maintenance scheduling problem. Andrade and Teixeira created a preventive maintenance scheduling program connected to track geometry quality using a bi-objective integer formulation that balanced renewal and maintenance costs with train delays. Vale et al. developed a binary linear program to schedule tamping taking into consideration track degradation over time, track layout, quality recovery of track and track quality limits based on standards.

Several mathematical programs for preventive maintenance scheduling problems have been formulated, and solutions have been proposed using multi-objective algorithms, artificial intelligence approaches, heuristic algorithms and other techniques. There is need to further address the optimisation allocation and utilisation of track possession time for maintenance to enhance operational capacity. The contribution of this paper is an investigation of differential deterioration along a length of a track and quantification of tamping intervention on a specific length of track over a finite horizon. Also, a methodology for optimum scheduling of tamping is proposed to minimise the direct cost of intervention and cost of track possession while maintaining geometrical quality within the desired level.

Theory and model formulation

The lifetime of track structure, as well as the quality of the track at any point in time, can be described in terms of deterioration and recovery phenomena. There are some basic principles and theories that are essential to the modelling of these phenomena.

Track degradation

The passage of a train over a track generates enormous forces. This leads to deformation and wear of track components such as rails, sleepers, fasteners, ballast and subgrade, and consequently, long-term deterioration of the track geometry. This phenomenon is one of the most important aspects of railway infrastructure maintenance. Thus, it is a vital requirement to adequately understand the pattern of deterioration of track geometry quality due to the
accumulation of plastic and elastic deformation as a result of traffic loading.

The geometry quality and irregularity of ballasted tracks are monitored by some key parameters including longitudinal level, alignment, gauge, cross level and twist. To manage track geometry problems, infrastructure managers (IMs) and academic researchers have monitored the evolution of principal parameters such as longitudinal level and alignment while others have used derived indices such as variation of acceleration (due to irregularity) and combinations of quality parameters to monitor the growth of track quality defects. Moreover, the standard deviation of the irregularity in the vertical direction of the track coordinate system has been shown to be sufficient to model the track geometry quality and also to support maintenance decisions and actions.

The life cycle behaviour of track has been explained using different empirical models based on measurement records and load or time. These models include the grey model, linear model, exponential model and other empirical models. The exponential model in equation (1) is preferred in this study considering the established behaviour of track; high-quality track or new track deteriorates slowly while low-quality track or ageing track deteriorates rapidly. The standard deviation of a vertical irregularity for a segment $s$ at time $t$ is given as

$$\sigma(s, t) = \sigma(s, 0)e^{bt}$$

where $\sigma(s, 0)$ is the initial standard deviation for segment $s$ estimated from vertical irregularity values from the recording car over a length of 200 m and $b(s)$ is the exponential constant or degradation rate for segment $s$ estimated from a series of measurements over time.

**Tamping and recovery**

The quality of the track geometry eventually deteriorates beyond the allowable threshold for maintenance and safety giving rise to the need for intervention to restore it to the design specifications. The intervention level depends on the tamping strategy deployed. Common strategies in use by IMs include correction of isolated defects and restoration of lines when specified thresholds are reached. The details of recommended intervention limits can be found in EN-13848-5. Ideally, from a life cycle perspective these thresholds should be dynamic, thereby becoming dependent on the age of the track structure or the number of interventions carried out. This practice will enhance the durability of track quality and also extend the lifespan of the track. Other factors considered in tamping are the availability of tamping machines and the maintenance philosophy of the owner of the asset. When a prognostic tamping strategy is to be deployed, the recovery or amount of improvement to be achieved by the tamping must be known in advance.

In reality, the recovery or efficiency of tamping depends on several factors such as track quality at tamping, age of track components, tamping technique, number of previous tamping operations, ballast condition and human factors. In the present study, an empirical regression model based on data collected in previous research on the investigated route has been developed. The model describes the relation between the standard deviation of the longitudinal level before tamping and the improvement following tamping after passage of some traffic for stabilisation of the track. The model is used to predict changes in the geometry parameter at any point in time when tamping is carried out. Figure 1 shows a plot of the observed recovery and quality at intervention. The regression model is given in equation (2), and it has an $R^2$ value close to 0.7. In order to improve the prediction accuracy of the simple regression model, a 90% prediction limit was estimated for the model to account for other parameters that could affect the recovery value.

$$\text{Recovery } R = 0.5445\sigma(s, t) - 0.8893$$

The linear model shown in Figure 1 suggests that recovery depends on the quality at the point of intervention. Only the observation data that fall within the region considered most likely for good substructure and ballast condition in the best-practice guide for optimum track geometry durability are shown in the figure.

**Assumptions**

The following assumptions underpin the model developed in this paper.

1. Deterioration follows an exponential model based on the explanation given earlier that high-quality track deteriorates rapidly and irreversibly (see equation (1)).
2. The degradation rate of each 200 m track section is considered to be constant within the time horizon considered for the scheduling in this study.
3. The track section is considered to have good ballast conditions because the track structure is relatively new. Thus, tamping recovery is assumed to lie within the region considered likely for efficient tamping in a good ballast condition and to follow the model described in the previous section. The same recovery model is used for all the segments.
4. Segments with switches and crossings and other critical units will be maintained using spot tamping considering them as isolated defects.
reason for this is because so many engineering works are carried out on these segments; thus to model their deterioration, additional measurement is required.

5. There are two tamping machines, one has limited availability, high tamping efficiency, and it is used for early or preventive intervention. The second one is more available but has a relatively low tamping efficiency and is suitable for late intervention or corrective tamping. Optimum allocation seeks a balanced mix of the two possibilities in terms of cost, quality and time on the track.

6. There are four different stations that can provide temporary parking for the machines before and after tamping.

Optimisation procedure

The activity breakdown structure shown in Figure 2 is a simplified intervention process for both preventive and corrective policies. The travelling time depends on the speed and location of the tamper before the shift, while the set-up and dismantling times are fixed.

A simplified representation of the optimization procedure is given by the flow chart shown in Figure 3. An algorithm was developed in FORTRAN to obtain a solution for the model formulated in equations (3) to (8). The notation is defined in Appendix 1.

The objective function for intervention decisions is

\[ t_e(N(t)) = \max\{t_1 + t_2 + t_3\} \]  

Figure 1. Recovery of track geometry quality after tamping interventions. (1*: preventive intervention threshold and 2*: corrective intervention threshold.).

Figure 2. Track possession time for interventions.
The decision function is
\[
g[t_w] = \begin{cases} 
1, & t_w(N(t)) = \max\{t_1 + t_2 + t_3\} \\
0, & \text{else}
\end{cases}
\]  
(4)

where
\[
t_1 = \min\{s(1) - p(0) + s(N) - p(j)\} + s(N) \\
-s(1) - (N - 1) \times \frac{d}{v}, \quad (i, j \in 1, \ldots, n_{\text{park}})
\]  
(5)
\[
t_2 = 2t_p(N - n_{\text{park}})
\]  
(6)

The number of corrective intervention shifts required to maintain the line section, given a specific number of preventive maintenance shifts is
\[
N_{\text{shift}}(T) = \sum_{t=1}^{T} g[t_w(N(t))]
\]  
(8)

### Case study

**Description of case study**

A line section in the network of the Swedish Transport Administration (Trafikverket) is considered in the case study. The line section is 130 km of single track from Kiruna to Riksgränsen. It is basically a freight line because the majority of the traffic is iron ore freight, although passenger trains and other freight trains also use the line. Thus, the
line has a high socio-economic importance and high maintenance requirements. The train speed on the line is between 80 and 120 km/h. The maximum allowable axle load on the line section is 30 t and the annual accumulated tonnage is over 22 MGT. The line has continuous welded rail, head hardened 60E1 rail type, with concrete sleepers and Pandrol fasteners. It is highlighted that the track structure on this line section was renewed between 2006 and 2009; this major work also included ballast renewal. It should also be noted that this line section operates in extreme climatic conditions which can influence the reliability, availability, maintainability and safety characteristics of the infrastructure. The winter season sees snowfall and extreme temperatures. The annual temperatures vary between $-40^\circ$C and $+25^\circ$C.

**Inspection data**

Inspection is an important element of any effective preventive maintenance programme. Track geometry inspection is needed for planning a tamping strategy that is optimum in the allocation and utilisation of track possession time. It also gives useful information to avoid tamping too early or too frequently, which degrades the ballast condition, and at the same time to warn against intervening too late, which can result in a temporary speed restriction or failure. Track inspection is done by the IM based on two factors, speed and annual accumulated tonnage on the section. For the case study, track quality inspection was done three to six times a year, generally between April and October. The available inspection data extend from 2007 to 2012, and for each 200 m segment, only data after the completion of renewal were considered. Several geometry parameters were recorded by the train measurement vehicle, but only the standard deviation of the longitudinal level over each 200 m track length was used for the geometry quality prognosis and maintenance optimisation. The standard deviations of the longitudinal level (3-25 m wavelength) for a 200 m track segment from four measurements on the 130 km line section investigated in this study are shown in Figure 4.

**Results and discussion**

The results and findings of the modelling and predictions of deterioration and recovery phenomena in this study are presented in this section.

**Non-homogeneity of track sections**

The degradation rate of the longitudinal level for each 200 m segment was estimated using the exponential model explained in the previous section and the 2007–2012 inspection data. The degradation rate is an indication of the evolution of the track geometry quality for each segment. The distribution of the degradation rates for all 592 segments is shown in Figure 5. It is heavily skewed to the right, indicating the existence of critical spots with rapid degradation in their geometric quality. More than 50% of the track section has an exponential degradation rate between 0.00024 and 0.00060. The differential degradation rate along the track sections reflects non-homogeneity and variation of track components along the track length. In fact the 200 m track segments can be regarded as non-identical units in terms of quality deterioration.

An inference that can be drawn from this plot is that continuous tamping of the whole length of the track section might not be the best strategy in terms of life cycle management of the track. An essential maintenance requirement revealed by the figure is the balance of preventive and corrective tamping, since a higher exponential rate will require more
interventions than a lower exponential rate, this is also noted by Arasteh Khouy et al.\textsuperscript{18}

Tamping strategy

Based on the observation of a differential degradation rate along the track length, there is a need to optimise the preventive and corrective tamping interventions on the track section. Using the model procedure described in the flow chart (Figure 3) and equation (8), the number of corrective interventions that will be required in 2 years for different numbers of allocated preventive tamping shifts was estimated, and the result is shown in Figure 6. The sensitivity of the result is also shown using the 90\% prediction limit of the recovery model fit. Increasing the number of allocated shifts for preventive tamping decreases the consequent number of corrective tamping shifts up to a point where there is no need for corrective tamping after the initial one which was necessary at the beginning when few sections were above both intervention thresholds.

Direct cost of intervention

Using the proposed optimisation procedure shown in Figure 3 with different cost ratios for the two tamping policies, the total cost for tamping interventions over a short period of 2 years is given in Figure 7. A high value of the ratio $c_p/c_c$ results in a higher cost of intervention when the number of preventive maintenance shifts increases.

For all cost ratios, the direct cost of intervention is constant after 16 shifts because no more segments will exceed the preventive maintenance threshold within the 2 year period of planning. For low corrective maintenance costs, the economic optimum plan in a short period will always be to carry out corrective maintenance. However, this is not always the best policy, particularly during the early life of the track, because it will reduce the service life. From Figure 8 the economic optimum policy considering a cost ratio of $c_p/c_c = 1$ is to have only a few preventive intervention shifts. However, track possession time and quality of the track are other parameters that need to be considered.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Distribution of exponential constant or degradation rate.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Corrective interventions and scheduled preventive interventions.}
\end{figure}
Cost of intervention and track possession

The present demand on railway infrastructure requires augmented allocation and utilisation of track possession time, and thus there is a need to implement optimum maintenance practice. In view of this, the global cost model proposed in Galar et al. was adapted to estimate the total cost of intervention by adding the direct and indirect costs of intervention. Following the model procedure outlined in Figure 3 and using stochastic simulation for the recovery model (equal chance of obtaining recovery
within the prediction limits, see Figure 1), robust estimation of track possession time was obtained and also the indirect cost of intervention (using $C_{DT} \approx 2c_e$). Figure 9 shows the total track possession time over a short period of 2 years for different numbers of preventive intervention shifts. Strategies with an $N_{p,shift}$ value of between five and eight are efficient because they are in the range of the values of the minimum track possession time. Furthermore, Figure 10 shows the total cost of intervention with different numbers of shifts allocated for preventive intervention. According to the results shown in Figure 10, selecting strategies with more than eight preventive maintenance shifts will result in additional cost due to overly frequent track possession. An optimum strategy should have a high economic performance, process efficiency and satisfy the required effectiveness in terms of track quality. In view of this, strategies with $N_{p,shift}$ values up to eight are cost-efficient because they are in the neighbourhood of the minimum total intervention costs. However, it is necessary to confirm the optimality of any of these strategies by assessing the resulting quality characteristics to see if they meet the quality requirement of the infrastructure manager. In this study $N_{p,shift} = 8$ is suggested because it is cost-efficient and produces better quality than other strategies with lower preventive maintenance shifts.

### Track quality characterisation

The tamping strategies were evaluated by characterising the predicted geometry quality using the procedure in the current state-of-the-art description for track geometry quality. Figure 11 presents the cumulative frequency distribution of the predicted longitudinal level defects over the length of the entire track section. The figure characterises the initial quality and also track quality after 2 years for three different scenarios: no tamping, $N_{p,shift} = 8$ and $N_{p,shift} = 16$. If it is required by the IM that at least 90% of the total segments on the track section should not exceed track quality class C ($\sigma_{LL} < 1.8\text{ mm}$) for safety, comfort, ride quality and life cycle management reasons, then from Figure 11 having $N_{p,shift} = 8$ is adequate.

Table 1 gives the detailed description and extended classification of the track section into track quality classes for the 2 years under consideration using the procedure mentioned previously. If the requirement of the IM puts a limit on the proportion of the track segments expected to be in each quality class within a certain time horizon, then this can be checked.

### Conclusions

The demand on the allocation and utilisation of track possession time is increasing, and there is a need to develop a model to support maintenance decisions, particularly for maintenance tasks with high possession requirements. This paper has presented an
optimisation tool to support the allocation of track possession time in a short-term plan for tamping. It considered the exponential function to model the deterioration of each 200 m segment in the case study.

The study found a varying degradation rate of the longitudinal level over the studied 130 km track section with about half of the segments having an exponential degradation rate of between 0.00024 and 0.0006. The objective of the optimisation model is to minimise the direct costs of intervention and track possession while geometry quality is maintained at a desirable level. In the case study, the optimum tamping strategy for a 2 year planning horizon will be to allocate eight shifts for preventive tamping while additional quality failures will be restored using a corrective intervention policy. This will support adequate planning and resource allocation including ordering of tamping machines. This approach provides knowledge of track behaviour; quantification of tamping requirements; and a suggested tamping strategy that reduces track possession time and associated costs.

Finally, in future work a long-term plan will be developed that considers dynamic intervention threshold levels for the two policies: low threshold levels at the early stage of the track life and high thresholds at the later part.

Funding

The authors wish to thank Trafikverket and Luleå Railway Research Centre for financial support.

Acknowledgements

The technical support provided by Professor Diego Galar, Dr Arne Nissen and Jens Jønsson was greatly appreciated by the authors.

References


Table 1. Percentage of the track falling into each quality class (A is the best, and E is the worst).}

<table>
<thead>
<tr>
<th>Scenario</th>
<th>A ≤ 0.75 mm</th>
<th>B 0.75–1.1 mm</th>
<th>C 1.1–1.8 mm</th>
<th>D 1.8–2.5 mm</th>
<th>E &gt; 2.5 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>24.12</td>
<td>41.48</td>
<td>30.02</td>
<td>4.21</td>
<td>0.17</td>
</tr>
<tr>
<td>Np(shift) = 8</td>
<td>5.40</td>
<td>26.98</td>
<td>58.68</td>
<td>8.94</td>
<td>0.00</td>
</tr>
<tr>
<td>Np(shift) = 16</td>
<td>5.40</td>
<td>37.27</td>
<td>57.16</td>
<td>0.17</td>
<td>0.00</td>
</tr>
<tr>
<td>No tamping</td>
<td>5.39</td>
<td>23.94</td>
<td>49.07</td>
<td>14.68</td>
<td>6.92</td>
</tr>
</tbody>
</table>
Appendix I

Notation

- $c_{p,c}$: cost of preventive and corrective intervention per segment
- $C_{DT}$: cost of downtime per hour (depends on line class)
- $C_{I-l}$: direct cost of intervention in $T$ days
- $d$: length of segment = 200 m
- $n_a$: number of adjacent segments among tamped segments
- $n_{park}$: total number times tamping machine is parked (in this study $n_{park} = 4$)
- $N$: number of segments tamped in preventive intervention shift ($N_p$) or corrective shift ($N_c$)
- $N_{c,shift}$: number of corrective intervention shifts required for a specified number of preventive intervention shifts $N_{p,shift}$
- $N(i)$: number of segments above preventive threshold ($N_{p,i}$) or corrective threshold ($N_{c,i}$) on day $i$

- $p$: location of start parking lot $p(i)$ and end parking lot $p(j)$ ($i, j \in 1, \ldots, n_{park}$)
- $s(i)$: location of tamped segments $S(i) \in S$ and $i = 1, \ldots, N$
- $t_{p}, t_{i}$: time for each set-up and disassemble = 10 min, total travelling time during a shift
- $t_{i}, t_{n}$: total set-up and disassemble time for a shift, duration of shift. (minimum duration $t_c = 6$ h)
- $TC$: total cost of intervention in $T$ days
- $TT$: total track possession time
- $\nu$: tamping speed for preventive tamper $\nu_p = 1.4$ km/h and corrective tamper $\nu_c = 0.8$ km/h
- $\nu'$: travelling speed for preventive tamper $\nu'_p = 90$ km/h and corrective tamper $\nu'_c = 80$ km/h
- $\sigma^*$: threshold for preventive interventions $\sigma^*_p = 1.8$ mm and corrective interventions $\sigma^*_c = 2.1$ mm from EN-13848-5.
Augmented utilisation of possession time: analysis for track geometry maintenance

Institution of Mechanical Engineers. Proceedings. Part F: Journal of Rail and Rapid Transit (Accepted for Publication).
Augmented utilisation of possession time: analysis for track geometry maintenance

Stephen M. Famurewa1,2, Ulla Juntti2,3, Arne Nissen4, Uday Kumar 1,2
1Division of Operation & Maintenance Engineering, Luleå University of Technology
2Luleå Railway Research Centre Luleå, Sweden
3Performance in Cold AB Luleå, Sweden
4Trafikverket, Sweden
Corresponding Author: stefam@ltu.se Phone number:

Abstract
The demand for more capacity on existing railway network is a challenge for many Europe-based infrastructure managers, and addressing this challenge would entail augmented utilisation of track possession time. It is considered that large scale maintenance tasks such as geometry maintenance can be improved to support the goal of reducing time on track and accommodate more freight traffic. In this study, analysis of track geometry maintenance was performed for reducing the possession time requirement. The procedure and models for planning and optimising track geometry maintenance are presented. A statistical model with simulation approach was used for prognosis of the geometry condition, while a schedule optimisation problem is formulated to support intervention decisions and optimise track possession time. The results of the case study show that optimising the maintenance shift length and cycle length are opportunities for reducing the possession requirement of track geometry maintenance. In addition, continuous improvement of tamping process through lean analysis promises about 45% reduction in possession requirement for a tamping cycle.

Keywords: railway track, scheduling optimization, tamping, track geometry degradation, maintenance, track possession time

1. Introduction
The demand for service quality improvement of railway operations and increased traffic on existing networks are among the major concerns of infrastructure managers. Effective track possession management is essential for supporting the design capacity of existing networks. It also plays a key role in meeting the additional freight and passenger traffic capacity demands without compromising safety and quality requirements. Data driven maintenance planning and scheduling is a requirement in effective track possession management. This facilitates optimum allocation and utilisation of track possession time for track maintenance and track renewal. In practice, some maintenance activities are not considered demanding in terms of
safety and possession requirements, if the event involved does not prevent the achievement of specified performance or cause a consequence (delay or cost) greater than the minimum allowable. Thus, it is adequate to adopt opportunity based maintenance approach to merge them with other activities. However, large scale activities such as grinding, turnout maintenance, tamping and other geometry maintenance measures have peculiar requirements in terms of cost, track possession time, quality demand, heavy machinery involved and scheduling challenges [1]. These requirements contribute to their large impact on railway network capacity. For example, in Sweden an average of over 10% of the total network is tamped annually, and this requires substantial track possession time. Figure 1 shows the length of track tamped in Sweden annually as recorded in the maintenance system.

Track geometry is a measure of its integrity and quality, and it has been adequately proven to be a necessity and not luxurious requirement in the design, construction, installation and maintenance of tracks [2]. Well maintained track geometry not only ensures safe train operation, passenger comfort and good vehicle ride quality but also extends the life of track structure and improve its availability for train operation. This makes geometry analysis and maintenance essential from the viewpoints of cost reduction and availability improvement [3] [4].

![Figure 1: Length of track tamped in Sweden annually](image)

### 1.1. Maintenance track possession time

Track maintenance requires that scheduled train paths be cancelled, redirected or postponed as is best suited to the case at hand. Maintenance tasks with high frequency or long maintenance window requirements have the greatest effect on track availability and network
capacity. Maintenance tasks related to geometry and rail profile correction are less frequent but require longer track possession windows. These tasks include, tamping, grinding, stone blowing, ballast cleaning, ballast completion and spot repairs. These tasks are difficult to squeeze into short possessions within the timetable and are therefore considered to have a great impact on track availability. These high impact maintenance tasks provide a good scope for improvement. In the shift towards augmented possession time utilisation, a realistic solution involves maximising availability of track by optimising possession time. This includes lean optimisation, maintenance window optimisation, subtask optimisation, maintenance interval optimisation and better planning.

1.2. Track possession for geometry maintenance

The time required to restore track geometry characteristics is significant when considering the capacity of an existing network. Depending on the maintenance philosophy and track management strategy, track possession demand for tamping could vary for similar track sections. If an effective tamping strategy is not deployed, the track design capacity might not be achieved. The parameters affecting the total track possession time for tamping are maintenance window, travelling speed of the machine, working speed, preparation time, time for safety arrangement and other logistic considerations [5].

In the past, different principles have guided the planning and implementation of geometry repairs, some of which are based on manufacturers’ recommendations, experience within the railway organisation, assumed deterioration, availability of maintenance equipment and other factors. However, these principles are unable to meet the growing demand for reduced time, improved safety, cost effectiveness and other quality factors related to railway transport. To meet this demand, several techniques and methods have been developed for optimum planning and scheduling of railway infrastructure maintenance. An overview of railway infrastructure maintenance planning highlighted two vital aspects of infrastructure maintenance planning: deterioration modelling and maintenance scheduling [6]. Significant parameters to be considered for track geometry deterioration or prognostics are initial quality, initial settlement and rate of deterioration [7], [5], [2], [8], [9]. The deterioration phenomena can be ascribed to track design and layout, rail profile, ballast condition, bearing capacity of the subgrade, drainage problems, axle load, traffic volume, etc. [2], [5], [10]. Track degradation phenomenon has been considered to be stochastic process and modelled using Petri net [11], Monte Carlo technique [12], Markov model [13], [14] and hierarchical Bayesian models [15]. In a recent study, artificial neural network was used as an alternative method to model significant relationship among relevant variables and to predict railway track geometry degradation [16]. Further, simple statistical regression models have been used to explain and predict the course of track geometry degradation. Most widely used among the statistical models are linear regression models [3], [5], [7], [9] and exponential models [2], [12], [17], [18]. Exponential
smoothing method is another relevant statistical approach which has been used to model track degradation [8].

Considering the scheduling aspect, Higgins [19] presented a model to determine the best allocation of maintenance activities and crews to minimise traffic disruption and completion time. Miwa [8] developed a mathematical programming model for generating an optimal tamping schedule, and indicating the track division for which tamping must be implemented within a specified period. Cheung et al. [20] developed a track possession assignment program for assigning railway tracks to a given set of scheduled maintenance tasks according to defined constraints. A preventive maintenance scheduling program was presented by Budai et al. [21] to merge routine tasks and projects for a link over a certain period such that the sum of possession costs and maintenance costs is minimised. Andrade et al. [9] addressed a preventive maintenance scheduling program connected to track geometry quality by using a bi-objective integer formulation that balances renewal and maintenance costs with train delays. Vale et al. [7] developed a binary linear program to schedule tamping considering track degradation over time, track recovery and quality limits based on standards. However optimisation of maintenance window duration for line tamping, maintenance interval and subtasks have not been addressed adequately for maximising track availability or minimising tamping possession time.

The contribution of this article is model development for augmented utilisation of possession time for geometry maintenance. Simulation approach has been used to estimate the optimum maintenance window, maintenance cycle length, and evaluate improvement potentials of tamping subtasks. In addition, a schedule optimisation problem was formulated to support intervention decisions, merge spot failure remedial actions and reduce the total time on track for geometry maintenance.

2. Maintenance task breakdown structure

The execution of maintenance tasks involves series of subtasks which are carried out serially or concurrently. During tamping operation, the subtasks carried out could contribute directly to the primary objective of restoring track geometry quality, support other value-adding subtasks or be mere waste or loss due to inefficient process. Thus, subtasks during possession for tamping are classified as value-added tasks, necessary non-value-added tasks and non-value-added tasks. Basically, the utilisation of possession time for large-scale maintenance tasks including tamping can be broken down into the seven generic sub-tasks which are described below:

- Confirmation: Ascertain if possession has been granted either at the beginning of the shift or during the shift when moving from one section to another. This adds no direct
value to the geometry restoration and can be eliminated with effective planning. Thus, it is a non-value added task

- Waiting: Waiting for personnel, equipment, traffic, or other logistic purposes. This subtask adds no direct value but could be necessary if the waiting is traffic-related.
- Communication: Conversation on phone to get information relevant for maintenance commencement and documentation. With an effective maintenance process and system, this can be eliminated. Thus, the task is considered as a non-value-added task.
- Preparation: The setting up and dismantling of heavy-duty equipment takes considerable time. In addition, it includes track safety and clearance measures. This subtask is necessary non-value-adding task.
- Active tamping time – This is the value-adding task involving rail lifting, insertion of tamping tines, squeezing and other actions for the restoration of track geometry quality.
- Transportation – This involves moving maintenance and other support equipment to the task location. It is an important necessary non-value-adding subtask.
- Pre and post measurements: These are carried out to check how far the track has settled from its reference position before tamping and to ensure relative or absolute restoration. However it is possible to use data from track measurement runs directly if the issues of location accuracy and relative-absolute coordinate system are resolved.

The result of process observation from the lean perspective for tamping carried out in an EU project called AUTOMAIN is shown in Figure 2. The figure shows a high level break down of track possession time typical for a tamping shift and highlights the composition of each subtask.

![Figure 2: Typical high level break down of track possession time for tamping in Europe [22]](image-url)
There is scope for improvement by applying effective maintenance principle such as lean optimisation, maintenance window optimisation, subtask optimisation, maintenance interval optimisation and use of improved planning tools.

3. Theories and model development

The approach employed in this article for augmentation of possession time utilisation involves data analysis, simulation and schedule optimisation. The tamping policies considered and the optimization procedures employed are described below.

3.1. Tamping policies

In the analysis and model development for effective geometry maintenance, the following tamping policies and approaches were considered:

1. Corrective policy: The geometry condition of a segment is restored when obvious defects are identified and reported. This could be after the intervention limit or the immediate action limit (late intervention, especially for isolated defects).

2. Predetermined policy: The geometry condition of a line is restored at predetermined intervals based on usage, i.e. tonnage or time. This is done to support geometrically even track and to avoid new local defects that could be created by frequent spot failure interventions.

3. Prognostic policy: Series of inspection data are used for modelling the chronological evolution of the track geometry quality of a short segment. Doing so creates the opportunity to intervene at any preferred threshold, identify spot failures and merge interventions in a shift without compromising safety.

An optimum strategy seeks the ideal combination of the policies such that both over-maintenance in the form of early intervention and under-maintenance in the form of late intervention are avoided.

3.2. Optimization procedure

The evaluation and optimization procedure to augment the utilisation of possession time for geometry maintenance is simplified in the flow chart shown in Figure 3 and explained thereafter.
1. Track simplification and characterisation: This involves mapping the track section with respect to its length, components, maintenance stabling points and estimating the degradation rate of each 200m-long segment using regression analysis.

2. Track degradation: The degradation of each segment is predicted by inserting the initial standard deviation of vertical irregularity $\sigma(s, 0)$ and degradation rate $\dot{b}(s)$ estimated from historical measurements into the formula given in equation 1. Where $T$ is time in days, $e$ is the prediction error, a random variable modelled using truncated normal distribution with mean $\mu_e=0$, standard deviation $\sigma_e$ and the two boundary values $a_e, b_e$ estimated from data.

$$\sigma(s, T) = \sigma(s, 0)e^{\dot{b}(s)T} + e$$  \hspace{1cm} (1)
3. Geometry restoration: Due to the amount of data available for modelling tamping operation efficiency, two approaches have been used. For the first restoration of each segment, first part of equation 2 is adopted and an assumption of imperfection in quality restoration is assumed for subsequent restoration attempts. This means that the quality after tamping $\sigma(s, T)^{\text{cum}}$ depends on the quality before tamping $\sigma(s, T)$, cumulative number of tamping operations (cum), quality achieved in last restoration $\sigma(s, T)^{\text{cum-1}}$ and the quality loss QL factor (taken to be 10% in the case study and based on expert opinion).

$$
\sigma(s, T)^{\text{cum}} = \begin{cases} 
0.4002\sigma(s, T) - 0.0307, & \text{cum} = 1 \\
\sigma(s, T)^{\text{cum-1}} + QL\left[\sigma(s, T) - \sigma(s, T)^{\text{cum-1}}\right], & \text{cum} > 1
\end{cases}
$$

(2)

Figure 4 shows a simulation of the longitudinal level of a segment on the track section which was performed using equations 1 and 2 and considering intervention only at a predefined threshold.

Figure 4: Simulation of longitudinal level evolution with intervention at 2.6 mm

4. Predetermined intervention decision: The decision regarding when to tamp is based on the cycle length. The exact day in the year for operation commencement can be selected freely based on other maintenance constraints and boundary conditions. Equation 3 and the associated formulations given in equations 4-6 are used for estimating the possession time required for a geometry maintenance cycle. The
estimated total possession time for predetermined maintenance over the planning horizon is presented in equation 7.

Estimated possession time required to maintain the entire track section during a maintenance cycle.

\[ TPC = \sum_{t=1}^{365.C} t_w(T) \]  

where

\[ t_w(T) = t_c(T) + t_p(T) + t_{proc}(T) \]

\[ t_w(T) \leq t_w^* \]

Tamping time \[ t_c(T) = \frac{L_s(T) - L_i(T)}{v} + \sum_{i=1}^{N} g[i] \cdot t_w(i)(T) \]  

where \[ g[i] = \begin{cases} 1, & L_s(T) < L_{w(i)} < L_r(T) \\ 0, & \text{else} \end{cases} \]

\[ L_s(T) \leq L_r(T) \]

Travelling time \[ t_p(T) = \frac{\min\left(\text{abs}\left(L_s(T) - L_p(i)\right)\right) + \min\left(\text{abs}\left(L_s(T) - L_p(j)\right)\right)}{v^*} \]  

Process time \[ t_{proc}(T) = t_{cw} + t_{cf} + \left(t_{wt} + t_{pre}\right) \cdot np(T) \]  

Estimated time required for predetermined maintenance over the planning horizon

\[ TPC_{PH} = \left\lfloor \frac{PH - 1}{C} \right\rfloor TPC \]  

5. Corrective intervention decision: This is made using the prediction model and intervention threshold taken from standard. For improved utilization of the possession time, locations with a quality level close to the threshold level or surpassing it are merged in the same window such that the overall corrective maintenance time over
the planning horizon is minimized. The objective function defined in equation 8 and the associated constraints are used for the optimisation.

Estimated time required for corrective maintenance over the planning horizon.

\[ TPS_{py} = \min \left( \sum_{T=1}^{365PH} t_v(T) \right) \]  
where \( t_v(T) = t_e(T) + t_w(T) + t_{proc}(T) \)  
\[ \text{s.t.} \]  
\( t_v(T) \leq t_v^* \) Window duration is not exceeded during a shift  
\( N(T) \leq N'(T) \) Only segments above the threshold are maintained  
\( \sigma(s,T) < \sigma^* + 0.3 \) No segment overly exceeds the threshold

\[ N'(T) = \sum_{s=1}^{S} f[\sigma(s,T) - \sigma^*] \]  
\[ f[x] = \begin{cases} 1, & x \geq 0 \\ 0, & \text{else} \end{cases} \]

Tamping time \( t_v(T) = \frac{N(T) \cdot d}{v} + \sum_{i=1}^{t_{wv}} f[ii] \cdot t_{wv}(ii) \)  
\[ f[ii] = \begin{cases} 1, & s(ii) = s(i), i = 1 \ldots N \\ 0, & \text{else} \end{cases} \]

Travelling time \( t_v(T) = \frac{\min|s(1) \cdot d - L_{v(i)}| + |s(N) \cdot d - L_{v(0)}| + (s(N) - s(1) - (N - 1)) \cdot d}{v'} \)  
\[ \text{Process time} \ t_{proc}(T) = t_{em} + t_{ef} + t_{at} + t_{proc} \cdot (N - n_u) \]  

6. The total track possession time required over the planning horizon is estimated using equation 13. To reduce the possession time, shift length, cycle length and other
Process parameters are varied, and the optimum conditions are obtained using the graphical approach.

\[ TPT = TPC_{PH} + TPS_{PH} \]  \hspace{1cm} (13)

Table 1. Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPT</td>
<td>Total track possession time required over the planning horizon</td>
</tr>
<tr>
<td>TPC&lt;sub&gt;PH&lt;/sub&gt;, TPS&lt;sub&gt;PH&lt;/sub&gt;</td>
<td>Total possession time for cycle maintenance, and total possession time for spot maintenance</td>
</tr>
<tr>
<td>N'(T)</td>
<td>Number of segments above or close to intervention threshold ( \sigma ) (2.6 mm) on day ( T )</td>
</tr>
<tr>
<td>N(T)</td>
<td>Number of segments tamped during corrective intervention</td>
</tr>
<tr>
<td>( t_{pre}, t_{c}, t_{v} )</td>
<td>Preparation time (setup, dismantling and safety measures), total travelling time during a shift, and total tamping time during a shift</td>
</tr>
<tr>
<td>( t_w )</td>
<td>Possession time for a cycle or corrective tamping shift (Max. duration ( t_w = 6 ) h)</td>
</tr>
<tr>
<td>( d, n_a )</td>
<td>Length of each segment = 200 m, and number of adjacent segments among tamped segments</td>
</tr>
<tr>
<td>L&lt;sub&gt;T&lt;/sub&gt;, L&lt;sub&gt;sc&lt;/sub&gt;</td>
<td>Total length of the track section, and location of turnouts on the track section</td>
</tr>
<tr>
<td>L&lt;sub&gt;c(T)&lt;/sub&gt;, L&lt;sub&gt;e(T)&lt;/sub&gt;</td>
<td>Locations of start point ( L_c ) and endpoint ( L_e ) on the track section during a tamping shift on day ( T )</td>
</tr>
<tr>
<td>L&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Locations on the track section designated as temporary machine pick up point ( L_p(i,j) ) and park at the end of a shift ( L_p(i,j) ) (( i, j \in [1, \ldots, 4] ))</td>
</tr>
<tr>
<td>( v, v' )</td>
<td>Working speed (0.8 km/h) and travelling speed (80 km/h)</td>
</tr>
<tr>
<td>( v_c )</td>
<td>Working speed during corrective interventions (0.5 km/h)</td>
</tr>
<tr>
<td>s(i)</td>
<td>Index of segments with quality above threshold, ( i = 1, \ldots, N )</td>
</tr>
<tr>
<td>s(ii)</td>
<td>Segment index of turnouts</td>
</tr>
<tr>
<td>np</td>
<td>Number of train passages during a shift</td>
</tr>
<tr>
<td>PH, ( Y )</td>
<td>Planning horizon (PH=14 years in the case study), ( Y = 1, \ldots, PH )</td>
</tr>
<tr>
<td>( t_{w(ii)}, N_{sc} )</td>
<td>Additional working time required for turnouts with index ( ii ), and number of turnouts on the section</td>
</tr>
<tr>
<td>C, ( t_{proc} )</td>
<td>Maintenance cycle, and process time</td>
</tr>
</tbody>
</table>

4. Case study
One of the track sections of the Swedish Transport Administration network (Trafikverket) is used as a case study for analysing and optimising track possession time for track geometry maintenance. The theories and the optimisation procedure explained in the previous section
are implemented using collected inspection data and adapted process data to demonstrate various aspect of augmented possession time utilisation. A 130-km-long single track section running from Kiruna to Riksgränsen is considered here. The traffic on this section is basically iron ore freight, with some passenger traffic and other freight. The train speed on the line is in 80–120 km/h category. The maximum allowable axle load on the line section is 30 tonnes and the annual accumulated tonnage is over 22 million gross tonnage (MGT). The line has continuous welded rail of the head hardened 60E1 rail type, with concrete sleepers and Pandrol e-clip fasteners.

4.1. Description of inspection data

The collection of track geometry data using a measurement train on the selected track section is carried out 3-6 times annually. The measurement train recorded several geometry parameters but only the standard deviation of the longitudinal level (1–25 m wavelength) over each 200-m track segment was used for the geometry quality prognosis and maintenance optimization. The inspection data used in the case study spans the years 2007 to 2012. The measurements are used to estimate the deterioration constant of each 200-m segment and also the error distribution parameters, which are then used for the geometry prognosis of each track segment according to equation (1).

4.2. Process data Description

Track geometry maintenance actions could be described using its high-level breakdown structure, with related time requirement and relevant machine parameters along with their values. The potential for improving track geometry maintenance in terms of possession requirement lies in the possibility of optimising or modifying any of these parameters. The simulation input data considers practical and realistic descriptions of the maintenance breakdown structure and machine parameters, which are listed in Table 2.

<table>
<thead>
<tr>
<th>Machine/process parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travelling speed v'</td>
<td>80 km/h</td>
</tr>
<tr>
<td>Tamping speed v</td>
<td>0.8 km/h</td>
</tr>
<tr>
<td>Switch tamping tsw</td>
<td>30–70min*</td>
</tr>
<tr>
<td>Preparation time tprev</td>
<td>10 min</td>
</tr>
<tr>
<td>Confirmation time tcf</td>
<td>30 min</td>
</tr>
<tr>
<td>Communication time tecm</td>
<td>20 min</td>
</tr>
<tr>
<td>Waiting time twt</td>
<td>10 min</td>
</tr>
</tbody>
</table>

*The tamping time of a switch depends on the type or model of the switch; for example, it takes 30–40 min to tamp a switch of 1:9 type (lift = 0-30mm) while it takes 55–65 minutes to tamp a switch of the 1:20 type using the same lift.
For analysing the optimum maintenance window duration during a tamping shift, the number of train passages within a specified window was taken into consideration as well. Variation in the number of train passages within a shift was modelled using the optimistic, most likely and pessimistic cases. The three different possible cases considered in the simulation are presented in Table 3.

<table>
<thead>
<tr>
<th>Maintenance window (h)</th>
<th>Parameters (o, m, p)</th>
<th>Maintenance window (h)</th>
<th>Parameters (o, m, p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>(0, 0, 0)</td>
<td>6</td>
<td>(1, 2, 4)</td>
</tr>
<tr>
<td>4</td>
<td>(0, 0, 1)</td>
<td>6.5</td>
<td>(2, 3, 6)</td>
</tr>
<tr>
<td>4.5</td>
<td>(0, 1, 1)</td>
<td>7</td>
<td>(3, 4, 7)</td>
</tr>
<tr>
<td>5</td>
<td>(1, 1, 2)</td>
<td>7.5</td>
<td>(4, 6, 8)</td>
</tr>
<tr>
<td>5.5</td>
<td>(1, 2, 3)</td>
<td>8</td>
<td>(5, 7, 9)</td>
</tr>
</tbody>
</table>

5. Results and Discussion
The information contained in the measurement data is summarised using relevant exploratory and visualisation tools. For exploring the geometric condition of each segment on the track section, the exponential deterioration constants estimated from the set of measurement data are grouped into three clusters using the k-means clustering technique. The first cluster represents the set of segments with a high deterioration constant owing to the presence of critical spots. The second cluster comprises the average deteriorating segments, and third cluster consists of the segments with low deterioration. Figure 5 shows a cluster plot of the three clusters along with the segment locations, standard deviations of the longitudinal level, and estimated exponential deterioration constants.
The analysis and optimisation results of augmented utilisation of track possession for track geometry maintenance are divided into two groups for simplicity and logical discussion. The results are presented from planning and process improvements viewpoints.

5.1. Improvement of geometry maintenance planning

Effective maintenance planning is cardinal to availability maximisation and track possession time reduction. This includes maintenance task analysis which provides useful specific information and requirements that ensure efficient track possession for geometry maintenance. The aspects of maintenance considered for improvement are maintenance window duration for tamping shift and tamping cycle.

5.1.1. Optimisation of maintenance window

The time required for predetermined tamping during a cycle with different shift lengths is estimated using equation 7 with the specified process and machine parameters. The result presented in Figure 6 indicates that the possession requirement decreases exponentially with an increase in the maintenance window duration when the section is completely closed to traffic. A short maintenance window requires several shifts and overly long track possession time to complete a tamping cycle on a track section. This suggests that maintenance shifts must be long enough to reduce the impact of non-value added tasks such as travelling time and other process subtasks. However, the capacity utilisation of an important line would not permit an uninterrupted white period for such long durations. The number of train passages to be allowed within the white period is shown in Figure 6 and listed in Table 3. This traffic requirement is often a major cause of extended time on the track. Figure 7 presents the mean
possession time for each window size considering the triangular distribution of train passages given in Table 3. For augmented track possession time utilisation, it is obvious from Figure 7 that the optimum possession in a shift should be between 5.5 and 6 h. These white periods are efficient because of reduced impact of non-value added tasks.

Figure 6: Maintenance window and track possession time for a tamping cycle and train passages that would be cancelled

Figure 7: Optimum maintenance window duration considering traffic requirement
5.1.2. Tamping cycle optimization

For reducing time on track for geometry maintenance, there is need to optimize the predetermined cycle length and sufficient lead time should be provided for scheduling and merging spot failure corrections. To this end, deterioration model, recovery models and maintenance scheduling models described earlier are used for determining optimum geometry maintenance strategy. An optimum strategy suggests appropriate predetermined tamping cycle and a bundling approach for the correction of isolated geometry defects in weak zones and critical spots. This favours the augmented utilisation of possession time in addition to sustaining track quality. The optimisation procedure described in the flow chart shown in Figure 3 is used to model the degradation process, predetermined and corrective restoration process.

The geometry quality evolution of each segment is estimated using exponential function given in equation 1, where the error term is estimated using Monte Carlo simulation technique with truncated normal distribution. Even though the simulation is computationally intensive, 10,000 simulation cycles were carried out. The result presented in Figure 8 shows the time required to perform both line tamping and spot failure restoration for different maintenance cycle lengths over a planning horizon of 14 years. This result follows the conventional maintenance optimisation pattern i.e. shorter maintenance cycle requires large preventive maintenance inputs such as cost and time but has a small corrective maintenance consequence.

In contrast, longer maintenance cycle requires lower preventive maintenance inputs but significant corrective maintenance demand to support a certain performance level. Obviously, for augmented track possession time utilisation, extreme maintenance cycles should be avoided from the quality, cost and track possession viewpoints. Over-maintenance in the form of very short maintenance cycle should be avoided to decrease the risk of ballast destruction and track life span reduction [17]. Similarly, under maintenance with too long maintenance cycle should be avoided to reduce the risk of irreversible loss of quality, uneven track quality and creation of additional local defects owing to excessive spot failure correction.
The total track possession time and the probabilistic characteristics of each maintenance cycle length scenario are estimated using equation 13. Figure 9 shows the estimated total possession requirement for geometry maintenance of the track section with 95% confidence interval such that the specified maintenance threshold is not exceeded over the planning horizon. The required possession duration is very high with too short maintenance cycle, but as the interval increases, the possession reduces to a point where it starts to increase again until it is more or less constant. Placing the maintenance cycle length between years 3 and 6 appears to be somewhat optimum from possession duration viewpoint. This is because many track segments were estimated to reach the specified threshold between years 4 and 6, and are thus restored under the scheduled maintenance. However, a cycle length of 3 years seems to be relatively small and could present a risk of reducing the lifespan of ballast and other track components.

Furthermore, the maintenance possession time seem to be constant for cycle lengths greater than 6 years. This is because many segments have already been maintained before the scheduled line maintenance and there is little opportunity for coinciding spot failure correction during the line maintenance. This approach presents the decision maker with the opportunity to choose the optimum strategy based on the infrastructure owner’s objectives. Choosing a cycle length of 4 years is optimum from the possession viewpoint, but extending the cycle length to 6 years can be an optimum solution from both the possession and the ballast life.
This is consistent with the submission of the best practice guide for optimum track geometry durability that repeated tamping actions themselves can cause additional ballast damage, and will therefore reduce its service life [17].

5.2. Improvement of geometry maintenance process
Following the high-level breakdown of tamping mentioned in sections 2 and 4, the benefit achievable with process improvement is analysed herein. The improvements include machine improvement, operators’ skill improvement, traffic management improvement, and resource scheduling and planning improvement. The action plans suggested for improvement of each subtask are given in Table 4. Figure 10 shows the possible potential reduction in track possession requirement for improving each subtask. The figure shows that improving tamping speed gives the largest potential for reducing track possession time. However, the improvement of tamping speed is limited by design if durable track quality is to be ensured. This is because increase in speed performance of the tamping machine might compromise quality performance since it is designed for a range of output performances under certain condition. This therefore makes small improvement in other subtasks important and appreciable. For the tamper under consideration, increase of the tamping speed more than a certain limit (taken to be 20% for the case study) will definitely require design modification or tamper replacement. In addition, reducing the number of train interruptions during a
maintenance shift has good potential for reducing the track possession time because the set up and waiting times are reduced considerably.

Table 4: Improvement suggestions for geometry maintenance subtasks

<table>
<thead>
<tr>
<th>Machine/process parameter</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travelling speed</td>
<td>Better scheduling tool, permission to travel at line speed</td>
</tr>
<tr>
<td>Tamping speed</td>
<td>Improvement in operator skill, automatic setting of tamping parameters, and selection of suitable tamper.</td>
</tr>
<tr>
<td>Preparation time</td>
<td>Reducing traffic interruption which increase the frequency of set up and dismantling, and standard operations procedure.</td>
</tr>
<tr>
<td>Confirmation time</td>
<td>Automatic track occupation control system, standard operations procedure</td>
</tr>
<tr>
<td>Communication time</td>
<td>Improved management system and communication process</td>
</tr>
<tr>
<td>Waiting time</td>
<td>Standard planning procedure, reduction of traffic interruptions, which leads to additional waiting time. Lean traffic management</td>
</tr>
</tbody>
</table>

Figure 10: Potentials for reduction in track possession with subtask improvement.
Furthermore, the tasks are grouped into value-added (active tamping), necessary non-value added (preparation and travelling) and non-value added tasks (waiting, confirmation, communication). Figure 11 shows that the improvement of value-added tasks with the existing tamping machine, can lead to 15% reduction in track possession time over a tamping cycle. Elimination of the non-value added tasks as possession requiring tasks will reduce possession time by 28%. The necessary non-value added tasks can be improved to reduce possession time by 13% without any machine change. A total reduction of up to 45% in possession time is achievable by using the existing tamping machine. In addition, if the tamping machine is replaced with a high-speed dynamic tamper, the potential reduction owing to modified tamping speed alone is 36%.

![Subtask optimisation](image)

**Figure 11: Augmentation of track possession time per maintenance cycle**

6. Conclusions

Effective track possession management and availability maximisation require data driven maintenance planning and scheduling. The analysis and modelling procedure for planning and optimising track geometry maintenance presented in this study will facilitate reduction of track possession time if implemented. The analysis demonstrated that optimisation of the tamping cycle length and shift duration, as well as tamping process improvement present opportunities for augmented of possession time utilisation. The degradation model and the formulated schedule optimisation problem were used to support intervention decisions, merge spot
failure remedial actions and reduce the total time on track for geometry maintenance over a specified planning horizon. The concluding remarks of the case study are as follows:

- In situation where the maintenance window is not limited by train traffic or machine parameters, the normal 8 h maintenance shift had the minimum possession for a tamping cycle.

- If train passages are to be accommodated during track maintenance, a maintenance window of 5.5–6 h would be optimum in terms of total possession requirement.

- Choosing a cycle length of 4 years is optimum from the possession viewpoint, but extending the cycle length to 6 years can be an optimum solution from the possession and ballast life span viewpoints.

- Improvement of tamping speed creates the largest potential for track possession time reduction. However, the increase in tamping speed is limited by design if durable track quality is to be ensured.

- A total reduction of up to 45% in track possession time per tamping cycle is achievable by using the existing tamping machine.

The methodology for planning and scheduling track geometry maintenance presented in this article can be adapted for other track sections and networks provided required data are available. In future, the methodology will extended for possession management of other large-scale maintenance tasks.

Acknowledgment
The authors wish to thank Trafikverket and Luleå Railway Research Centre for financial support.

References
Analysis and possession scheduling of maintenance tasks: a case study of conditional failures on Swedish iron ore line

Analysis and possession scheduling of maintenance tasks: a case study of potential failures on the Swedish iron ore line
Stephen M. Famurewa, Arne Nissen and Uday Kumar

Abstract
Condition based maintenance scheduling is a promising approach towards effective track possession management in railway transport. The increased use of condition monitoring and inspection technologies on railways has improved our knowledge of the actual status of the system. This has significantly helped in moving from predetermined maintenance of railways to condition based maintenance. However, some of the inspection remarks require that intervention must occur before a set deadline to prevent reduction of the operational capacity of the railway. Therefore there is need for a data-driven scheduling approach to efficiently use available train-free periods for restoration of potential failures such that availability and capacity are maximised. In this article the authors present an analysis of inspection remarks. In addition a short-term maintenance scheduling problem is formulated to support the effective and efficient scheduling of maintenance works that are not accommodated in the long-term plan. The formulated problem focuses on reducing the sum of maintenance cost, possession cost, window start-up cost and penalty cost. It is modelled as a quadratically constrained mixed integer programming problem and solved using a branch and cut algorithm. A case study on the Swedish iron ore line is used to demonstrate the use of the method for effective track possession management. The scheduling procedure yielded an optimal solution for supporting effective maintenance scheduling and possession utilisation.

Keywords: scheduling, maintenance, infrastructure condition, integer programming, track possession, train-free window

Introduction
Railway transportation is a sustainable, safe and cost-effective mode of transportation with notable contribution towards economic expansion and mobility of people and goods. The expansion of economic activities and the increasing mobility of people have led to higher axle loads, increased speeds and tighter train movements that leave little room for daytime maintenance. In EU 15 countries rail freight tonnage-kilometre and passenger-kilometre have increased by 15% and 28% respectively, over 1990–2007 [1]. Therefore infrastructure managers (IM) are concerned with increasing the competitiveness of railway transport through capacity and service quality enhancement.

Effective track possession management is essential for supporting the design capacity of existing networks. Moreover, it plays a key role in meeting additional freight and passenger traffic capacity demands without compromising safety and quality requirements. The methodology suggested in the standard [2] measures the consumed capacity of a line section by adding infrastructure occupation within a defined time, time supplements for timetable stabilisation and maintenance requirements. Thus this approach describes the state of a line section in terms of its possession and availability, as shown in Figure 1.
Furthermore, track possession for maintenance and renewal of railway infrastructure varies, depending on the type of work, required resources and machinery. Generally, the possession requirement for track works can be summarised by adapting the conventional maintenance overview in the railway transport context, as shown in Figure 2. In other words, possessions for track works include possession for inspection and restoration of potential failure, possession for immobilising or functional failure, possession for large-scale or planned tasks and possession for renewal works.

Maintenance planning and scheduling are essential elements of the maintenance management process which defines the tasks to be performed, analyses them to determine the required information and resources, and identifies and assigns the needed support efficiently [3]. The scheduling of these tasks is based on a priority system to ensure that the most urgent and important tasks are performed first and resources are utilized efficiently. An overview of the general structuring of the railway infrastructure maintenance planning process was presented in [4]. It gave a state-of-the-art view on two vital aspects: degradation modelling and scheduling of works for track possession.
A study across the railway IM in Europe confirms that long possession periods for maintenance are planned 18–24 months in advance to ensure minimal disruption to traffic [6]. However, short possession periods are requested within short timescales to restore potential failures reported during inspection and condition monitoring. Such inspections include visual inspection or non-destructive testing such as ultrasonic inspections, eddy current check, track geometry measurement and laser inspections [7]-[9]. Generally, inspection and condition monitoring of railways are primarily based on the traffic volume and line speed [7]. In addition, inspection and condition monitoring frequency are influenced by type of transport (e.g. dangerous goods), climate and inherent environmental conditions, geotechnical conditions, technical structure, designed performance, age, etc. The reported potential failures or remarks are classified into different priority levels (acute, weekly, monthly, next inspection and yearly based) based on the actual condition and risk of not fulfilling the expected functional performance. Further, restoration works are often not accommodated in the long term maintenance plan, and therefore require short-term maintenance schedule such that there is minimum influence on traffic operation and possession time is used efficiently.

Data-driven maintenance planning and scheduling models for effective track possession management in railways have been addressed in different studies. Higgins [10] addressed an aspect of the possession problem for determining the best allocation of railway maintenance activities and crew to minimize train disruption. A methodology for dividing a railway network into working zones that will be taken out of service to carry out maintenance activities was presented by van den Hertog et al [11]. Cheung et al. [12] developed a track possession assignment programme for assigning railway tracks to a given set of scheduled maintenance tasks considering defined constraints. A time-space network model was presented by Peng et al [13] to solve the track maintenance scheduling problem by minimizing the total travel costs of the maintenance teams and the impact of maintenance projects on railroad operation. Miwa [14], Andrade et al. [15] and Famurewa et al [16] addressed preventive maintenance scheduling programme related to track geometry quality and tamping operation using different approaches.

Furthermore, a preventive maintenance scheduling programme was presented by Budai et al. [17] to merge routine tasks and projects for a link over a certain period such that the sum of possession costs and maintenance costs is minimised. Zhang et al [18] developed a maintenance cost model and suggested an enhanced genetic algorithm approach to produce an optimal monthly schedule for maintenance works of one or more teams assuming that the deterioration of the segments are probabilistic. Malin et al [19] developed a mixed integer programming model that optimises a production plan and suggests the best possible traffic flow given a fixed set of planned maintenance activities. An optimisation-based possession assessment and capacity evaluation decision support tool was designed by Savelsbergh et al [20], to evaluate schedules of planned maintenance and renewal work on rail infrastructure. Finally, a review of planning and scheduling techniques of preventive maintenance activities of railway can be read in [21], [22].

There is need to use a model based approach for short-term scheduling of inspections, potential failure and deferrable failure maintenances, which are not included in the long term maintenance plan. To this end, this article presents an analysis of potential failure reports and an approach for the use of available train-free possession windows for maintenance. A short maintenance schedule problem is formulated to handle railway inspection remarks such that the challenge of temporary speed limit can
be reduced while using possession time efficiently. The formulated problem focuses on reducing the sum of maintenance cost, possession cost and penalty cost.

Case study

Track and traffic description:
A track section in the network of the Swedish Transport Administration (Trafikverket) is considered in the case study. The line section is 130 km long single track from Kiruna to Riksgränsen. The traffic on the line is mixed, with speed of 60 km/h for loaded iron ore freight and up to 120 km/h for passenger trains. The maximum allowable axle load on the line section is 30 t, the average daily tonnage is approximately 90,000 t and the annual accumulated tonnage is approximately 30 MGT. Heavy, long and slow running trains make track possession and capacity enhancement a challenging issue for the IM. In addition, the anticipated increase in traffic volume on this track section requires an efficient maintenance practice such as availability on demand. This requires that maintenance works be fitted within short track possession periods and around the demands of freight and passenger traffic.

Data description
The data used in this study include historical potential failure data, expert assessments of failure records, train movement data and cost data. The potential failure data on the selected track section between 2010 and 2013 and the train movement data in 2013 were obtained from the database of the Swedish Transport administration. The train position data recorded at some operational zones were processed to determine train-free windows that can be used for maintenance. For generating a short-term condition based maintenance schedule, 51 windows over a period of one month were considered usable from traffic, safety and resource-availability perspectives. The selected windows vary in duration from 1 to 3 h with an average size of approximately 1½ h and about 75% of the windows were smaller than this average value. Furthermore, 50 maintenance tasks were selected from the historical records of potential failure; these represent the expected monthly workload. The tasks included in the monthly workload are S&C, overhead wire, rail, fastener and signal repairs, as well as ballast and sub-ballast spot tamping. Based on expert experience and records, the possession requirement of each task is estimated and the requirement varied between ¼ and 3 h, depending on the type of work and estimated extent of damage. The maintenance cost per hour is estimated to be 217 € based on expert information and existing contracts of the infrastructure manager (IM). The possession cost per hour is approximately 80% of the maintenance cost while the penalty cost per day is estimated to be approximately 5 times the maintenance cost per hour. The fixed cost per window and fixed cost per task are half of the maintenance cost per hour.

Method
The method employed in this article can be divided into two parts; graphical analysis and the development of a short-term maintenance scheduling model. Historical potential failure records were analysed to extract the key features of the records and obtain an overview of the information content in a way that is useful for maintenance planning. The second part is the core of the study, and it involves formulation of a short term maintenance scheduling model that can be used for possession
management of potential failures. The track section is divided into 10 maintenance segments for logistic and operational purposes. Only the segment used for maintenance is considered occupied during a given window; thus only tasks on the same segment can be merged during a window to prevent shutting down of the entire section and to avoid too long travelling times.

Model formulation

A short-term maintenance scheduling problem is formulated to handle railway inspection remarks such that the challenge of temporary speed limit or other capacity limiting measures can be reduced while ensuring efficient use of possession time. The formulation of the objective function and constraints of the model are described below.

The objective function minimizes the total cost, which is the sum of the direct and indirect maintenance costs. The direct maintenance cost is simply referred to as maintenance cost and it consists fixed cost per task and variable labour cost that depends on the estimated possession time required for each task. The indirect cost consists of variable possession cost, fixed window start-up cost and penalty cost. This objective function is given by equation (1)

\[
\min \sum_{m \in M} \sum_{w \in W} f_{mw} x_{mw} \\
M = 1, 2, \ldots, M, \ W = 1, 2, \ldots, W
\]

Where \( f_{mw} \) is the aggregate cost of using window \( w \) on day \( d_m \) for task \( m \) with deadline on day \( D_m \). \( x_{mw} \) is the decision variable for carrying out task \( m \) during window \( w \). \( x_{mw} \) is a binary variable where 1 means that task \( m \) is implemented in window \( w \) and 0 means otherwise. The aggregate cost of using window \( w \) for task \( m \) depends on the size of the task, deadline of the task and the cost parameters given in equation 2. In explicit terms, \( f_{mw} \) depends on the time \( t_m \) required for implementing task \( m \), maintenance cost per hour \( C_m \), fixed cost for starting a task \( C_{mst} \), hourly cost for window possession \( C_w \), fixed charge as window start-up cost \( C_{ws} \) and the daily penalty cost \( C_p \) for exceeding the task deadline. The first two terms in equation 2 are the direct maintenance cost while the last three are the indirect maintenance cost.

\[
f_{mw} = c_m t_m + c_m v + c_m k_m + F(x_{mw})(c_m v) + F(d_m - D_m) c_p
\]

where

\[
F(x, a) = \begin{cases} \frac{(d_m - D_m)}{a}, & d_m - D_m > 0 \\ 0, & d_m - D_m \leq 0 \end{cases}
\]

\[
F(x_{mw}) \in (0, 1)
\]

s.t.

\[
\sum_{m \in M} F(x_{mw}) = \begin{cases} 1, & \sum_{m \in M} x_{mw} \geq 1 \\ 0, & w \in W \end{cases}
\]
An important aspect of this model is penalty cost modelling. Penalty cost is introduced to efficiently use an available window such that planned works are distributed and merged into the window. The conventional practice during inspection is to assess the infrastructure condition and provide priority remarks on the urgency of required intervention. This can be taken as limit which when exceeded requires extra measures that decrease the capacity/performance of the concerned segment. The daily penalty cost adopted is calculated from the estimated delay consequences based on reduction of line speed from 120 km/h to 70 km/h. In the case study, a penalty cost of 1000 € per day is imposed.

The objective function is subject to the constraints explained below:

Constraint 1: Implementation of maintenance tasks within a window should not exceed the window duration. This is presented by equation 5, where $t_m$ is the time required to fix remark $m$, $t_w$ is the duration of window $w$ and $Trt_{mm'}$ is the time required to travel between the locations of two task $m$ and $m'$ on the same segment. An average travelling time of 10 minutes is used in the case study.

\[
\sum_{m,m' \in M \atop m' \neq m} (t_m \cdot x_{mw} + Trt_{mm'}) \leq t_w, \quad w \in W
\]

Constraint 2: All tasks must be completed, i.e. a task expected to take $t_m$ hours should have a total sum of $t_m$ hours. This constraint is defined by equation 6.

\[
\sum_{m \in W} t_m \cdot x_{mw} = t_m, \quad m \in M
\]

Constraint 3: This constraint is introduced to reduce travelling within a possession window. It ensures that only repair tasks that are close to each other and on the same segments are merged in a window. This is practical for operation viewpoint, because a segment can then be occupied for maintenance without completely stopping traffic on the entire line. The possibility of rerouting and redirecting will be slim if two or more segments are occupied for maintenance in the same window. Equation 7 describes this constraint, where $m$ and $m'$ are two different tasks on segments $s_m$ and $s_{m'}$ respectively. $N_s$ is the total number of segments. This constraint is handled as a quadratic constraint.

\[
x_{mw} \cdot x_{m'w} \begin{cases} 0, & s_m \neq s_{m'} \\ x_{mw} \cdot x_{m'w} & \end{cases}
\]

\[
m,m' \in M, \quad m \neq m' \quad w \in W \quad s_m, s_{m'} \in S \quad S = \{1, 2, \ldots, N_s\}
\]

The boundary condition of the variables is defined in equation 8 below.

\[
x_{mw} \in \{0,1\} \quad m \in M \text{ and } w \in W
\]

However, for the alternative approach where the problem is solved as a simple mixed integer linear program, the quadratic constraint in equation 7 is replaced with a new linear constraint presented in equation 9. The new constraint ensures that only one task can be carried out in a window. In addition,
the constant term $f_w$ is modified and presented in equation 10. This alternative approach only gives
the baseline solution for comparison.

\[ \sum_{m \in M} x_{mw} \leq 1 \quad m \in M \text{ and } w \in W \quad (9) \]

\[ f_{mw} = F(x_{mw})(c_{m}t_{m} + c_{mw}) + F(x_{mw})(c_{m}t_{m} + c_{mw}) + F(d_{w} - D_{w})c_{p} \quad m \in M \text{ and } w \in W \quad (10) \]

**Solution**

The proposed model has a linear objective function and a combination of linear and quadratic constraints. Given that the variables are binary, the model is treated as a mixed-integer quadratic constraint program (MIQCP), a special case of mixed-integer program (MIP). Many solvers can be used to solve this problem, however Gurobi optimizer was selected in this study because of its accessibility and performance record on the public benchmark test set, e.g. fast solve time to feasibility and optimality.

The MIP models referred to hereafter as models 1 and 2, are solved using a branch and cut algorithm that combines the advantages of a pure branch and bound scheme and the cutting planes scheme. The branch-and-bound algorithm involves systematic enumeration and exploration of a set of candidate solutions or branches that are subsets of the solution or tree and application of the lower bounding method to each candidate solution. The cutting planes tighten the formulation by removing undesirable fractional solutions during the solution process without creating additional sub-problems. A detailed description of branch and cut algorithm can be found in [23], [24]. Additional guidelines for implementation of the algorithm within the Gurobi optimizer are available in the reference literature of the optimiser [25]. The optimizer uses either the linearized outer approximation approach with the simplex algorithm or the continuous QCP relaxation approach with the barrier algorithm for both the root and other nodes in the branch and cut tree.

In model 1, the continuous QCP relaxation approach is used at the root and other nodes of the tree, and the sub-problem at the nodes are solved using the barrier algorithm. In model 2, the operation of the optimisation engine is modified for reliable improvement of the solution by pre-linearizing all quadratic terms in the model. This is achieved by introducing new variables to replace the quadratic terms and new constraints such that the original problem remains unchanged. The sub-problems at the tree nodes are then solved using continuous LP relaxation with simplex algorithm. Furthermore, an alternative approach named model 3 is used to model the problem using the simple mixed integer linear progam (MILP). This is done by removing the quadratic term in the quadratic constraints as explained earlier to obtain a baseline solution for comparison.

**Results and discussions**

The result of the analysis and possession scheduling of potential failures is presented in this section with further discussions. The first part of the results presents important information and key features extracted from historical inspection records in a context useful for maintenance planning. The second
part presents the result of using the proposed model for maintenance scheduling and efficient possession management.

Analysis of potential failure records

The state and maintenance needs of railway infrastructure can be assessed using historical data of field inspections and condition monitoring runs. The frequency of annual potential failures observed and reported for the track section between 2008 and 2013 are shown in Figure 5. The figure also shows the proportions of different priority levels that can be considered as indicator of the urgency of necessary intervention work. There is a variation in the total frequency of potential failure over the years, with the highest frequency being recorded in 2009 and lowest in 2013. This can be an indication of infrastructure performance improvement despite the reported increase in the traffic volume over the years, provided the functional failure reports has a similar trend. Remarks that should be fixed within 3 months and 2 weeks have the highest proportions with annual averages of 39% and 59% of the annual total remarks, respectively. In addition, an average of 2.5% of the annual total remarks require immediate intervention because the risk of not fulfilling the expected functional performance and that of harm are assessed to be intolerable. In such cases, the intervention measures are line closure or speed restriction until restoration.

![Figure 3: Count of condition failures with priority levels on studied track section](image)

Railway infrastructure being a linear asset, the authors suggest the division of the studied track into maintenance segments for scheduling purpose. Thus, the potential failure frequencies of the 10 maintenance segments on the track section are shown in Figure 4. Furthermore, the monthly count of potential failures with acute, weekly and monthly priorities for the 10 maintenance segments are shown in Figure 5. In principle, this supports the identification of maintenance-significant segments from logistic support and operational viewpoints. In Figure 4, the two extreme occasions are shown, i.e. 2009 with the maximum potential failure counts and 2013 with the minimum potential failure counts. Segment 8, 1 and 10 are maintenance-significant segments because the potential failure counts of these segments are the highest. This information is useful for improving maintenance logistics, planning and scheduling.
The bars in Figure 5 represent the range of monthly counts of potential failures that are assessed to require either immediate intervention or intervention within one month for each maintenance segment. This is an indication of monthly workload or maintenance need on each segment in connection to standard inspections procedures for different items on the network. It is required that these track works are efficiently scheduled using available train free windows so that operational availability and capacity are not affected. Segments 8, 1, 10 and 7 have higher maintenance needs and could eventually turn out to be bottlenecks if not given adequate attention. In essence, this characterisation provides the overall picture of the status of the track section, as well as useful information about track possession planning for maintenance works.

Figure 5: Monthly workload per maintenance segment from potential failure records between 2008 and 2013
Further analysis of all potential failure reports down to the items/assemblies level is shown in Figure 6. The maintenance-significant items/assemblies in the two extreme years—2009 and 2013 are shown in the figure with switches and crossings (S&C) noticeable clearly. The proportion of S&C in the total report spanning 6 years is approximately 61%. Point machine, switch-blade position detector and crossing are the most maintenance-significant items based on the counts of reports related to them. The failure modes responsible for a large proportion of the potential failure reports of S&C on the track section are shown in Figure 7. The respective intervention measures for the reports are shown in the figure, with adjustment, welding and grinding being the most reported. This information is useful for organising the maintenance team and for effective logistic planning to cater for similar maintenance work orders.

Figure 6: Characteristic features of remarks in 2009 and 2013

![Figure 6: Characteristic features of remarks in 2009 and 2013]

Figure 7: Characteristic features of S&C remarks

![Figure 7: Characteristic features of S&C remarks]
**Maintenance schedule for potential failure**

The results of using the proposed model and alternative model for efficient possession management of potential failure and deferrable failure maintenances are presented below. The short term maintenance schedules of these models are evaluated in terms of their computational times, solutions obtained, constraint violations, optimum values, number of delayed tasks, number of days with capacity reduction, average window utilisation and number of windows used.

The overall performance of the models are summarised in Table 1 and further elaborated thereafter. Models 2 and 3 generated optimal solutions in less than 1 min while model 1 yielded the best feasible solution in approximately 8 min. The optimality of the solutions of models 2 and 3 were proven because the gap between the best feasible solution and the incumbent optimal solution in the optimisation algorithm is equal to zero. Model 1, which used the continuous QCP relaxation with barrier algorithm, gave a feasible solution however optimality was not reached before the time limit of 3 h due to large gap between the feasible and the incumbent optimal solutions. None of the models violated their respective constraints, i.e. all tasks were completed, window durations were not exceeded and tasks in different segments were not scheduled together.

The performances of the models were analysed further by studying the solutions they yielded. Model 3 is an expensive approach, in the sense that it does not permit combinations of tasks on the same segment into one window. Therefore, the associated total maintenance cost is the highest among the three models. Model 2 is associated with the minimum total maintenance cost while model 1 is very close to model 2 and far better than model 3 in terms of cost. Looking further into the schedule generated by each model and comparing it with their respective deadlines, model 2 has the best performance with all works scheduled and no task delayed. In model 1, only one task would be implemented after the deadline, while in model 3, 4 tasks would be delayed.

In terms of the number of days for which capacity would be affected owing to infrastructure conditions, model 3 has the worst performance while model 2 has the best performance, i.e. no reduction in capacity. The average window utilisation is the highest for model 2 owing to the possibility of merging maintenance tasks in a single window. Even though none of the models led to 100% window utilisation, the proposed model (model 2) showed better performance and can even be improved if some tasks can be broken down and the constraints of task combination are relaxed. In terms of the number of windows utilised, models 1 and 2 utilise less windows to complete all the tasks, leaving behind four unused windows that can be used for other purposes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution</td>
<td>MIQCP</td>
<td>MIQCP</td>
<td>MILP</td>
</tr>
<tr>
<td>Constraint violation</td>
<td>Nil</td>
<td>Optimal</td>
<td>Optimal</td>
</tr>
<tr>
<td>Optimum value (€)</td>
<td>34420</td>
<td>33267</td>
<td>40168</td>
</tr>
<tr>
<td>Number of delayed works</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Number of affected days</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Average window utilisation</td>
<td>85%</td>
<td>87%</td>
<td>82%</td>
</tr>
<tr>
<td>Number of windows used</td>
<td>47</td>
<td>47</td>
<td>50</td>
</tr>
</tbody>
</table>

* A feasible solution was obtained at 8 min however the optimisation was terminated after a set time limit of 3 h
In addition to the overall performance evaluation of the models given in Table 1, a breakdown of the total maintenance cost for the optimal task schedules generated by the three models is given in Figure 8. The total direct maintenance cost $C_{\text{maint}}$ and possession cost $C_{\text{poss}}$ are similar for all models because these cost elements are functions of estimated repair time, and all tasks are expected to be completed in a window. However, the distinct differences between the optimality of the models are the total penalty cost $C_{\text{pen}}$ and window start-up costs $C_{\text{wind-st}}$. The schedule generated by model 2 has no penalty cost because no task is delayed and its window start-up cost is small because the schedule minimises the number of used windows.

![Figure 8: Breakdown of the total maintenance cost for the proposed model and alternative models](image)

The initial window duration and left-over time in each window for model 2 are shown in Figure 9 for visual assessment of the possession allocation efficiency. The obviously high left-over times represent the unused windows. Approximately 80% of the remaining windows can be considered practically unusable for new repair works because they are too small to accommodate travel times and start a new task.

Furthermore, the left over window duration can be analysed and classified, as shown in Figure 10 for unplanned opportunity based track works such as manual inspection and routine checks. The class A windows are efficiently used and perhaps not usable for other works if encroachment into the maintenance withdrawal time before the next train is to be avoided. The class B windows can still be used for opportunity based maintenance involving small-scale track works, routine checks or inspection on the same segment where the window time it was originally used. The class C windows are unused and can thus be used for any type of work on any segment provided other utilisation constraints are not violated.
An important aspect of the proposed approach for possession management is the analysis of the optimal schedule or reason for infeasibility. In instances where not all tasks can be scheduled in available windows owing to the number or the size of the windows, a review of the task can be conducted. For instance, the review could entail the possible break-up of some tasks into smaller chunks or removal of the less significant works that will be later spread over the left-over usable windows. The model can be adapted with little improvement to support other scheduling cases, including night possession with long duration, where merging of tasks in different segments is allowed within the same window. In future, the model will be extended to consider task implementation order and other technical or logistic conditions related to different tasks. It will also be extended to multiple track scenarios with additional information about the track layout from the asset information system.
Conclusion
This article presents an analysis of historical inspection reports from a freight corridor to extract useful information for maintenance planning and preparation. In addition, it describes the formulation of a short-term maintenance-scheduling problem to support the effective and efficient scheduling of maintenance works that are not accommodated in the long-term plan. The formulated problem focuses on reducing the total maintenance cost. The conclusions from the case study are as follows:

i. Maintenance segment 8 is a maintenance-significant segment and should thus be considered during maintenance planning, scheduling and other logistic support tasks.

ii. Adjustment, welding and grinding of S&C are the most frequent maintenance tasks on the line section under consideration; thus, their logistic support in terms of personnel and resources should be given priority.

iii. Possession scheduling for maintenance works can be supported with the proposed MIQCP model presented in this work.

iv. The MIQCP model with continuous LP relaxation approach gives the best performance with the lowest cost, zero task delay and zero capacity loss due to infrastructure condition for the case study.

v. The use of maintenance windows for routine works or condition-based maintenance is a promising approach for possession management especially in corridors where complete night dedication for maintenance is impractical.

Acknowledgement
The authors wish to thank Trafikverket and Luleå Railway Research Centre for financial support. We also appreciate Prof. Inge Söderqvist for his fruitful discussions.

References

22. T. Lidén, "Survey of railway maintenance activities from a planning perspective and literature review concerning the use of mathematical algorithms for solving such planning and scheduling problems," Linköping University, Linköping, Sweden 2014.