RISK-BASED INSPECTION PLANNING OF RAIL INFRASTRUCTURE CONSIDERING OPERATIONAL RESILIENCE

by

MOHD HANIFF BIN OSMAN

A thesis submitted to
the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Civil Engineering
College of Engineering and Physical Sciences
University of Birmingham
January 2020

ABSTRACT

This research proposes a response model for a disrupted railway track inspection plan. The proposed model takes the form of an active acceptance risk strategy while having been developed under the disruption risk management framework. The response model entails two components working in a series; an integrated Nonlinear Autoregressive model with eXogenous input Neural Network (iNARXNN), alongside a risk-based value measure for predicting track measurements data and an output valuation. The neural network fuses itself to Bayesian inference, risk aversion and a data-driven modelling approach, as a means of ensuring the utmost standard of prediction ability. Testing on a real dataset indicates that the iNARXNN model provides a mean prediction accuracy rate of 95%, while also successfully preserving data characteristics across both time and frequency domains. This research also proposes a network-based model that highlights the value of accepting iNARXNN's outputs. The value is formulated as the ratio of rescheduling cost to a change in the risk level from a missed opportunity to repair a defective track, i.e., late defect detection. The value model demonstrates how the resilience action is useful for determining a rescheduling strategy that has (negative) value when dealing with a disrupted track inspection plan.

I dedicate this thesis to my parents, Osman Haji Abdul Rashid and Salmiah Hasan, my wife, Maria Ibrahim and my children Muhammad Ashavin and Ashavina for their endless love, support, and encouragement during all these years.

ACKNOWLEDGEMENTS

First and foremost, I am grateful to The Almighty God for the wisdom he bestowed upon me to get through the whole task of my PhD work.

I wish to thank my supervisor, Dr Sakdirat Kaewunruen for providing me with the opportunity to initiate my PhD at the University of Birmingham and for imparting his knowledge and expertise in this study. I also thank my co-supervisor, Professor Anson Jack for your contribution and practical advice.

To Mr Martin Li, The Swedish Transport Administration (*Trafikverket*), thanks for the help you provided during the acquisition of track data; track measurements and tamping record.

I take this opportunity to express my deep sense of gratitude to my beloved wife, Maria for her constant encouragement, and my wonderful children, Ashavin and Ashavina who served my inspiration and strength to finish this study. I also thank my parent and family members for their endless encouragement and support.

To Mr Amir Bin Ibrahim, thank you for copy editing my thesis for conventions of language, spelling and grammar. My thanks and appreciations also go to one and all who, directly or indirectly, helped me out with their abilities in my PhD journey.

Finally, I place on record, my sense of gratitude to the Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia for the financial supports, which led to the completion of this study.

CONTENTS LISTINGS

LIST OF PUBLICATIONS

- **Osman, M. H.** and Kaewunruen, S. (2018a) 'Value of rescheduling of rail inspection', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, pp. 1–10.
- **Osman, M. H.** and Kaewunruen, S. (2018b) 'Uncertainty propagation assessment in railway-track degradation model using Bayes linear theory', *Journal of Transportation Engineering, Part A: Systems. American Society of Civil Engineers*, 144(7), p. 04018026.
- **Osman, M. H.** and Kaewunruen, S. (2018c) 'Reschedule or not? use of benefit-cost indicator for railway track inspection', in Caspeele, R., Taerwe, L., and M. Frangopol, D. (eds) *Life-Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision*. Leiden, The Netherlands: CRC Press, Taylor & Francis Group, pp. 2013–2019.
- **Osman, M. H.** and Kaewunruen, S. (2018d) 'Execution time estimation of recovery actions for a disrupted railway track inspection schedule', in Caspeele, R., Taerwe, L., and M. Frangopol, D. (eds) *Life-Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision*. Leiden, The Netherlands: CRC Press, Taylor & Francis Group, pp. 796–802.
- **Osman, M. H.** and Kaewunruen, S. (2018e) 'Value added strategy for unplanned rail track inspections', *Proceedings*. doi: 10.3390/proceedings2161141.

- **Osman, M. H.** and Kaewunruen, S. (2018f) ‘Role of pre-processing in textual data fusion: learn from the Croydon tram tragedy’, *Frontiers in Built Environment*, p. 30.
- **Osman, M. H.**, Kaewunruen, S. and Jack, A. (2017) ‘Optimisation of schedules for the inspection of railway tracks’, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. IMECHE, p. 095440971772163.
- **Osman, M. H.,** Kaewunruen, S., Ann, M. and Serdar, D. (2016a) ‘Disruption: a new component in the track inspection schedule’, in 2016 *IEEE International Conference on Intelligent Rail Transportation (ICIRT)*. IEEE, pp. 249–253.
- **Osman, M. H.,** Kaewunruen, S., Ann, M. and Serdar, D. (2016b) ‘Disruption management of resource schedule in transportation sector: understanding the concept and strategy’, *Procedia Engineering*, 161, pp. 1295–1299.
- **Osman, M. H.**, Kaewunruen, S., Jack, A. and Sussman, J. (2016) ‘Need and opportunities for a “Plan B” in rail track inspection schedules’, *Procedia Engineering*, 161(Supplement C), pp. 264–268.
- Sakdirat, K., **Osman, M. H.** and Wong, H. C. E. (2019) ‘Risk-based maintenance planning for rail fastening systems’, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. American Society of Civil Engineers*, 5(3), p. 4019007.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF ABBREVIATIONS

AD	Anderson-Darling
AL	Alert Limit
CBM	Condition-based Maintenance
CDF	Cumulative Distribution Function
CPT	Cumulative Prospect Theory
DM	Disruption Management
FRA	Federal Railway Administration
GJT	Generalised Journey Time
HILP	High Impact, Low Probability
IAL	Intermediate Action Limit
IL	Intervention Limit
iNARXNN	integrated Nonlinear Autoregressive model with eXogenous input Neural Network
LCC	Life Cycle Cost
LTM	Late Time Multiplier

MECA	Minimum Expected Cost Analysis
MRE	Marginal Revenue Effect
MTTF	Mean Time to Failure
PM	Preventive Maintenance
RIC	Rail Infrastructure Company
RIM	Rail Infrastructure Manager
TIS	Track Inspection Schedule
TIP	Track Inspection Plan
TQI	Track Quality Index
SD	Standard Deviation
UT	Utility Theory
VRP	Vehicle Routing Problem

CHAPTER 1 INTRODUCTION

1.1 Study background

The appropriate completion of safety-critical inspections is vital to a condition-based maintenance (CBM) program. As its name suggests, CBM is a practice of diagnosing the health of a component in a deteriorating system using measurements of the condition of the component of interest. In practice of railway safety, a track system is regularly inspected by way of various deterministic inspection activities, ranging from human visual inspections to ultrasonic vehicle and track geometry measurements for maintenance priorities. Whether it is preventive maintenance, renewal work, or no action, an appropriate decision should be justified with decision support systems, using both the results of the visual inspection and the recorded measurements. Furthermore, as stated under the Health and Safety at Work Act of 1974, it becomes a duty of the railway infrastructure manager (RIM) to provide a reliable track system, which in turn ensures the safety of both passengers and staff (RSSB, 2014). As a result, the track inspection is safety-critical and mandatory, which means that RIMs have to perform such inspection works and set a budget for inspection activities. For example, Network Rail, a railway administrator in the United Kingdom, spends almost £9 million on periodic inspections annually (Office of Rail Regulation, 2012). According to Daniels (2008), costs grow exponentially to railway network size and inspection requirements. Indeed, track inspection is a proactive action to assure track’s safety and operability, but from a business viewpoint, track possession for inspections seems to be an obstacle for RIMs to offer an additional track usage to freight companies. Note that RIMs and train companies in the European region are not eligible for inclusion under one organisation, following an endorsement of a reform model in 1990 (Caetano and Teixeira, 2013). Therefore, to gain as many benefits as possible from track inspection and measurement, scheduling theory has been incorporated into prioritising inspection tasks.

The theory of scheduling (as well as its predecessor; planning) generally enables users to gain optimal benefits from predetermined activities or tasks that are subject to a set of constraints (Tanaev, Gordon and Shafransky, 1994). From a railway management perspective, the main goal of scheduling track inspections is to minimise risks of the late detection of rail and track defects and irregularities (e.g. longitudinal track defects, poorly positioned ballast and rail cracks) by ordering tracks to be inspected (Podofillini, Zio and Vatn, 2006). Such inspections address issues related to the cost-effectiveness of track maintenance. In recent decades, the scheduling of track inspection tasks has become more challenging as a large volume of inspections is often required due to the increasing number of failures (Turner *et al.*, 2015). Hundreds of inspection tasks are performed to examine rails, including for both external and internal rail defects such as thermite welds defects (Chen *et al.*, 2006), wheel burn defects, rolling contact fatigue cracks (Kaewunruen and Remennikov, 2009; Farhangi *et al.*, 2015), and issues with track geometry alignment and component conditions (Konur *et al.*, 2014; Santos, Fonseca Teixeira and Pais Antunes, 2015). Fortunately, the periodic-style inspection being practiced nowadays reduces the complexity underlying the scheduling process.

Periodic inspections allow a track supervisor, who is responsible for various aspects of inspection (Office of Rail Regulation, 2012), to apply a master (one-off) approach for the track inspection plan (TIP). With this approach, the schedule can be prepared several months before the year of operation and instantly after the disclosure of freight planning, passenger timetable, and major possessions (Lidén, 2014). A key feature of having a master schedule is that the overall quality of inspection tasks arrangement is known in advance; company resources, such as equipment, inspection vehicles, and manpower, are ready; and track possessions can be arranged effectively.

Track inspection activities are planned according to a planner's objectives (e.g. rail safety, inspection costs) and are subject to many technical, safety, and business constraints (Lake, Ferreira and Murray, 2000; Andrade and Teixeira, 2013; Farhangi *et al.*, 2015). Because of their nature, the inspection problems are commonly formulated as constrained optimisation problems and are solved either using heuristic or exact methods to obtain an optimal solution. As different requirements exist from one track inspection problem to another—based on inspection order, railway network size and number of track components among other factors—more sophisticated and problem-dependent optimisation methods have been developed.

The presentation of constraints in the optimisation model formulation demonstrates that a planner (a scheduler) has a good knowledge of the limitations and restrictions that the plan/schedule could face in reality. However, planners have to be selective and realistic at this stage, as problem complexity increases as they incorporate problem-specific constraints (Snowling, 2000). Realising a positive relationship between uncertainty and complexity in a track inspection problem formulation, thus not all uncertainties, in particular, those related to dynamic changes in their model's real-time environment, will be taken into consideration. Trading model complexity with uncertainties, especially those with very low information, soundly exposes a one-off inspection schedule to disruption risks.

Disruptions are unexpected, have a low probability of occurrence and their consequence level is ambiguous. Furthermore, their presence is abrupt and mostly beyond the control of the system owner. Some might argue about the relevance of such disruptive events—for example, a worker strike—to TIP. It must be considered, however, that some rail cracks that are not detected by an ultrasonic inspection vehicle might then be detected through manual visual inspection. In the case of a worker strike, those cracks would remain undetected, leading to aggressive impact loading and subsequent track damage, which could then cause a train accident (Remennikov and Kaewunruen, 2008). Imagine that a freight train carrying hazardous materials or petroleum products gets involved in an accident near a plantation area. Fire is easily triggered from the accident and turns the area into ashes. Losses occur not only on the rail management side but also to third parties outside the railway domain. As stated in Lee, Preston and Green (2012), three days of operations closure is enough to trigger significant social and economic impacts.

Climate-related extreme weather events are also in a potential list of disruption (Kaewunruen, Sussman and Matsumoto, 2016) where the Malaysian east coast train line closure in late 2014 is among the recent examples (Davies, 2015). The natural disaster from heavy floods not only disrupted passenger timetables but also scheduled inspection and maintenance activities. Landslides (CNN, 2010), train accidents (The Guardian, 2016) and poor contingency plans (BBC UK, 2015) are real disruptions that have happened in the Europe region in the last decade. Terrorist attacks, which have not been a threat to railway operations until 20 years ago, could prevent some tracks from receiving inspections. During the closure time, unaffected tracks will be increasingly used to minimize losses in business, and consequently, those tracks require additional inspection and maintenance as well. Losing crews, machine breakdown and special inspection requests potentially present disruption to inspection schedule execution in various ways.

In both flight and train services, the effectiveness of the business operation—i.e. an optimal revenue point—begins to deteriorate if there is a disruption during the day of operations

(Zhen *et al.*, 2016). Under disruption attack(s), the feasibility of prescribed schedules experiencing performance deteriorates. Hence, the development and maintenance of robust schedules and contingency plans are a few respond strategies for disruption risks. According to Cacchiani et al. (2014), contingency plans are beneficial in handling disruptions to railway timetables. In the Netherlands, there are 1,000 pre-planned periodic timetables for various types of disruption. However, the time spent on the preparation of those plans is definitely a loss in the year (or operational period) without disruption. Theoretically, the initial cost of planning goes up as the level of protection or security increases. Besides, only an experienced planner is ultimately capable of completing a plan selection. Meanwhile, the establishment of a plan that behaves robustly towards different disruptions lacks practicality and becomes harder to implement as the time for recovery is often limited in railway systems (Fang, Yang and Yao, 2015). Realising that most disruptions are unforeseeable, many previous studies focusing on reducing the consequences of disruptions have been conducted (Chopra and Sodhi, 2004), and the focus of this study is the concept of resilience.

A system (or subsystem) is said to be resilient if adjustments to how it works before, during, and after disruptions and opportunities can be made, regardless of whether those events were expected or not (Hollnagel, Woods and Leveson, 2006; Ayyub, 2015). Generally speaking, resilience demands creative and rapid thinking from an organisation (or community) to effectively respond to disruptive events by absorbing the impacts and making changes in the existing structure without causing too many further alterations. Interestingly, managing disruptive events under the resilience framework, in principle, could be conducted without considering sources and types of events. For example, a system manager would not need to know what types of events can occur nor express their likelihoods as required in traditional risk assessments. However, in cases with high level of uncertainties related to the types of events that might occur, this is important because such risk assessments are not able to produce reliable probability estimates.

From an engineering perspective, the concept of resilience offers an alternative to conventional risk management approaches. According to Steen and Arven (2011), several features of the conventional risk assessment approach such as hindsight knowledge, failure reporting, and risk assessments calculating historical data-based probabilities are found to be inadequate for usage in the context of present-day systems. In contrast to a basic risk perspective where a probability or probability distribution is the main component, an alternative perspective shift in thinking from probability estimation to uncertainty assessments is warranted. By restricting risk to the probability assignments alone, elements of uncertainty and risk are not fully exposed. There could be a lack of understanding about the underlying phenomena, and strong assumptions may have been made to determine probability. While probability is a tool used to express uncertainties, it is not a 'perfect' tool. In the case of a very small number of historical data or rare events, traditional risk assessments fail to ensure resilience.

1.2 Motivations and problem statement

While the TIP is prepared under uncertainty for real-world settings, unexpected events linked to such uncertainties may also occur in reality. Indeed, the non-availability of inspection crews due to strike, time delays for vehicle or equipment maintenance, extreme weathers, budget reallocations or ad-hoc inspection requests have a low likelihood of occurring in reality but may occur suddenly and without warning during planned operations. In situations whereby a decision-maker has to confront an unexpected event, risks in the operation of the TIP must be identified and analysed properly before advancing to a response model selection. For a system (or subsystem) exposed to low-probability events, a decision-maker's attitude towards risk in the process of evaluating the severity of consequences and the probability of occurrence is highly recommended to be considered (Weber, Blais and Betz, 2002; Abadie, Galarraga and de Murieta, 2017). Understanding the decision-makers' risk attitudes is relevant for a risk management team to design an affordable resilience strategy against low-probability risk. A risk-averse decision-maker is one who is reluctant to accept risks, while the term 'risk-accepting' better represents a decision-maker's attitude in the opposite direction. A risk-neutral decision-maker will likely be deterred by the existence of risks alone.

The application of consequences-based decision models such as expected utility theory to risk management frameworks has supported engineering decisions in quantifying decision-makers' attitudes towards risks (Cha and Ellingwood, 2013a). The expected values of a utility function that characterise a decision-maker's attitude in a risk evaluation form the basis to categorise decision-makers quantitatively based on their willingness to accept risks. Apart from the use of decision weights, which allows for probabilities of event occurrence to be weighted in accordance to a decision-maker's perception, the utility-based decision model in principle can take any form of function to measure the consequences of an event. The life cycle cost (LCC) is a common measure of consequence in consequence-based decision models (Cha and Ellingwood, 2013a). However, the LCC cost function, which is in favour of long-term (15–30 years) analysis, is not sufficient for measuring the impacts of unexpected events, particularly on a plan or schedule, in a short to immediate time frame. For example, the damage to a railway infrastructure manager's reputation due to timetable delay or train cancellations such as the chaotic incident at the King's Cross railway station (BBC UK, 2015) was only for few days. The public will stop mentioning or commenting about an incident once its related stories are no longer discussed in the media. In this setting, the costs of protecting a company's reputation simply outweigh the related benefits if both quantities were calculated based on an LCC analysis concept. As a result, a decision-maker might struggle with or overuse limited resources to design an affordable strategy to respond to unexpected events. Treating low-probability risks according to traditional risk analysis mostly leads organisations to accept a risk by default (before business operations). To address this limitation, a measure or cost function sensitive in multidimensional directions (spaces) against the impacts of sudden changes in an operation of such systems should be incorporated into the utility function.

While an occurrence of disruptive events cannot be perfectly predicted due to epistemic or aleatory (stochastic) uncertainties, its impacts in the system under attack can be evaluated as a point in the system's performance metric (measure). For the event to either cause a gradual or abrupt reduction in the system's performance, the system needs to be restored before the system completely fails, depending on the resources employed. Framing status transitions between the levels of a system's performance within a resilience framework allows for the performance metric to be characterised by the evolution of a system's status over time. This approach provides a generic platform to incorporate temporal dynamics into the performance metric. At times, a resilience-based performance metric is used to define a utility function and a precise quantification of a decision-maker's utility against low-probability risk can then be presented. As a result, an accurate tipping point of the risk-aversion parameter—the point at which the response strategy becomes preferable to maintaining the status quo based on maximum expected utility approach—is determined, thereby protecting the RIM from excessive spending in the area of risk-reduction costs. This calls for the need to formulate a TIP as a system or subsystem in a system of track maintenance and followed by the incorporation of resilience science in the assessment and management of low-probability risk events.

1.3 Aim and objectives

The aim of the study described in this thesis is to develop a risk-resilience response and recovery model for a disrupted planned rail track inspection. To help achieve that aim, the following objectives have been completed:

1. A literature review has been conducted on mathematical optimisation model formulation for track inspection plan and schedule as well as relevant subjects such as condition-based track maintenance, risk analysis and evaluation, operational resilience, disruption management, deterioration analysis, and artificial neural network and Bayesian inference.
2. A risk-resilience methodology for managing disruptive events in a TIP has been established.
3. An affordable time-series prediction model for track geometry measurement has been developed to absorb an initial impact of disruptive events. The absorptive capability of the proposed model is given by a time-dependent resilience metric and also the value of the method of fusion.
4. A recovery model (resilience action) in the formulation of graph-based track inspection rescheduling has been developed; this model will also be used to measure the applicability of the prediction model (refer to the third objective) in promoting operational resilience.

5. The response and recovery models built from the outcome of the third and fourth objectives were validated and tested by carrying out an integration of simulation and sensitivity analyses and using real data.

1.4 Research questions

The central question of this thesis is: *How can a method of fusion and data analysis be embedded to develop an affordable response and recovery model in the occurrence of disruptive events in a planned TIP?* As a consequence, the following sub-questions, which have been tailored according to the research objectives, must be answered:

1. To what extent has the scheduling approach been applied to construct a periodic plan (and schedule) for track inspection?
2. What are the characteristics that confer the uniqueness and/or similarities of TIPs when compared to other periodic plans?
3. How can uncertainties in an operation of TIP be determined by formulating the inspection problem as a mathematical optimisation problem?
4. To what extent are the components of the resilience concept meaningful to an existing risk management framework, particularly when dealing with low-probability events?
5. What is a reliable platform to develop a prediction model for a temporal data series? Is it possible for the model to be accurate and have good generalisation but in a simple and functional form?
6. What are the requirements for incorporating a rescheduling framework into the recovery phase of a resilient system?
7. What time-varying value function should be used to compare the costs involved and the expected benefits from re-planning an affected inspection plan?

8. What are the trade-offs between the level of managerial decision and the model values given that this study aims to propose a resilient track inspection system?

1.5 Thesis statement

The establishment of a smart partnership between artificial intelligence and disruption management offers an affordable response and recovery model for a disrupted TIP, which can lead to a positive impact in increasing the value of inspections, greater track access for freight train business, an improved environment (low carbon footprint), and a reduction of workplace accidents while also providing a reasonably practicable level of track safety.

1.6 Research hypotheses

The following hypotheses are constructed to answer the quantitative research questions:

1. A tipping point of affordability capability in the development of the response model can be determined from an expected utility theory.

2. The use of time-varying resilience metrics as risk-acceptance criteria provides flexibility to decision-makers whether to immediately implement or to delay the making of a response to a disrupted inspection plan, depending on the real-time evaluation of a company's strengths and limitations.
3. An application of artificial intelligence for spatiotemporal data prediction is reliably proven to respond to the absence of track measurement data.
4. Extending an inspection interval with or without a reduction in inspection frequency creates positive value in an inspection scheduling system.
5. The degree of changes in the established TIP depends upon the managerial decision preferences in a cost benefit trade-off analysis.

1.7 Contributions of the study

One of the ‘missing’ components in the formulation of rail track inspection plan (or schedule) is a plug-in method to respond to a disruptive event in the operation of track inspection plan. Comparing to the past decades where the number and sources of disruptions in railway operations are small and limited, embedding a track inspection plan with the plug-in method to exhibit a resilience function is an exception. Nowadays, the scenario has changed greatly; not only a disruption has an increasing trend in the number of occurrence but it also occurred in new various ways. Recent developments in respect to railway operations enable organisations to rethink the resilience of track inspection plan. One of potential strategies in promoting resilience function is turning a plug-in method a must-have component to complement conventional methodologies for planning track inspections. Hereafter, the term plug-in method is referred as a resilience model in this study.

The resilience model proposed in this study takes the form of an active acceptance risk and resilience strategy, while having been developed under the disruption risk management framework. The proposed model entails two components working in a series; an integrated Nonlinear Autoregressive model with eXogenous input Neural Network (iNARXNN), alongside a risk aversion-based value measure, respectively, for responding to and recovering from disruptions in real time. iNARXNN is a prediction model acts as a response to absorb the performance losses in a track inspection plan due to negative impacts from a disruptive event. The neural network fuses itself to time series analysis, deterioration analysis, Bayesian inference, and a data-driven modelling approach, as a means of ensuring the utmost standard of prediction ability, while also enabling a simulation of the effects of a sudden shift in track deterioration. At this point, a track engineer who responsible for making a decision at an operational level will be provided with logical sense explanation on the valuation of iNARXNN based on information it carried. In this research, value of information is proportional to a deviation of upper bound of uncertainty of a deterioration rate of track quality measures. Since the study does not deal with a precision calculation of uncertainty, unnecessary computational and elicitation effort is avoided through an application of Bayes Linear Method as an agent of information processor. With risk-return relationship is a core in business decision making thus the benefit of the response model is formulated in respect to risk aversion of low probability events and of course, in terms of monetary value. The risk aversion-based measure is formulated in order to quantitatively evaluate the benefits and costs of accepting the iNARXNN's outputs for an input of track quality assessment. Rescheduling framework was introduced to demonstrate of the performance of the iNARXNN results in the context of recoverability capacity. Various problem scenarios have been setup to address any limitations in the rescheduling framework.

An ultimate achievement of the proposed resilience model is to flexibly adapt resilience strategies to a planned inspection schedule that has been exposed to disruptions. Flexibility in decision-making about how and when to respond to disruptive events subject to (near to) real-time evaluations of organisations' financial status, operational capacity, and safety; traffic volume; and train passengers' sentiments regarding viral issues such as carbon emissions, passenger train delays, etc. In a specific situation, the risk-resilience model would able to recommend a reduction in the total number of inspections without compromising track safety requirements that should be presented over a period of time. Considering cost savings, it has been estimated that around £40 per one inspection reduction could be saved from the original inspection expenses. In parallel, rail infrastructure managers will enjoy substantial benefits from the reduction such as more track access for freight train services and the elimination of some contributors to train delays. Interestingly, no capital investment for installing, replacing, or purchasing any new equipment is required to embed the risk-resilience model into a planned track inspection plan. Finally, the development of both response and recovery models also promotes data innovation in railway management maintenance. In addition, use of freely available data in the RIM's data repository reduces time and costs of resilience.

CHAPTER 2 LITERATURE REVIEW

2.1 Railway track inspection

Railway infrastructure managers from around the globe spend millions of dollars annually on CBM, in order to avoid service failures from occurring (Peng, 2011). As its name suggests, an up-to-date i.e., near to real time condition status of the track (or track components) is gathered at meaningful time points to verify whether the inspected track is no longer in nominal condition or a fault is impeding. The sooner the failure symptoms (defect) can be detected, the longer time engineers have to design an effective maintenance upon an inspected track. For deteriorating assets, there are always detectable symptoms or warning signs prior to the state of failure (Blann, 2013). Thus, a failure can be viewed as a terminal point in a process that starts with a failure symptom(s) where it evolves progressively over time (or measurable quantity in general). Due to this mechanism, the condition of track (or any deteriorating asset) is inspected and measurements are recorded through predetermined inspection activities which can range from human visual inspections to track (geometry) measurement. Appropriate maintenance works prior to failure are then decided in a decision making process, where analysis results of the measurements as well as inspection data are the primary source of information. A track inspection does not only assure the track's safety and operability, but also contributes to the ride comfort improvement.

In general, for railway track maintenance, presenting a well-prepared inspection strategy is one of the prerequisites to gain maximum benefits from the investments made in inspection dependency maintenance (Khouy *et al.*, 2016; Stenström *et al.*, 2016). Periodic inspection is basically performed for a repairable asset e.g. railway tracks, to verify its safety and performance at determined time intervals. Maximum length of inspection interval (minimum number of inspections) is basically subject to accumulated traffic tonnage and speed category (Esveld, 2001). Periodic inspection is relatively simple to implement and probably is the most common inspection strategy applied in practice to repairable systems (Peng, Ouyang and Somani, 2013). For example, in the United Kingdom, periodic inspections is still the primary way to measure and gather track geometric characteristics and track structure condition data. On contrary, online condition monitoring may be the best for critical, high-values assets and has a short potential-functional failure interval (Roberts *et al.*, 2014). With none of the features, RIM will suffer a high capital investment for system acquisition, office arrangement and safety, data management and personnel training.

2.2 Track inspection plan

Unlike previous decades, track inspection is more demanding and expensive nowadays, due to the increasing number of failures (Lidén, 2015; Turner *et al.*, 2015). Hundreds of inspection tasks where each task roughly takes one to four hours are performed in order to maintain the target levels of track reliability, availability, safety and ride comfort (Al-Nazer *et al.*, 2011). Hence, it is important to perform inspection tasks systematically and objectively as it incurs a track possession cost. Longer possession interval results in higher possession cost, particularly on heavily traffic section/line. For rail lines with mixed passenger and freight traffic, unavailable time slots were likely to have been sold to a freight train operator (Putallaz and Rivier, 2003; Budai, Huisman and Dekker, 2006). On top of that, train timetables are given priority to maintenance-related tasks in track possession. This long-run policy states that track possession for maintenance works is allocated last when the identified tracks are unattended by both passengers and freight trains unless RIM is willing to pay a compensation to an affected train timetable e.g. train cancellation and/or delay charge (Network Rail, 2008). To gain maximum benefits from the investments made in track inspection, e.g., perform track monitoring and repair at the right time, inspection activities are properly scheduled with respect to the railway company's objectives and are subject to many complex constraints.

A schedule may be described as a sequence of tasks or activities that will be sequentially performed for a given time period. The feature gives two options for the track supervisor: to either prepare a prescribed (master) inspection schedule or do it partially as an interval-based routine. The former scheduling mode is the practice of producing a complete schedule before the beginning of a business operation period. Under the time-rigid category, identified tracks under RIM supervision will know in advance about the time and inspection tasks that will be performed on them. In addition, a prescribed schedule offers other benefits, such as the schedule's objectives being known prior, in real-time status of company resources, e.g. manpower and equipment are always available and the planning team have to experience the exhaustive schedule design process only once.

In any cases, from an RIM perspective, the main goal of scheduling track inspections is to maximize the probability of recording irregularities in track condition data from inspection activities by optimally ordering the tracks to be inspected (Podofillini, Zio and Vatn, 2006). Scheduling theory enables users to gain optimal benefits from predetermined activities or tasks subject to a set of constraints (Tanaev, Gordon and Shafransky, 2012). To date, track inspection plan and/or track inspection schedule (TIS) problems are conveniently modelled as an optimisation problem which incorporates various parameters of temporal, spatial, as well as physical i.e. human and machine characteristics. As different requirements exist from one track inspection problem to another-based on the inspection order, inspection interval and the size of the railway network, among other factors-more sophisticated and problem-dependent methods have been developed. Following subsection provides a review on TIS optimisation problem with the aim is to identify sources of disruptions in track inspection. The ultimate goal here is to use the review findings to develop a response and recovery model for TIS in the occasions of disruption.

2.2.1 Optimisation problems of rail track inspection

A majority of researchers formulate TIS and/or TIP as an optimisation problem. In doing so, the TIP problem should present at least one objective function to be optimised and a set of constraints, if possible. Although an unconstrained optimisation problem is less complicated to solve, the given solution might become less feasible if some changes occur during execution times. Presenting constraints in the problem formulation restricts the search for solutions only in a feasible region defined by the limitations and challenges that the schedule could face in reality.

Recent studies of a constrained optimisation problem for TIS are presented by Konur et al. (2014) and Farhangi et al. (2015). Two objectives are considered in their model: the first one is to minimise the total inspection times to complete the predetermined number of inspections in the given inspection period. The total time is a summation of total times to inspect all the tracks and travel times among the inspected tracks. In order to benefit as much as possible from the travelling decisions, a quality measure was introduced. The measure is a degree of safety importance of inspections and the study aims to maximize the safety measurement as well. An integration of Pareto optimality concepts and genetic algorithm was applied to the bi-objective optimization problem. The suggested solutions are a whole year's track inspection activities, in which case financial, logistic and human resources as well as track possession can be managed more effectively, when compared to an incremental approach proposed which is proposed by Peng et al. (2013). Nevertheless, both objectives are hardly to solve either separately or simultaneously in the presence of nine technical constraints. Among all the constraints, a time interval between two consecutive inspections on a same track has been determined by an industry as a critical element in track inspection.

An introduction of this constraint can be viewed as an achievement of past experiment-based studies on railway asset management. For example, Lam and Banjevic (2015) imposed an inspection interval in an intelligent asset health monitoring system. This system alerts an asset manager with an optimal situation to conduct asset inspection before proper maintenance jobs are assigned. A decision is made based on the level of risk to failure which uses information about the hazard of asset as an input for the system. Kim and Frangopol (2011) conducted research with a similar purpose but they used a probabilistic approach to a fatigue-sensitive structure. The statistical-based model generates an inspection schedule that requires a low inspection cost but is able to guarantee inspection quality, at least at an acceptable level. In their proposed model, the cost is calculated based on costs of inspection and expected cost of failure. Benefits of the proposed model are evident not only on the inspection section but they also extend to monitoring scheduling. A similar concept can be found in Kashima (2004), in which a condition-based inspection regime was proposed which in turn means that an optimal inspection time interval is determined quantitatively using a structural reliability theory. A series of life-cost analyses shows the effectiveness of the proposed method. This work contributes significantly to the development of track inspection frequencies guideline, while it was later used as problem constraints in Konur et al. (2014).

Reliability techniques were also applied in large-scale railway network systems, as presented in Carretero et al. (2003). Generally, the reliability centred maintenance techniques offer ground benefits, such as technical insight into planning of preventive maintenance (PM), which allows various levels of adjustments in selected maintenance processes, and clear decision diagrams. The authors demonstrated a wide range of specific benefits, such as reduction in time taken for information extraction, an increase in equipment life that positively affects corrective maintenance costs, and an overall improvement in company productivity. Lin et al. (2015) proposed a system reliability-based methodology to construct a non-periodic PM schedule for deteriorating complex repairable systems. The methodology makes an estimate of system reliability as the condition variable functions differently depending on the current scenario in the system. In each scenario, an optimal PM schedule is obtained by solving constrained minimisation problem, which incorporates properties of a specific reliability-based PM model.

A model formulation of the TIS problem by Higgins et al. (1999) where it defines an inspection cost as a problem constraint is suitable for an organisation that has a limited budget for inspections. This work also put forward job sequence, track authorisation, and travel time as constraints that need to be satisfied when solving the optimisation problem. Two objectives are involved: minimisation of disruptions to train services and completion time. The former objective was introduced due to the fact that trains must follow speed restrictions when approaching the inspected area, and to this extent it might cause delays. Too many delays could create a bad perception from the public which is certainly not welcome in a passenger transportation business (Balcombe *et al.*, 2004). Budai et al. (2006) extended the work by introducing generalised costs of track possession as an objective to be minimised. This study is unique as it generates an optimal schedule which involves both preventive and routine maintenance works.

In work by Konur et al. (2014), inspection results are assigned as an input to a risk of failure analysis of tracks. In the same line, reliability and a crack growth approach have been studied as a case study to effectively export track inspection results to a rail-related failure risk measurement analysis. Their primary concern is to optimally utilize track inspection data in the context of track inspection but is not intended to be a primary source of data. The risk of failure is controlled by introducing two penalty cost functions for exceeding maintenance thresholds into the total cost of TIS model (Soleimanmeigouni *et al.*, 2016). With these functions, a different inspection policy, i.e. interval could turn out non-optimal due to the function changes. The findings point out that the effect of changes in model inputs on total cost formulation could generate a different inspection strategy. Meanwhile, Khouy et al. (2016) propose a risk of accident cost function which is derived from the cost of derailment and the probability of safety fault occurrence that can cause derailment in the interval between maintenance execution and the next inspection. However, the use of the proposed risk function is limited under certain assumptions, namely tracks are identical regardless of geometric characteristics, location (curve or tangent), substructure characteristics and construction time, and maintenance history. Further sensitivity analysis is strongly suggested to justify the claim that tracks with higher degradation rates require more frequent inspections and PM.

It is not an exaggeration to say that both inspection and measurement vehicles are a great creation for track inspection and maintenance. A train borne with plain line pattern recognition technology, for example, not only increases inspection integrity but also reduces inspection times as compared to a foot patrol (Network Rail, 2014). However, it is crucial to assign those vehicles on tracks at low expenditures without comprising the railway safety principles. To achieve both objectives, Podofillini et al. (2006) developed a risk-informed methodology to determine the optimal strategies for assigning ultrasonic inspection vehicles. Realising the restrictions that underlie the inspection and maintenance procedure in the real world, the study developed a model to verify the workability of the proposed solutions. In addition, no technical constraints have been presented in the problem formulation, unlike in Peng et al. (2013). Periodicity constraints, penalty costs imposed due to unfinished inspections within the allocated time windows and avoiding task completion by an unauthorised inspection team were taken into account with regard to an optimisation problem for an inspection vehicle. By taking into account the complexity of the abovementioned realistic issues, the single objective problem was formulated as a vehicle routing problem (VRP).

VRP is a popular methodology to serve a known number of orders/clients on the given network with a fleet of vehicles of minimum cost while satisfying side constraints such as time windows . A solution of the proposed model was found to be superior than the one produced through a manual procedure when it was tested for a short-term schedule, i.e. a partial complete schedule. Meanwhile, Lannez et al. (2015) also proposed a single objective VRP but a solution of the problem has a minimum total deadhead distance while satisfying twelve constraints, where at least two constraints are related to vehicle limitations.

An attempt to move away from a periodical practice in managing railway assets can also be observed from the way an optimisation problem is formulated. Ottomanelli et al. (2002) developed a fuzzy-logic-based decision-making to facilitate rail tracks maintenance which provides a track supervisor more flexibility in terms of deciding which tracks should be accessed and when. With the proposed model, there are no more crisp and rigid decisions, and the system will generate a membership value to six maintenance modules, which includes delay of the maintenance as one of them. A maintenance module with a higher score is more trust-worthy but an action to be taken is not limited to the highest one. This study also offers a new scheme of track categorisation which aims to study tracks behaviour towards the end of their respective service time, with respect to loading effects. On the other hand, Peng et al. (2013) proposed an incremental approach to generate non-fixed inspection interval solution. It is a good solution approach, as maintenance expenditures can be reduced, which in turn helps railway businesses to be more competitive, especially within the public transport sector. Although the approach considers real-time situations, including current condition of equipment when deciding the next inspection task, it is obviously exhaustive and difficult to implement. One can think about computer simulation every time for track selection but managerial issues such as crew planning, logistic decision and track possession needs to be repetitively managed. This situation itself might spark disturbances in a railway maintenance system. In work presented in Yan et al. (2015), the whole degradation process will be divided into a normal and warning section. A warning threshold divides the sections in which the latter section has typically shorter (time) interval. Inspections will be more frequent in the warning section. Theoretically, a non-periodic inspection policy not only improves inspection efficiency but can also avoid unnecessary inspection/maintenance costs. However, an implementation of a non-periodic inspection policy for railway track inspection is very challenging due to periodicity of train timetables, prioritisation on track access given to freight companies, and of course, resource constraints.

2.2.2 Uncertainties

A solution for TIS is obtained by solving a constrained optimisation problem that postulates the inspection vehicle, equipment, crew, and track as variables. Each decision variable is, in general, non-deterministic and has a different number of parameters. However, several input variables would be certain; for example, the financial budget for the track inspection is likely secured until the end of the schedule timeline, irrespective of the other variables. While the method of solution for TIS is differ (Osman, Kaewunruen and Jack, 2017), the basic relationship among input variables in TIS can be viewed from an influence diagram. An influence diagram is not a flow chart but is a simple way to understand the relationship among input uncertainties, structure and decision values (Renooij and Van der Gaag, 1998). Fig. 2.1 shows an influence diagram associated with the TIS problem where the oval-shaped block represents (potential) uncertainties in the TIS model.

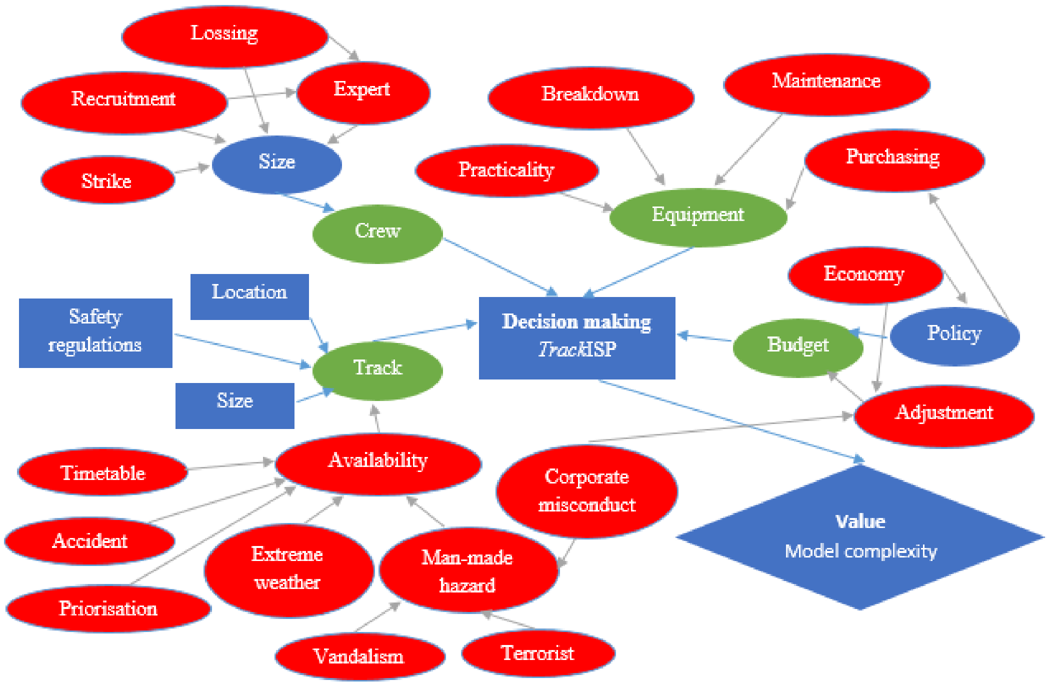


Fig. 2.1 An influence diagram corresponds to track inspection schedules

According to Zhang and Mahadevan (2000), different types of uncertainties arise in modelling non-destructive inspections as optimisation problems, which in principle demands parallel and/or complex hierarchical analysis due to several constraints, such as lack of quality data, model characteristics and design, knowledge–interpretation clash and inadequate information. This situation could relate to the positive relationship between modelling uncertainty and model complexity (Snowling, 2000). In any cases of TIS, a complexity of the associated optimisation model can be calculated using the following equation:

$$I_C = \sum_{j=1}^{N} \sum_{i=1}^{n_j} p_i r_i \tag{2.1}$$

where N, n_j, p_i and r_i is number of state variables, number of processes flowing to or from state variable *j*, number of parameters used to describe process *i* and number of mathematical operations used to describe process i, respectively. It is clear that the complexity of the model increases as the number of input parameters increases. Hence, complex models are generally robust compared with simpler models. In addition, the probability calculation of uncertainty (refer to the arc in an influence diagram) also leads to increasing model complexity. In a case where probabilities of some events are known to be low, it is decided not to include them in the TIS model. When TIS problem is solved under batch environment the model complexity issue is dealt before a method of solution (e.g. genetic algorithm) operates a solution search. Comprehensive discussion with regard to handling uncertainty can be found in Brugnach et al. (2008).

2.2.3 Disruptive events in rail track inspection

Track inspection is scheduled with respect to RIM's objectives e.g. track safety and costs, and it takes place at the tactical level of railway management. A desktop review on scheduling for track inspection (as discussed in the Section 2.2.1) has shown that TIS is conveniently formulated and solved under an optimisation model. Taking practicality and maintainability concerns into account some input uncertainties (red oval in Fig. 2.1) have not been integrated within existing optimisation models.

Some might argue about the relevance of such uncertainties—for example, a worker strike—to TIS. It must be considered, however, that some cracks in rail that are not detected by an ultrasonic inspection vehicle might then be detected through manual visual inspection. In the case of a worker strike, those cracks would remain undetected, leading to aggressive impact loading and subsequent track damage, which could then cause a train accident (Liu *et al.*, 2014). Imagine that a freight train carrying hazardous materials or petroleum products gets involved in an accident near a plantation area. Fire is easily triggered from the accident and turns the area into ashes. Losses occur not only on the rail management side but also to third parties outside the railway domain.

Terrorist attacks, which have not been a threat to railway operations until 20 years ago, could prevent some tracks from receiving inspections. During the closing time, unaffected tracks will be increasingly used to minimize losses in business, and consequently, those tracks require additional inspection and maintenance too. Machine breakdown and special inspection requests potentially present disruption to inspection schedule execution in various ways.

Climate-related extreme weather events are also in a potential list of disruption (Kaewunruen, Sussman and Matsumoto, 2016). For example, the Malaysian east coast train line closure in late 2014 is a notable illustration of this issue. Natural disaster from heavy floods not only disrupted the passenger timetables but also the scheduled inspection and maintenance activities. Consequently, the event led to a significant decline in the RIM profits due to operation closure in the east-coast region and large spending on infrastructure repair and/or renewal work (Davies, 2015).

Railway capacity utilization function takes number of trains that running on the infrastructures in the unit of time as a basis to determine a balance point between access charges to the rail infrastructure and short run marginal costs (Khadem Sameni, 2012). However, uncertainty demand forecasting in passenger and freight trains may temporarily affect the existing capacity value which leads railway infrastructure company (RIC) to reduce operational costs in order to protect a profit margin. The cost reduction is feasible to achieve through maintaining tracks with lesser standards. This approach is legal as long as RIC gets an approval from railway regulatory body to downgrade the existing track classification (for a short period) e.g. from Class 4 (60 mph for freight, 80 mph for passenger) to Class 2 track (25 mph for freight, 30 mph for passenger). Re-classified tracks with lower class will permit RIC to assign a less stringent inspection regime. This decision stretches a time interval for an affected track to receive its first (and subsequent) inspection maybe by few months due to longer time taken for accumulated traffic exposure to reach an inspection requirement.

While TIS is prepared under uncertainty for real world setting, a disruption links to these uncertainties may occur in reality. Indeed, non-availability of inspection crews due to strike, time delays for vehicle or equipment maintenance, or ad-hoc inspection requests have a low likelihood of occurring in reality but it could occur suddenly and without warning during scheduled operations. Disruptions that may occur during the schedule operation could affect the schedule performance negatively i.e. deteriorating objective measures. The impact of disruption might propagate to other states (cascading effect). For example, a worker strike is not a monthly event but if it occurs it can delay some inspection activities for a couple of days. As stated by Cacchiani et al. (2014), three days of operations closure is enough to trigger significant social and economic impacts.

In this sense, the scenario has a distinct similarity to the 9/11 attacks, the tsunami in Japan, and the Manhattan bridge collapse, in that the probability of occurrence is relatively low in the context. Such situations are easily understood, as nothing in the past could have convincingly pointed to the possibility of them occurring. Indeed, some organizations use the "acts of God" defence as a reason not to act to avoid losses caused by unexpected events. Importantly, the impacts of such events are further exacerbated by our tendency "to act as if they do not exist." Described as high impact, low probability (HILP), such events pose multi-level risks, not only economic but also societal and political (Beddington, 2012). HILP events unpredictably occur during normal operation of systems. Their presence manifests abruptly and mostly beyond the control of the system owner. This characteristic might indicate that risks associated with some HILP events can only be responded by addressing the consequences rather than reducing the probability of occurrence.

2.3 Low probability risks/events

Identification and analysis of detected risks are crucial for risk-informed decision-makers to establish robust decision-aiding models. Regardless of the type of risks, each risk element, probability and consequence have a clearly defined role in a decision model. More specifically, concerning risks of low probability event, the decision model should accommodate a decision-maker's attitude towards the risk in the process of evaluating consequence (Camerer and Kunreuther, 1989). As stated by Cha and Ellingwood (2013a), the authors of (Von Neumann and Morgenstern, 1944; Kahneman and Tversky, 1979) provided evidences from cognitive psychology and behavioural science that the attitude of individuals, groups or organisations towards risk does have a significant influence on how important elements of risk management are dealt with; such as risk preference, evaluation and strategy. According to Cha and Ellingwood (2014), structural configuration (including building type and height), loss characteristics (magnitude of possible economic losses or extent of casualties), the decision-maker's role as a public or private entity, resources available to the decision-maker and societal impact (whether direct or indirect) are major factors which can affect decision makers' risk attitudes in structural engineering decisions. Not only is the probability of occurrence of a hazardous event and its potential consequences necessary and relevant, but so is the attitude of individuals, groups or organisations towards risk for risk management of low probability event. Hence, understanding the types of attitudes of decision-makers have toward risks is crucial information for a risk management team to design an affordable response strategy against low-probability risk.

In general, attitudes towards risks can be characterised by the degree of willingness of decision-makers to accept the risk, as follows;

i. Risk averse

Individuals and small groups are often reluctance to accept a risk when the outcome of activities, projects or investments are less uncertain or faced with potential loss or consequence. They are said to be risk averse. The social standing of the decision-maker, nature of the potential risk (e.g. if it involves loss of human life or injury *vs* purely economic loss), magnitude of the risk, and the return period of the event are all influencing factors when assuming a risk (Cha and Ellingwood, 2013b).

ii. Risk-accepting

This category represents the attitude when a decision-maker is willing to acquire investments of negative consequences with lower expected return. According to Cha and Ellingwood (2014), risk accepting is a decision-maker's willingness to accept a risk for various reasons ranging from practical limitations of resources to ignorance. Accepting risk decisions is usually followed by monitoring actions in which an alert will be transmitted if any changes in initial risk classification are detected.

iii. Risk-neutral

A decision maker is different, not falling in the risk averse or risk accepting categories, if an assessment of risk elements is unbiased without any special perceptions or fears; government agencies or corporations with virtually unlimited resources typically adopt this attitude (Sunstein and Zeckhauser, 2011; Cha and Ellingwood, 2013b).

2.3.1 Incorporating risk attitudes to a decision making model

Different types of attitudes toward risks can be quantitatively evaluated by applying consequences-based decision models (e.g. utility theory, expected cost analysis, etc.) to the risk management framework (Cha and Ellingwood, 2012).

Minimum expected cost analysis (MECA) is the most commonly adopted consequence-based decision model. MECA presumes a risk-neutral attitude on the part of the decision-maker, with its underlying assumption that the economic valuation of risk is based on technical considerations which can be monetised, and the monetary evaluation is not distorted (Tversky and Kahneman, 1992; Farrow, 2017). However, not all risks (e.g. reputation damage due to flooded railway tracks in the eyes of green-peace activists) can be entirely monetised (Cha and Ellingwood, 2013a; Farrow, 2017). Compared to a decision model that tends to consider nontechnical factors such as environmental and political impact, and social risk perception, MECA provides a more elegant decision (Kendall, Keoleian and Helfand, 2008).

Utility can be defined as a measure of the desirability of a potential consequence for decision making under risks. Consequences in the presence of uncertainty will be mapped into utility using a utility function $U(X)$. The mapping is the procedure of incorporating subjective factors in risk evaluation and a decision maker's risk attitudes into the decision process. The decision-makers' degree of risk aversion is reflected in the shape of the utility function where it would be convex for risk-averse, linear for risk neutral and concave for risk-accepting; as shown in Fig. 2.2. The figure compares the utility functions embedded in MECA, cumulative prospect theory (CPT) and utility theory (UT). For wealthy organisations with large resources, the utility function is linear over consequences of the risk and of course, in terms of monetary value describing risk-neutral behaviour. The linear function is preferred preferences for MECA. A decision based on CPT uses *S*-shaped function (concave for gains and convex for losses), as illustrated in Fig. 2.2. The utility function for losses is steeper than the utility function for gains, reflecting that more risk-aversion in the domain of losses. With regard to decision making by regulators, a utility function is convex for both losses and gains domains, for the function corresponding to utility theory (to be described subsequently). Importantly, a decision-maker must possess a utility function that postulates axioms on a preference order: completeness, transitivity, continuity, and independence, see (Levin, 1989) for a review.

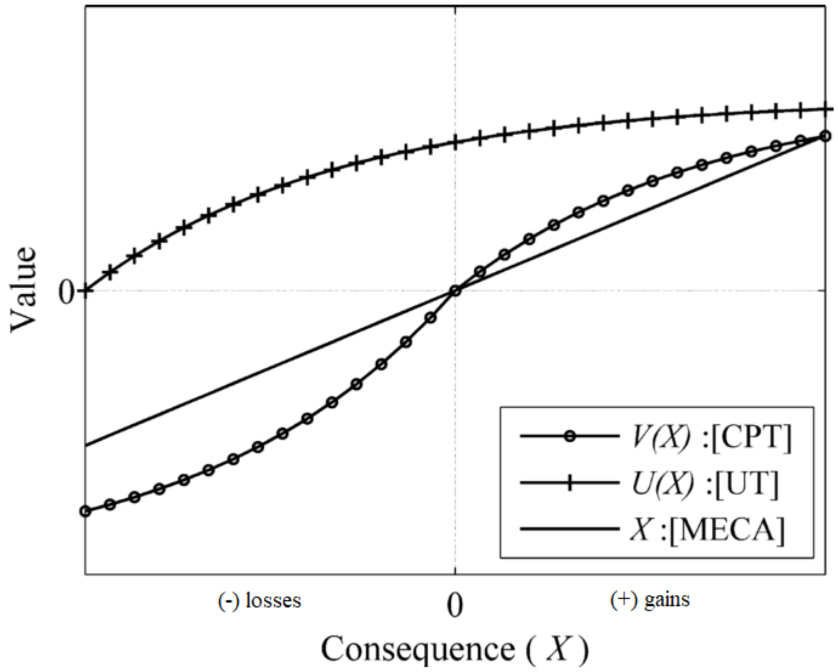


Fig. 2.2 Subjective evaluation of consequence

On the practical level, UT is concerned with people's choices and decisions. The theory is also concerned with people's preferences and with judgments of preferability, worth, value, goodness or any number of similar concepts. In UT, an individual or a decision-maker must often compare expected utility values $E[U]$, i.e. the weighted sums are obtained by adding the utility values of outcomes multiplied by their respective probabilities from potential prospects. Let *x* be an outcome, which can be interpreted as the possible monetary outcome or possible wealth level, and $X = \{x_1, x_2, \ldots, x_n\}$ be the set of possible outcomes. Let *p* be a simple probability measure on *X*, thus $p = (p(x_1), p(x_2), \ldots, p(x_n))$ where $p(x_i)$ is the probability of outcome $x_i \in X\ (i = 1, \ldots, n)$ to occur. Note that there are finite elements $x \in X$ for which $p(x_i) \geq 0$ for all $i = 1, \ldots, n$ and $\sum_{i=1}^{n} p(x_i) = 1$. The axiomatic hypothesis of expected utility is that the decision-maker can make a possibility distribution over possible outcomes of activities (Li *et al.*, 2009). The expected utility over the set of outcomes X can be expressed as follows:

$$U(X) = \sum_{i=1}^{n} p_i u(x_i) \tag{2.2}$$

where $u(\cdot)$ is the utility function. Let $\geq_h$ be a binary relation over U so that $X \geq_h Y \Leftrightarrow U(X) \geq U(Y)$, which means that X is preferred to, or equivalent to, Y if and only if $U(X) \geq U(Y)$. Similarly, $X >_h Y \Leftrightarrow U(X) > U(Y)$ and $X \sim Y \Leftrightarrow U(X) = U(Y)$. Under the assumption that a decision-maker will choose a prospect (e.g. a maintenance policy) at which its expected value of the utility is maximum, UT can be used to determine possible preferences over alternatives. This scene limits the number of alternatives which greatly helps in accelerating the decision-making process.

Applying UT to regulatory decision-making requires an assumption that the regulator has a utility function (or preference) over customer safety or public wealth (Li *et al.*, 2009). This assumption is reasonable because the objective of the regulators is to regulate specific activities or orders on behalf of customers or the public. Notice that this assumption is different from the general assumption in economics that the utility function of an individual is the evaluation of his/her own wealth. In economics, individual decision-makers are assumed to be self-interested.

As stated in Royal's (2017), the UT decision-maker responds to hazards by first forming beliefs about the probability that the hazard will occur and then weighing the expected benefits of reducing hazard exposure against the cost of taking protective measures. UT naturally combines with models of belief formation that fall under headings such as rational expectations, adaptive expectations, or Bayesian updating. These common approaches share the assumption that a decision-maker rationally integrates all available information and acts on beliefs that minimise expected error. The decision involves choosing the alternative with the maximum expected utility.

2.3.2 Tipping point: How much additional investment should be?

Consider a scenario of operational performance-related decision making under risk (e.g. in regard to a response to a disruptive event) where the risk is the possible realisation of unplanned maintenance. Suppose that the hazard (or disruptive events) may lead to a loss of performance $W_o - W_n$ measured by a monetary value), where W_o denotes the steady-state performance and W_n the reduced performance if the hazard has occurred. In order to keep the risk within the acceptable range, an amount of money *C* is going to be invested. The objective of regulatory decision making is to find the optimal amount of investment that maximises the third-party e.g. customer or public, interest.

For a deteriorating system that operates for a certain period, let say, *T* and potentially adds *r* to W_o at the end of the period, the wealth of a system is $U_o = u\left(W_o + r\right)$. When unplanned maintenance is imposed to the system it indicates that benefits of planned maintenance are diminished away. This situation is likely related to an occurrence of sudden shift in a deterioration path. The shift drifts a system's condition reaches a certain condition threshold earlier than estimation. In this setting where the system could facing a loss *L* due to unplanned maintenance with probability p_{um}, the expected utility of U_o at the end of the period will be

$$E\left[U_{end}\right] = p_{um}u\left(W_o + r - L\right) + \left(1 - p_{um}\right)u\left(W_o + r\right). \quad (2.3)$$

In practice, an effort to mitigate this particular hazard can be executed through an inspection performed at a discrete point $k; k = 1, 2, \ldots, n_I$ where n_I is a number of planned inspection. Under rational thinking, a decision maker would place a positive value for an inspection cost, let say, C_I if at the end of the observed period the condition of $U_{+I} = u\left(W_o - C_I\right) \geq E\left[U_{end}\right]$ can be satisfied. However, an inspection cannot guarantee a detection of sudden deterioration but at least it provides engineers with recent measurements (or inspection data) for near-to-real time system's condition evaluation. Thus, it is fair to treat an inspection as a lottery with a random economic consequence δ and is defined in the gain domain as:

$$\delta = \begin{cases} h_1; \text{with probability } \phi \\ h_2; \text{with probability } 1 - \phi \end{cases} \quad (2.4)$$

Thus, updating a formulation of U_{+I} with participating of δ gives the expected value $E\left[U_{+I}\right]$ of U_{+I} at the end of the period

$$E\left[U_{+I}\right] = \phi u\left(W_o + r - C_I + h_1\right) + \left(1 - \phi\right)u\left(W_o + r - C_I + h_2\right). \quad (2.5)$$

Hence,

$$\begin{aligned} E\left[U_{+I}\right] &= \phi u\left(W_o + r - C_I + h_1\right) + \left(1 - \phi\right)u\left(W_o + r - C_I + h_2\right) \geq \\ E\left[U_o\right] &= p_{um}u\left(W_o + r - L\right) + \left(1 - p_{um}\right)u\left(W_o + r\right). \end{aligned} \quad (2.6)$$

An introduction of $W_a = W_o + r$ which (at this moment) takes only a positive and certain value is only for the sake of simplicity of the series expansion. Following the introduction, the Eqn. (2.6) becomes

$$\phi u\left(W_a - C_I + h_1\right) + \left(1-\phi\right) u\left(W_a - C_I + h_2\right) \geq p_{um} u\left(W_a - L\right) + \left(1 - p_{um}\right) u\left(W_a\right) \tag{2.7}$$

Expanding the three terms $u\left(W_a - C_I + h_1\right), u\left(W_a - C_I + h_2\right)$ and $u\left(W_a - L\right)$ using the first three terms of a Taylor series around W_a yields

$$u\left(W_a - C_I + h_1\right) \approx u\left(W_a\right) + \left(-C_I + h_1\right) u'\left(W_a\right) + \frac{\left(-C_I + h_1\right)^2}{2} u''\left(W_a\right) \tag{2.8}$$

$$u\left(W_a - C_I + h_2\right) \approx u\left(W_a\right) + \left(-C_I + h_2\right) u'\left(W_a\right) + \frac{\left(-C_I + h_2\right)^2}{2} u''\left(W_a\right) \tag{2.9}$$

$$u\left(W_a - L\right) \approx u\left(W_a\right) - Lu'\left(W_a\right) + \frac{1}{2} L^2 u''\left(W_a\right) \tag{2.10}$$

Substituting Eqns. (2.8), (2.9) and (2.10).into Eqn. (2.7) turns it into as follows:

$$\begin{aligned} LHS &= \phi u\left(W_a\right) + \phi\left(-C_I + h_1\right) u'\left(W_a\right) + \frac{\phi}{2}\left(-C_I + h_1\right)^2 u''\left(W_a\right) \\ &\quad + \left(1-\phi\right)\left[u\left(W_a\right) + \left(-C_I + h_2\right) u'\left(W_a\right) + \frac{\left(-C_I + h_2\right)^2}{2} u''\left(W_a\right)\right] \\ &= u\left(W_a\right) + u'\left(W_a\right)\left[\phi\left(-C_I + h_1\right) + \left(1-\phi\right)\left(-C_I + h_2\right)\right] \\ &\quad + u''\left(W_a\right)\left[\frac{\phi}{2}\left(-C_I + h_1\right)^2 + \left(1-\phi\right)\frac{\left(-C_I + h_2\right)^2}{2}\right]. \end{aligned} \tag{2.11}$$

Analogously, the following is derived from the right hand side of the inequality in Eqn. (2.7)

$$RHS = u\left(W_a\right) - p_{um} L u'\left(W_a\right) + \frac{p_{um} L^2}{2} u''\left(W_a\right). \tag{2.12}$$

Plugging expressions in Eqns. (2.11) and (2.12) into the original inequality in Eqn. (2.7) gives us

$$
\begin{aligned}
& u(W_a)+u'(W_a)\left[\phi(-C_I+h_1)+(1-\phi)(-C_I+h_2)\right] \\
& +u''(W_a)\left[\frac{\phi}{2}(-C_I+h_1)^2+(1-\phi)\frac{(-C_I+h_2)^2}{2}\right] \\
& \geq u(W_a)-p_{um}u'(W_a)+\frac{p_{um}L^2}{2}u''(W_a)
\end{aligned}
\tag{2.13}
$$

which will be

$$
\begin{aligned}
& \frac{1}{2}u''(W_a)\left[\phi(-C_I+h_1)^2+(1-\phi)(-C_I+h_2)^2-p_{um}L^2\right] \\
& \geq u'(W_a)\left[-\left(\phi(-C_I+h_1)+(1-\phi)(-C_I+h_2)\right)-p_{um}L\right] \\
& = u'(W_a)\left[-\left(\phi h_1+(1-\phi)h_2-C_I\right)-p_{um}L\right] \\
& = u'(W_a)\left[\left(C_I-E[\delta]\right)-p_{um}L\right]
\end{aligned}
\tag{2.14}
$$

Dividing both sides of Eqn. (2.14) by L^2, the inequality becomes

$$
\begin{aligned}
& \frac{1}{2}u''(W_a)\left[\frac{\phi(-C_I+h_1)^2+(1-\phi)(-C_I+h_2)^2}{L^2}-p_{um}\right] \\
& \geq u'(W_a)\left(\frac{1}{L}\right)\left[\left(\frac{C_I-E[\delta]}{L}\right)-p_{um}\right].
\end{aligned}
\tag{2.15}
$$

Consider there is an instrument which mimics the concept of *lottery* with a random economic consequence $\hat{\delta}$ and is defined in the gain domain as:

$$
\hat{\delta}=\begin{cases} (-C_I+h_1)^2 \text{; with probability } \phi \\ (-C_I+h_2)^2 \text{; with probability } 1-\phi \end{cases}
\tag{2.16}
$$

where an expected value of $\hat{\delta}, E\left[\hat{\delta}\right]$ exists in the left hand side of Eqn. (2.15). Thus, the equation is rewritten in simpler form as follows;

$$u''(W_a)\left[\frac{E\left[\hat{\delta}\right]}{L^2}-p_{um}\right]\geq\frac{2}{L}\left[\left(\frac{C_I-E[\delta]}{L}\right)-p_{um}\right]u'(W_a) \tag{2.17}$$

which may be rearranged into the form

$$\frac{u''(W_a)}{u'(W_a)}\leq\frac{2}{L}\frac{\frac{C_I-E[\delta]}{L}-p_{um}}{\frac{E\left[\hat{\delta}\right]}{L^2}-p_{um}} \qquad \text{for}\,\frac{C_I-E[\delta]}{L}<\sqrt{p} \tag{2.18}$$

subject to

i. $u'(W_a)>0$,

ii. $\frac{E\left[\hat{\delta}\right]}{L^2}<\frac{C_I-E[\delta]}{L}$,

iii. $p_{um}<\sqrt{p_{um}}$,

iv. $C_I-E[\delta]<E[L]=p_{um}L$, and

v. $E\left[\hat{\delta}\right]<p_{um}L$.

Using the definition of Arrow-Pratt's risk aversion (Thomas, 2016), the lowest value of risk aversion that a track manager assigns to a profitable track section with an initial monetary value W_o and that will gain a profit of *r* if his company pays for an inspection cost of C_I is

$$Ra_{\min}=\frac{2w_o}{L}\frac{\frac{C_I-E[\delta]}{L}-p_{um}}{E\left[\hat{\delta}\right]-p_{um}}. \tag{2.19}$$

Preferences of such track manages towards the risk of unplanned risk maintenance can be determined from Eqn. (2.19). In a situation where $Ra_{\min}=0$ then a decision maker is risk neutral. For risk averters, they will make sure $Ra > 0$ is true for any selection of an inspection policy. In an opposite direction, we will have a risk-accepting decision maker.

2.4 Resilience

According to Klein et al. (2003) resilience can be traced back to the Latin word "resilio" which means "to jump or bounce back". In the scientific world, resilience is firstly defined as a measure of the persistence of systems and of their ability to absorb change and disturbance while maintaining the same relationships between populations or state variables (Holling, 1973; D'Lima and Medda, 2015). A resilient system is capable of absorbing the magnitude of disturbance before the system's condition moves far from any stable steady-state. A change in state causes instabilities in the system that could redefine its original structure and configuration. In many practical problems such as engineering/safety systems and risk management, various definitions of resilience have been proposed, making resilience more inclusive (Hollnagel, Woods and Leveson, 2006; Henry and Emmanuel Ramirez-Marquez, 2012). Regarding the infrastructure system, Doherty et al. (2012) interprets resilience as the ability of an asset (or more extensively, a system) to remain operational in a range of circumstances/ perturbations. Adaptive resilience emphasises the adaptive capability which refers to behavioural considerations after perturbations through absorbing disturbance before changing state; the ability to adapt in order to avoid potential losses by creating security/safety in an environment and also to rapidly recover from perturbations (Adjetey-Bahun *et al.*, 2016a).

Systems (of interest) that possess good operational resilience capabilities tend to prevent undesired events from causing any disruption incur in the normal system operations (Adjetey-Bahun *et al.*, 2016b). These events have low probability of occurrence and possess either a high or low impact. This is based on the assumptions that the other combinations should have been covered either up-front (high- probability/high-impact) or during normal operations (high-probability/low-impact) (Stolker, Karydas and Rouvroye, 2008). This could be one of the reasons why the resilience concept has been introduced to deal with low-probability events with ambiguous level of impacts (Sheffi, 2005; Stolker, Karydas and Rouvroye, 2008). Moreover, an operationally resilient system (or subsystem) must have the ability to effectively and quickly respond, recover from disruptions and exploit opportunities from unanticipated events effectively (Birkie, 2016).

2.4.1 An illustration of resilience concept

Bruneau et al. (2003) proposed an illustration of a *Resilience Triangle* to illustrate the key features of the definition of resilience, see Fig. 2.3. The y-axis represents the quality of infrastructure $Q(t)$, varies over time. The measure is on the percentage scale ranged from 0-100 where 100% means no degradation in service and 0% means no service is available. A dipping vertical line in the figure indicates a decrease in the quality of a system (e.g. civil infrastructures) when any disaster, in any form strikes. A resilience action is expected to restore an affected system to normalcy as time passes (shown along horizontal axis). Ultimately, system owners would be interested in minimising economic losses due to the drops in the system quality. Mathematically, this can be achieved by limiting the area associated with resilience triangle; a small value for both the height and depth of the triangle. To reduce the depth, a system needs the ability to resist the direct initial impact (like static resilience) and avoid immediate damages (Bruneau *et al.*, 2003; Pant, Barker and Zobel, 2014). This is termed as robustness and is one of the main characteristics of a resilient system (Bruneau *et al.*, 2003). Meanwhile, the height of the triangle represents time taken for a system to attain a desired state and should be less i.e. $t_o \to 0$ which is more likely to occur at higher recovery speed. The speed or rate of the recovery action relates to the capability of a system to diagnose and prioritise disruptions and initiate recovery actions in order to overcome disruptions by managing all resources (monetary, material, technological, human, etc.). External resources may be mobilised to deal with an extreme disruption which can be one of the elements of resourcefulness. Besides robustness, rapidity and resourcefulness, a resilient system can be further characterised by redundancy. Redundancy in a system sustains functional requirements to withstand disruptions.

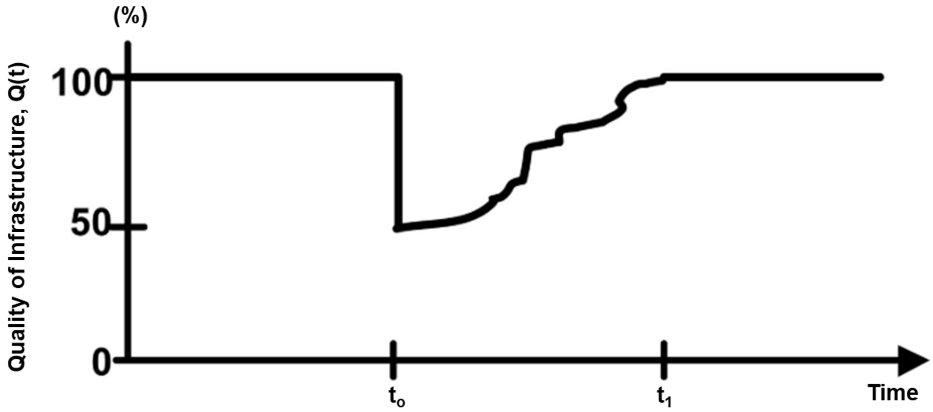


Fig. 2.3 Measure of seismic resilience—conceptual definition (Source: Bruneau et al. (2003)).

Henry and Ramirez-Marquez (2012) described resilience as the ability to restore a system, *S* from the disrupted state S_d to a stable recovered state S_f, see Fig. 2.4. When a disruptive event occurs (due to internal and/or external factors) in the system, the system's quality will drop and enter the S_d. Resilience of a system at time *t* is exhibited if and only if there is a resilience action taking place at time t_s. A resilience action is expected to mend the affected system to S_f with delivery function $F(t_f)$ at time t_f. In such situations, the S_f may not be the same as S_o, since the new state may reach an alternative (lower, or perhaps higher) equilibrium level (e.g., for economic systems (Pant *et al.*, 2014)).

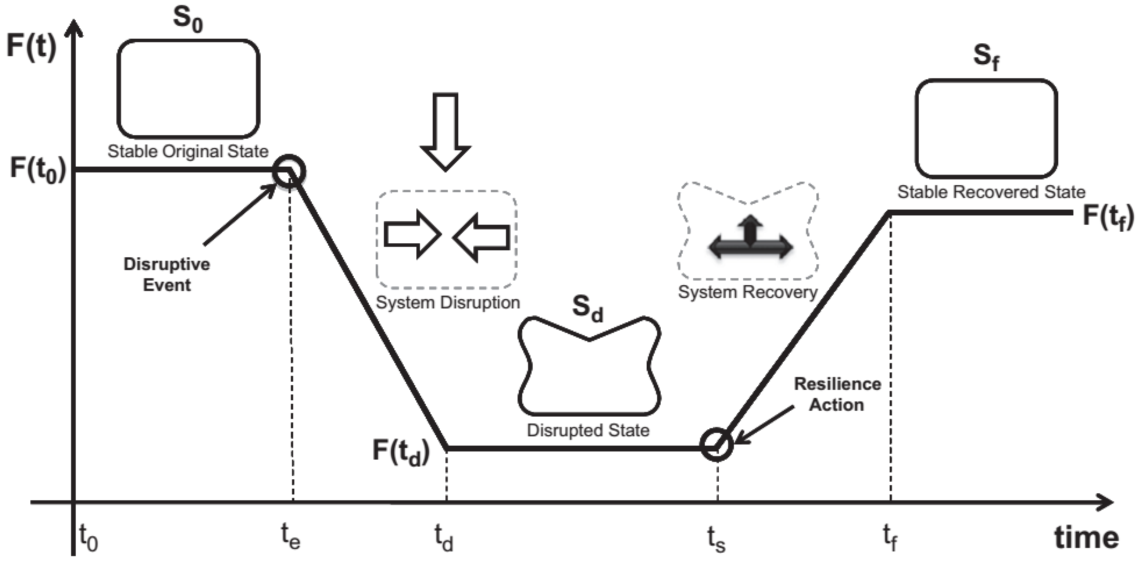


Fig. 2.4 Delivery function transition in resilience (Source: Henry and Ramirez-Marquez (2012))

Nan and Sansavini (2017) improvised the illustration of transitions in system resilience as depicted in Fig. 2.4 with a description of system capability (susceptibility, absorptive, adaptive and restorative and recovery) at different system phases. As illustrated in Fig. 2.5, in some cases the negative impacts of disruptive events are not necessarily observed at the time instant it incurs at t_o i.e. $t_d < t_o$. This can be observed in a system with susceptible capability; the inability of a system to avoid being hit by a threat mechanism. The selection of performance metric and type of disruptive event play an important role in improving this system capability. A system in the disruptive phase $t_d < t_r$ since the system performance begins to drop until it reaches the lowest level at time t_r. However, the initial magnitude of performance loss can be avoided through its absorptive capability. Vugrin et al. (2011) defined absorptive capability as the degree to which a system can absorb the impacts of system perturbations and minimise consequences with little effort. This capability can be enhanced by improving system redundancy, which provides an alternative way for the system to operate. Hosseini (2016) highlighted that any activity that must be taken to absorb shocks of disruptions should be identified in advance.

During the recovery phase $\left(t_r \le t < t_f\right)$, the system can apply its adaptive capability to adapt itself and attempt to overcome a disruption without any recovery activity but incurring some extra effort and cost (Vugrin, Warren and Ehlen, 2011). Adaptive capacity is distinguished from absorptive capacity in that adaptive systems change through self- organisation in response to adverse impacts, especially if absorptive capacity has been exceeded. A system's adaptive capacity can be enhanced with emergency systems (Francis and Bekera, 2014). Adaptive capacity is a temporary feature (i.e. a system continues operating through a nonstandard manner at non economical operating points). Restorative capacity refers to the ability of a system to permanently repair or restore damages from a disruption (Vugrin, Warren and Ehlen, 2011). According to Francis and Bekera (2014), restorative capability of a resilient system is often characterised by the rapid return to normal or improved operations and system reliability. This capacity should be assessed against a defined set of requirements derived from a desirable level of service or control. Note that the effects of adaptive and restorative capacities overlap.

Upon completion of the resilience action, the system performance reaches and maintains a new steady level for $t > t_f$. The action provides possibility of having a newly attained steady state for recovered system at a lower or higher level than the original one (Nan and Sansavini, 2017).

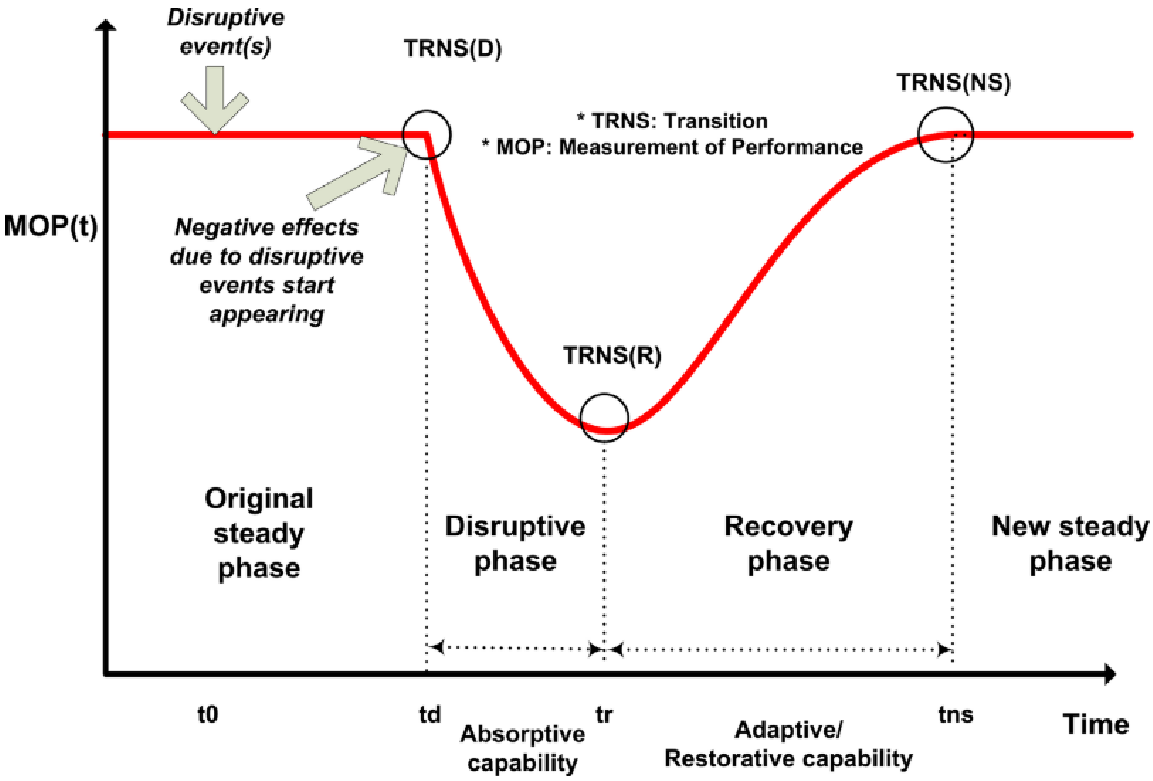


Fig. 2.5 System resilience transitions and phases (Source: Nan and Sansavini, (2017))

2.4.2 Generic resilience metric

Such resilience metric would greatly enable development of resilient systems, allow comparisons of resilience strategies and support resilience related decisions during the design and operation (Henry and Emmanuel Ramirez-Marquez, 2012). Several proposals have been introduced for generic resilience matric to address the issue of resilience quantification across any application domain. Following sub-section provides a review on the generic resilience metric where focus is on deteriorating systems, safety as well as planning and scheduling.

Regardless of the system structure, a system owner initially uses generic resilience metric to quantitatively assess a system's resilience by measuring performance of the system. The resilience metric is universal which requires general information about the system's structure to compare the performance of the system before and after disruption. Although the generic resilience metric does not require details on the general structure of a system, a quantitative examination compels attentive thinking about behaviour of the system to deal with disruptions (Hosseini, 2016). Fundamentally, a system can be classified either as resilient or still requiring improvements if it lacks resilience.

Early research work on the development of the resilience metric adopted the concept of *Resilience Triangular*, as illustrated in Fig. 2.3 (Bruneau *et al.*, 2003). Three resilience elements that can be extracted from the triangle are i) loss of functionality from disruption or disturbance, ii) pattern of restoration, and iii) recovery time. These elements can further be used to measure the functionality of a system after a disruption and also resilience loss, *RL*. In the scene of earthquake, *RL* with respect to that specific earthquake, can be measured by the size of the expected degradation in quality (probability of failure), over time (that is, time to recovery) (D'Lima and Medda, 2015). Mathematically, it is defined by

$$RL = \int_{t_o}^{t_1} \left(100 - Q\left(t\right)\right) dt \tag{2.20}$$

In this sense, a resilient system's performance can be improved through a reduction in the triangle area. It can either be static (no change in an initial recovery length) or dynamic (shift the t_f closer to t_d) direction. The 'resilience loss' metric in Eqn. (2.20) is however not suitable for a system that has not experienced a dramatic drop in its performance upon occurrence of disruption. Moreover, the metric only focuses on a single resilience capability which may be inadequate for systems which other phases and transitions of resilience are apparent in them. Perhaps resilience metrics concerned with changes in system's performance at every transition points (disruptive events, recovery and restoration) are needed. Resilience metric is thus defined as the time dependent function in the context of systems (Henry and Emmanuel Ramirez-Marquez, 2012).

According to Henry and Emmanuel Ramirez-Marquez (2012), time-dependent resilience metric $\mathfrak{R}(t)$ is formulated as a ratio of recovery at time t to loss suffered by the system at some previous point in time t_d i.e. $\mathfrak{R}(t) = \text{Recovery}(t)/\text{Lost}(t)$. The system is said to be fully resilient if recovery is equal to the loss i.e. $\mathfrak{R}(t) = 1$. In contrast, a system exhibits no resilience if there is no recovery $\left(\text{Recovery}(t) = 0\right)$. The following is the basic guidelines that define the parameters involved in both the numerator and denominator of $\mathfrak{R}$ (Henry and Emmanuel Ramirez-Marquez, 2012):

1. Define a system under studied:

The transition path shape encodes key states of a resilient system. The path begins at a point where a disruptive event causes a system under studied to leave its original state ξ_o and enters the disrupted state ξ_d. If a resilience action has been taken and is effective, the system will recover from the disrupted state to the recovered state ξ_f; this is where the path ends.

2. Figure-of-merit (Resilience metric)

Developing a quantifiable and time-dependent delivery function or figure-of-merit $F(\cdot)$ for the computation of resilience is necessary. Depending on the nature of system consideration, the function mainly represents the performance of the system such as availability, safety or train delay minutes. Fig. 2.4 illustrates the overall figure-of-merit over time during resilience. Any undesired changes in the value $F(\cdot)$ indicates the system is likely losing its performance/capability caused by a disruptive event. In some cases, $\xi_d \to 0$ i.e. the system disruption stops at time when the resilience action is initiated.

3. Disruptive event

An event e_j is called a disruptive event for *S* if it affects *S* such that the value of $F(\cdot)$ is less than $F(t_o)$ after time t_o. Only then a system in state ξ_d. The value of $F(t_o)$ decreases faster when under attack by a set of disruptive event defined as $\mathfrak{D} = \left\{ e_j \in \Upsilon \middle| F\left(t_d \middle| e_j\right) \right\}$ where $\Upsilon = \{e_1, e_2, \ldots, e_m\}$ represent the set of all disruptive events.

4. Resilience action

A disrupted system can move out from a disrupted state ξ_d if a resilience action is able to increase the value of $F(\cdot)$ from $F(t_d)$ to $F(t_f)$ between times t_s and t_f as illustrated in Fig. 2.4. For that purpose, both recovery policy and the overall resilience strategy must be defined in advance. The recovery policy (or mechanism) focuses on how the damaged part(s) of the system of interest could be replaced or could be repaired; with or without removal from the system (Henry and Emmanuel Ramirez-Marquez, 2012). The overall resilience strategy is more on the implementation of the recovery policy at the system level. Resources, operating constraints as well as geographical location will be determined with respect to economic decisions.

Based on the descriptions of (1-4), the value of resilience $\mathfrak{R}_F\left(t_r \middle| e_j\right)$ can be formulated in the following form:

$$\mathfrak{R}_F\left(t_r|e_j\right)=\frac{F\left(t_r|e_j\right)-F\left(t_d|e_j\right)}{F\left(t_o\right)-F\left(t_d|e_j\right)};\forall e_j\in\mathfrak{D} \tag{2.21}$$

where $F\left(t_r|e_j\right)$ is a specific figure-of-merit evaluated at time $t_r, t_r \in \left(t_d, t_f\right)$. Here, it is necessary to select a quantifiable figure-of-merit. Actually, $\mathfrak{R}_F\left(t_r|e_j\right)$ indicate the impacts of a resilience action on the recovery of the disrupted system. In the situation where the system is unable to "bounce back" to pre-disruption state i.e. $F\left(t_r|e_j\right)=F\left(t_d|e_j\right)$ the value of $\mathfrak{R}_F\left(t_r|e_j\right)$ will equal zero. This could imply that any resilience action that has been taken is totally ineffective. The value of $\mathfrak{R}_F\left(t_r|e_j\right)$ equals one whenever the resilience action causes the system bounce back to the original stable state S_o from its disrupted state. As mentioned earlier, resilience actions are sufficient to recover a system from its disrupted state. Similarly, the value of $\mathfrak{R}_F\left(t_r|e_j\right)$ is greater than one for a system in a better state than a desirable state after recovery has been completed. Importantly, the formula in Eqn. (2.22) is undefined if a system does not suffer any loss i.e. $F\left(t_d|e_j\right)=F\left(t_o\right)$ under disruptive event e_j which implies that no resilience action is needed.

2.4.3 Time and cost of resilience

Any system S is naturally decomposed into components $\left\{s_1, s_2, \ldots, s_{n_S}\right\}$ each with specific characteristics including its relationship with $F\left(\cdot\right)$. When one or more system components are disrupted, it would cause $F\left(\cdot\right)$ to drop from $F\left(t_o\right)$ to $F\left(t_d|e_j\right)$. Therefore, an application of resilience action serves to restore the disrupted components in which $F\left(\cdot\right)>F\left(t_d|e_j\right)$.

Assuming each disrupted component is restored sequentially, the corresponding time of resilience T_{S_j} under disruptive event e_j, $e_j \in \mathfrak{D}$ may be computed as

$$T_{S_j}\left(e_j\right) = \sum\nolimits_{s_i \in S_j} t\left(s_i\right) \tag{2.22}$$

where $t\left(s_i\right)$ is the time elapsed for a resilience action to recover s_i from its disrupted state $\xi_{s_i,d}$ to $\xi_{s_i,f}$ which implies $F\left(t_f\right) > F\left(t_r \middle| e_j\right)$. S_j will be the set of components that are disrupted due to event e_j.

Similarly, let $C_{S_j}\left(e_j\right)$ be the cost incurred in implementing the resilience action to all disrupted components, is given as

$$C_{S_j}\left(e_j\right) = \sum_{s_i \in S_j} c\left(s_i\right) \tag{2.23}$$

where an individual resilience cost is to a specific component.

In an occasion of disruption, instability in the system may cause losses as it operates under a nominal condition for $t_f - t_o$ units of time. This loss could be combination of system's performance loss, financial loss, human health loss, environmental loss, etc. (denoted as C_{LossRs}) as the nature of system under studied (Khan and Haddara, 2003). The total loss due to e_j is hereafter denoted by C_Ψ then the total cost incurred by the system disruption is

$$C_\Psi = C_{S_j} + C_{LossRs} \tag{2.24}$$

Evidently, Eqns. (2.23) and (2.24) would be affected by both the component recovery mechanism and the overall resilience strategy.

2.5 Disruption management

Disruption management (DM) is an art of dealing with disruption situation. Generally, a disrupted situation (often just denoted a disruption) is a state during the execution of the current operation (e.g. the schedule, plan, or distribution trip), in which its impact is large enough to urge the planners to revise the original operation (Jespersen-Groth *et al.*, 2009; Clausen *et al.*, 2010). Some disruptions are known in advance, but human may decide to be passive due to operational complexity and/or limited resources (Chopra and Sodhi, 2014). A fast response within a reasonable period of time could avoid permanent damage to the mission objective.

Based on this definition, disruptions can be viewed as high impact, low probability events. HILP events unpredictably occur during a normal day of operations of such systems. Their presence is abrupt and mostly beyond the control of the system owner. For example, the probability of the 9/11 attacks, the tsunami in Japan and a Manhattan bridge collapse occurrence was low as nothing in the past could convincingly point to this possibility. Unfortunately, their impacts are further exacerbated by our tendency “to act as if they do not exist.” For example, Japan lost about 10 per cent of its capital stock due to nuclear plant closures because of the tsunami. This unique characteristic means that the risks associated with HILP events can best be mitigated by addressing the consequences rather than reducing the probability of occurrences. These measures include investing in contingency management, increasing resilience, improving knowledge, etc. In the context of operational management, those actions are known as ‘responses,’ which aim to restore the system to normality after a disruption has materialised. To effectively recover from a disruption, available resources and plans should be systematically synchronised and have the capability to adapt to dynamic changes in the environment.

In the transportation sector, two applications in disruption management are broadly discussed: vehicle routing and schedule operation.

Vehicles in the transportation system are carefully coordinated as transporting occupies one-third of the logistics costs and also is the key to maximize stakeholder's utility; e.g. manufacturer, logistic provider, customers, etc. (Chopra and Sodhi, 2014). In order to obtain a set of vehicle utilities i.e. goods and distribution routes on minimum cost routes, the problem is solved as a VRP with location, delivery time windows and customer demand as constraints. For a deeper inquiry into VRP, one may refer to studies by Eksioglu, Vural and Reisman (2009). When such disruptions happen such as vehicle breakdowns, traffic congestion due to accidents, delays in receiving supplies at the depot, etc., planners must start searching for solution(s); i.e. vehicles reassignment, within and limited to the relative size of the transport (Wang, Wu and Hu, 2010). In the case of delivery of a single commodity (such as gas containers or oil) that is the same item for all customers, a rerouting order is given manually based on the planner's past experiences or common sense choosing among available vehicles. It is however far more complex to deal with as the best response to the disruption in VRP (designated as a rerouting vehicle problem) will depend on the problem characteristics such as a single or multiple depot(s), type of disruption, time window requirements, etc.

Most work on the rerouting vehicle problem involves road vehicles operated by a single or two person(s) with no passengers, which is relatively less complicated than DM in airline and railway operations, which involve a crew schedule and passenger timetable. In both flight and train services, the effectiveness of the business operation i.e. an optimal revenue point, begins to deteriorate if there is a disruption during the day of operations (Kohl *et al.*, 2007; Nielsen, 2011). Infrastructure malfunctions, mechanical failures, accidents, vehicle (airplane or rolling stock) breakdowns and inclement weather are some of the few events that are capable of causing flight/train delays and/or cancellations, which in turn make the planned timetable, vehicle, and/or crew schedules become unfeasible. For such disrupted situations, planners must take necessary actions to modify the timetable and the resource schedules (Spliet, Gabor and Dekker, 2014). In work by Jespersen-Groth et al. (2009), it is reported that the total direct disruption cost to these two modes of transportation is millions of US dollars every year. In the Netherlands, there is daily frequency of disruptions. Realising that most disruptions are unforeseeable, many studies focusing on reducing the consequences of disruptions have been conducted (Chopra and Sodhi, 2014), and the focus of one study one focus is on the theory of DM.

2.5.1 Managing a disrupted schedule or plan

The main goal of disruption management in a schedule (or plan) is to obtain solutions that minimise deterioration in the objectives of an original schedule due to the changes in the current environment in which the planned schedule is operated. Generally, the solution is not necessarily the lowest cost; recently, costs associated with solution adaptation to disruption have been included as an additional goal of disruption management in scheduling. The time taken to get a schedule back to normal operations was also involved considering disruption is a real-time event. As presented in a recent review paper (Cacchiani *et al.*, 2014), the solutions that will adapt an existing schedule to a modified situation can be found by rescheduling the resources.

2.5.2 Rescheduling

Scheduling in a railway system can be studied at least from two management standpoints: operation and maintenance. The operational management team handles train timetables, rolling stock and crew schedules closely as they have an interdependent relationship, which is vital to the reliability of railway systems (Cacchiani *et al.*, 2014). A train timetable can be seen as a set of passages that contain information such as departure and arrival times, dedicated routes and stations, humans or freight, etc. carried out by trains in the railway system. With increasing capacity demands on passenger and freight services in almost railway networks worldwide, the timetable generation is currently executed using computers programmed with a complex problem solver (Yang, Li and Gao, 2009). Presenting a feasible passenger train timetable is a challenging task due to optimisation factors such as costs and efficiency, while the feasibility of both rolling stock and crew schedules must be taken into account (Kroon *et al.*, 2009). Existence of complex environments in association with train scheduling problems justify why most disruption factors are neglected. On top of that, disruptions are unexpected, have a low probability of occurrence and their consequence level is ambiguous in nature.

Under disruption attack(s), the feasibility of resource schedules as well as timetables will possible experience performance deterioration. Landslides (CNN, 2010), train accidents (The Guardian, 2016) and poor contingency plans (BBC UK, 2015) are real disruptions that have happened in the Europe region recently. As consequences, passenger and freight train services have suffered one or a combination of service delays, and cancelled or rerouted trains. According to Sato and Fukumura (2012), those components have been found as disqualifiers when using trains and this has caused railway companies a loss of millions of pounds in revenue. The situation is expected to become worse as a delay compensation regulation will be implemented (UK Office of Rail and Road, 2014). From a financial aspect and reputations notwithstanding, rescheduling is carried out in order to minimise impacts from disruptions and assure the survival of the initial timetable and resource schedules until the end of their period. Alwadood et al. (2012) points out that impacts of minimisation can be achieved through immediate responses in generating the provisional schedules. In parallel, rescheduling has been widely studied in the area of railway operational management where the existence of a substantial number of survey/review articles, for example (Fang, Yang and Yao, 2015), is proof.

In the context of railway rescheduling, an unexpected event can be classified into two categories: disturbance or disruption, prior to effects of mitigation. According to Cacchiani et al. (2014), a disturbance has a small degree of perturbations of a railway system and mostly affects the timetable only. Adding slack time to the affected timetable and applying small adjustments to resources are some common repair actions that can be used in this situation. In work by Alwadood et al. (2012), disturbance is known as a minor disruption. In a case of disruption where it causes major deviations, rescheduling actions are generally performed in sequential order starting with a timetable followed with resource schedules (Narayanaswami and Rangaraj, 2011). The order is necessary as a provisional timetable applies large changes to an initial timetable (i.e. cancelled or rerouted trains), which causes the existing schedules to be invalidated partially or completely. As explained by Cacchiani et al. (2014), the nature of disruption is a basis of when to decide which part of a timetable requires modification. Normally, the modification is applied according to a scenario-based policy. Following this, a rolling stock schedule receives an adaptation with respect to an updated timetable. Operational constraints involve train capacity, rolling stock circulation, etc. must be satisfied at the same time to conform to the feasibility of the new schedule in real time. If possible, there should be an aim to minimise the size of changes especially related to rolling stock circulation, as it demands adjustments on traffic coordination that have been planned for a year(s) before the operation period (Lidén, 2015). Finally, the affected crews, involving drivers and conductors, would get a new working shift to reflect changes proposed in a timetable. Again, there are operational constraints such as working regulations involved that restrict rescheduling crew problems from getting a straightforward solution. If the results of rescheduling the resource schedule are unsatisfactory then improvements in the proposed provisional timetable are a necessity.

To avoid a recurring situation, an integration (simultaneous) approach has been introduced for reorganising train schedules. The main drawback of rescheduling an integrated operation (e.g. timetable and rolling stock (Veelenturf, Kroon and Mar´oti, 2014), timetable and crew (Walker, Snowdon and Ryan, 2005) and rolling stock and crew (Nielsen, Kroon and Maróti, 2012)) is dealing with a large size and complex problem constraints, which usually requires an advanced heuristic algorithm (Cacchiani, Hemmelmayr and Tricoire, 2014).

2.6 Summary

This research review's purpose was to help the reader understand the importance of managing disruptions in a planned track inspection plan. The phenomenon can be simply interpreted, based on input uncertainties in optimisation models for a track inspection. An influence diagram in Fig. 2.1 depicts the relationship among input uncertainties, structure, and decision values associated with the rail track inspection. Given the strong association between model complexity and uncertainties, especially those with less information, it is better to understand how to manage a disrupted inspection plan. This is significant because the performance of the track inspection plan begins to deteriorate until it reaches the lowest level, where it cannot perform its intended function.

A review of the literature undertaken in the area of disruption management found that most studies have focused on reducing the consequences of disruptions instead of reducing their probability of occurrence. Special attention has been paid to incorporating a decision-maker's attitude in the risks within the process of evaluating consequences. Studies have suggested that the decision-maker's role as a public or private entity, resources available to the decision-maker, and societal impact (whether direct or indirect) are the major factors that can affect decision-makers' risk attitudes in safety-related decisions. The benefits, however, will be specifically related to designing an affordable solution for low-probability risk. Moreover, the literature offers no clear methodology to evaluate the consequences of low-probability events with an ambiguous level of impacts and a lack of flexibility in implementing different model solutions at different phases of resilience management. This may cause the existing methodology to respond inappropriately under the application of the track inspection plan. These limitations justify the development of a risk-resilience response and recovery model for disrupted planned track inspections, further described in Chapters 6 and 7.

CHAPTER 3 TRACK GEOMETRY

3.1 Track geometry parameters

Track geometry defects are characterised by deviations of track parameters from a designated grade and in the range of millimetres can significantly decrease the performance of railway track. In extreme cases, an undetected track geometry defect leads to track-caused train derailments. A risk assessment by Qing, et al., (2014) showed that in 2009 the second cause of freight railroad derailments in the USA was a result of the failure of track geometry (the first one being broken rails). By regularly updating track geometry quality (some authors/organisations use condition instead of quality), track management indicators such as safety, ride quality, operational and maintenance costs and train punctuality can be effectively evaluated resulting in preserving track performances at a desired level. In line with safety regulation, deviations of each track geometrical defects are measured in the loaded condition with a track recording car. Longitudinal level, alignment, track gauge, cant, twist and curvature are typical geometry parameters measured in a railway line during periodic track geometry measurements (BVF 807.2, 2005).

a) Longitudinal level

Longitudinal level is deviation z_p in *z*-direction (see Fig. 3.1) of consecutive running table levels on any rail, expressed as an excursion from the mean vertical position (reference line), covering the wavelength ranges stipulated below and is calculated from successive measurements (European Standards, 2009). These track defects are described by the short-wave longitudinal level of right and left rail and by the mean value of both rails as long-wave longitudinal level (Lander and Petersson, 2012). A rolling movement in the wagon's direction is likely to result if heights deviations are equal for the left and right rails. In contrast, a deviation of longitudinal levelling will generate a rotating movement in the wagon. In short, longitudinal level can be defined as the evenness or uniformity of track in short distances along the top of the rails.

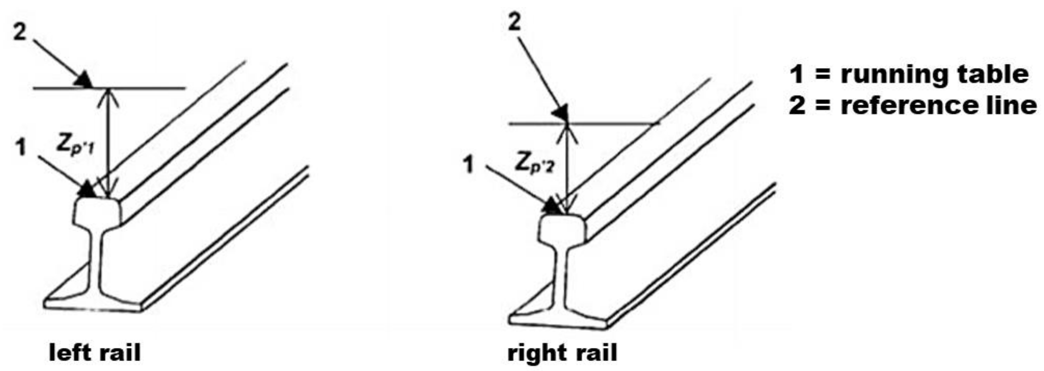


Fig. 3.1 Longitudinal level (EN 13848-1)

b) Alignment

In the standard (European Standards, 2009), alignment is defined as a deviation y_p in *y*-direction of consecutive positions of point *P* on any rail, expressed as an excursion from the mean horizontal position (reference line) covering the wavelength ranges stipulated below and calculated from successive measurements (Fig. 3.2).

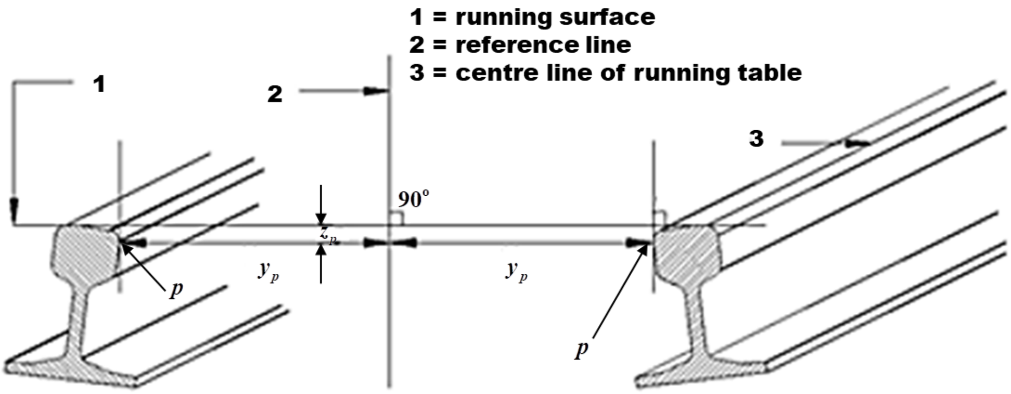


Fig. 3.2 Alignment (EN 13848-1)

c) Cant (cross-level)

Cant is defined in (European Standards, 2009) as the difference in height of the adjacent running tables calculated from the angle between the running surface and a horizontal reference plane. It is expressed as the height of the vertical leg of the right-angled triangle having a hypotenuse that relates to the nominal track gauge plus the width of the rail head rounded to the nearest 10 mm (Fig. 3.3). The hypotenuse for a nominal gauge of 1435 mm is 1500 mm long (European Standards, 2009).

Maximum cant is usually regulated to control the unloading of the high rail wheels at low speeds.

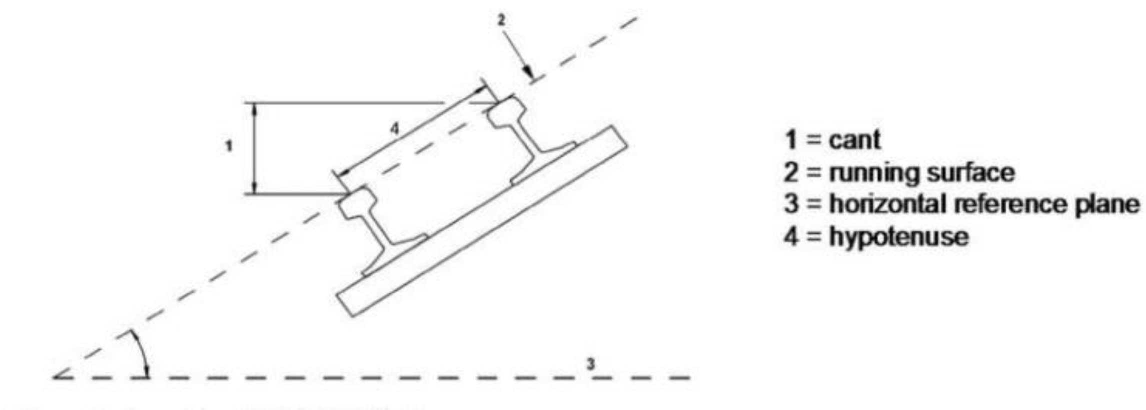


Fig. 3.3 Track cant (EN 13848-1)

d) Twist

Twist is defined as the algebraic difference between two cross levels taken a defined distance apart, usually expressed as a gradient between the two points of measurement. It should be noted that, twist may be expressed as a ratio $(\%\ \text{or}\ \text{mm/m})$ between the differences in cant over a base length (m), e.g. 3 and 6 m (Fig. 3.4). When the twist is higher than the expected value, that is considered as a fault on the railway track (European Standards, 2009). More specifically, twist failure of track was selected based on its criticality with regard to safety (Bergquist and Söderholm, 2015).

Uneven changes in horizontal and vertical position due to uncontrolled poor ballast condition can cause twist. Compared to straight sections, twist irregularities occur more often at curves and transition curves. Therefore it is of great importance to connect the twist irregularities to the track layout e.g. curvature (Lander and Petersson, 2012).

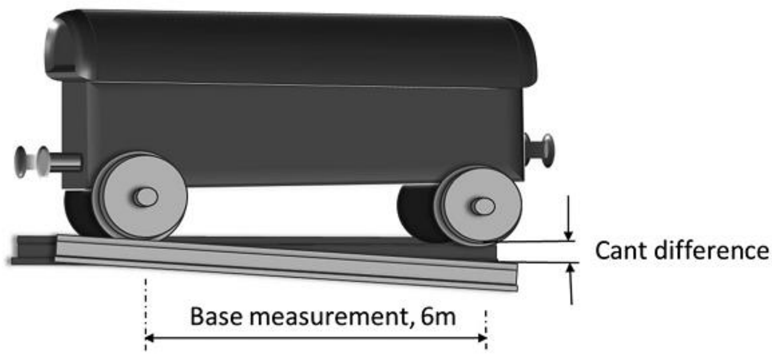


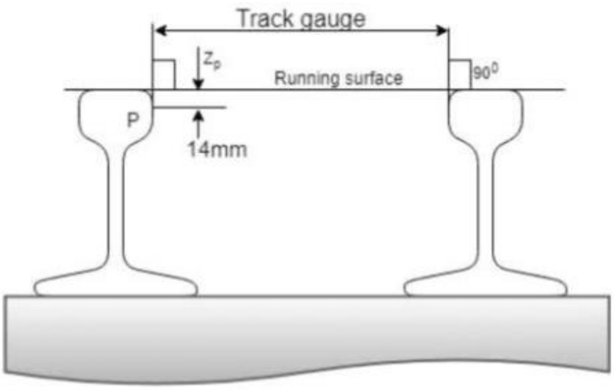
Fig. 3.4 A twist measure (Bergquist and Söderholm, 2015)

e) Gauge

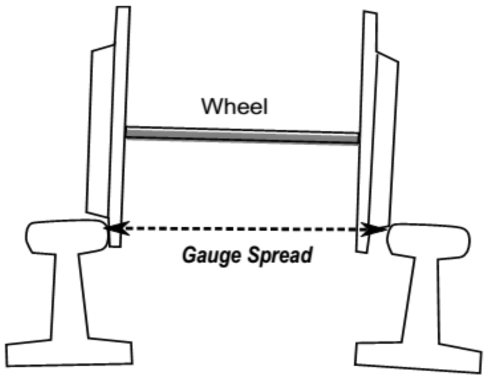
In the standard (European Standards, 2009), track gauge is the smallest distance between lines perpendicular to the running surface intersecting each rail head profile at gauge face point P in a range from 0 to $z_p = 14$ mm below the running surface (Fig. 3.5(a)). The nominal track gauge for a standard track which is used in 60% over worldwide railways is 1435 mm (European Standards, 2009). Track gauge deviation (spread) information is also necessary to appropriately keep a desired state of track safety, since excessive deviations can cause train derailment.

Similar to other geometrical track parameters, gauge deviation can be classified to be static gauge spread and dynamic gauge spread respectively. Measured gauge with the load is defined to be the dynamic track gauge (Fig. 3.5(b)), while the measured gauge without the load is static track gauge. While a loaded measurement of track gauge is available, it will be assessed according to standards to determine an appropriate action (European Standards, 2008). For example, in UK, if the dynamic gauge in the range of 1435-1465 mm; for this low gauge spread, the defect can be controlled by speed restriction or restrain controls including tie-bar and spot re-sleepering. If the dynamic gauge exceeds 1465 mm (medium gauge spread), speed restriction and restrain controls should be carried out to control the defects. If the gauge spread is high (the loaded gauge is over 1480 mm), the line should be closed and the track should be renewed.

Displacement of the rail on the sleeper and rail wear are two common causes of track gauge irregularities (Lander and Petersson, 2012).



(a)



(b)

Fig. 3.5 Track gauge; a) definition according to EN 13848-1, and b) loaded gauge deviations

3.2 An assessment track geometry quality

Following track measurements, measured track geometry will be analysed to determine track irregularities. Lander and Peterson (2012) explained that deviations from the nominal track geometry exceeding tolerated levels define track irregularity (see Fig. 3.6). The nominal or projected track geometry is the designated geometry of the railway track, referring both to horizontal-plane geometry and vertical alignment in the cross-section. For a newly built track, there are minimal track irregularities. Growth in rail traffic however has a great influence on the amount of track irregularities, horizontally due to track movements and vertically due to crushing of the ballast. These irregularities change and evolve to a possibly state where rail tracks cannot be operated in an economically wise manner. Hence, track irregularities are important to maintain regularly.

Depending on the extent under which the track irregularity is, track irregularities can be classified in various groups.

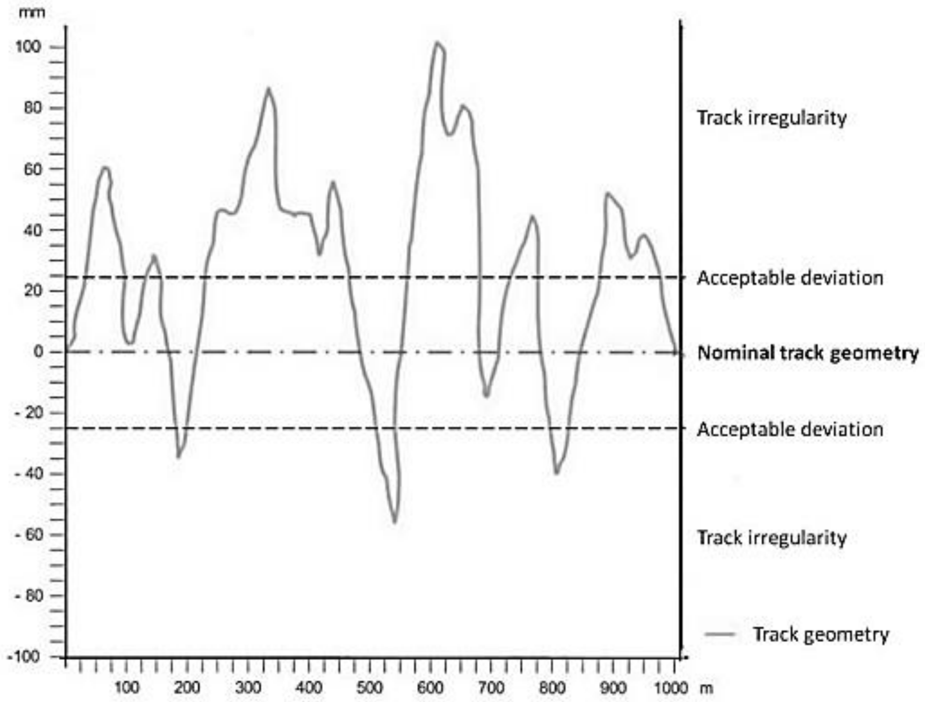


Fig. 3.6 Illustration of track irregularities. The dashed and dotted line indicates the nominal track geometry.

In the standards minimum requirements are given for track geometry for the safe operation of trains based on isolated (geometric) defects. An isolated defect is a single point value where the track deviation norms are exceeded and can concern cant, twist or track gauge (Lander and Petersson, 2012). Isolated defects are shortest irregularities (1-3 m) and can induce dynamic forces on the moving vehicles leading to passenger discomfort. If it is not controlled and corrected in time, they can lead to extreme track wear that can lead to fatigue on the fasteners, sleepers, and rail, which is risky and can cause derailments (Lander and Petersson, 2012). Corrective maintenance is often scheduled to restore track sections with isolated defects.

Short wave irregularities are defined as irregularities with a wavelength of 3-25 m, here alignment and longitudinal level are sorted. Normally, short wavelength irregularity produces short wavelength periodical disturbance that will be hardly noticed and cause the problem to low-speed trains. This is normally indicated by small amplitude and short wavelength. According to Yu, Li, & Wang (2016), short wavelength irregularity can either originate from the rail manufacturing or installation of the sleepers or alignment of sleepers. In most cases when there is sleeper deviation emerging from laying process, there are short-wave irregularities that are associated with the deviation. Despite having similar consequences of isolated defects, short wave irregularities are distinguishable as it creates vibrations of a couple of seconds than inducing dynamic forces (Lander and Petersson, 2012). Nevertheless, isolated defects and short wavelengths are rectified by the tamping process.

Long wave irregularities are defined as irregularities with a wavelength exceeding 25 m which can be difficult to detect with the accelerometers that are used as measuring equipment. Interestingly, long wave irregularities do not pose any safety threats but it would create swaying movements of the wagons resulting in passenger discomfort. No maintenance action is required but long wave irregularities serve to determine of lining is necessary (Esveld, 2001).

For all types of track irregularities, severity of track geometry deviations is evaluated to define a maintenance level (European Standards, 2008):

- Intermediate Action Limit (IAL): this is a safety limit; if the deviation exceeds this limit, there is a risk of derailment. The risk can be reduced by closing the line, reducing the speed or correcting track geometry.
- Intervention Limit (IL): this is a corrective maintenance limit; if the deviation goes beyond this limit, corrective maintenance should be performed so that the immediate action limit will not be reached before the next inspection.
- Alert Limit (AL): this is a preventive maintenance limit; if the deviation exceeds this limit, the track geometry condition should be analysed and included in the regularly planned maintenance operations.

The values of these maintenance levels are given as a function of speed, a significant parameter in the assessment of track geometry quality (European Standards, 2008). For example, the Swedish railway network uses maintenance thresholds in Table 3.1 in the evaluation of short wave irregularities (as well as for segment-based track condition). Briefly stated in Khouy et al. (2014), C-faults identify the limits for the execution of corrective maintenance (IL) whereas B-faults identify the limits for the execution of preventive maintenance (AL). For track segments of 200 m, the UIC ride comfort graph can be used to specify the IL for the longitudinal levels of track segment based on the maximum allowable operating speed (see Fig. 3.7).

Table 3.1 Comparison of the allowable limits between two quality classes; K2 and K3

		B-fault limits	C-fault limit	Comfort limit	
Quality class	Maximum allowable speed (in km/h) for local trains	Alert limit for 25-cm interval (1–25m wavelength) (mm)	Intervention limit for 25-cm interval (1–25m wavelength) (mm)	$\sigma_{H_{\lim}}$ The comfort limit for standard deviation of longitudinal level (mm)	$\sigma_{S_{\lim}}$ The comfort limit for sum of standard deviations of the cant error and the lateral position error of the high rail (mm)
K2	105-120	7	12	1.5	1.9
K3	75-100	10	16	1.9	2.4

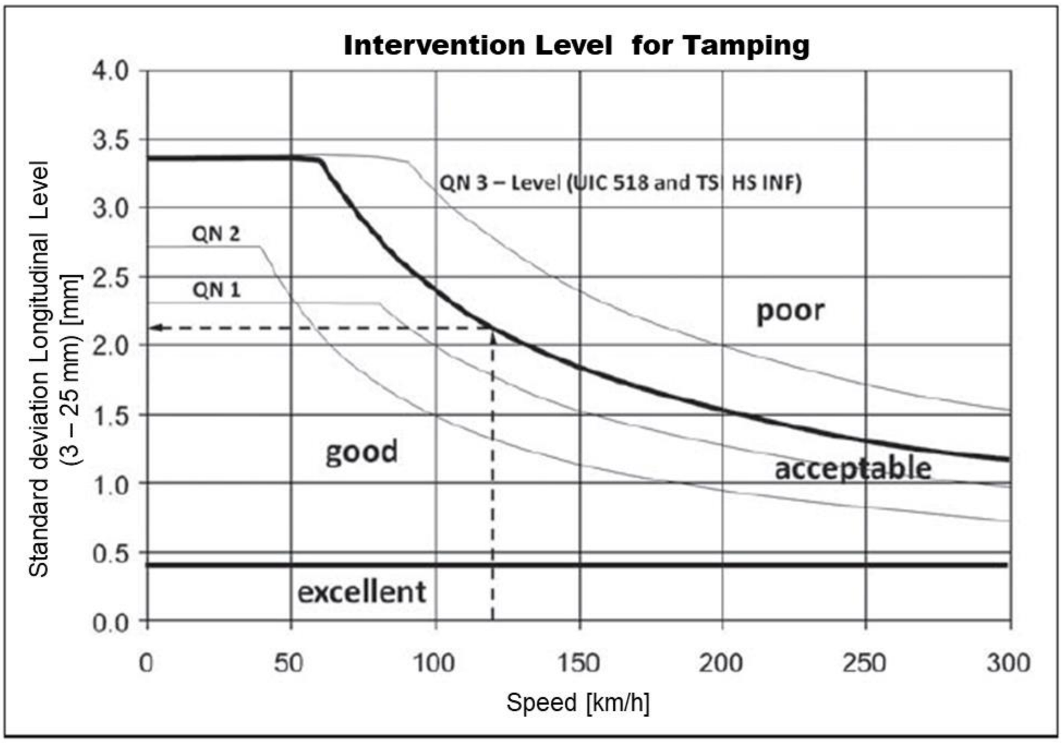


Fig. 3.7 Lines of constant riding comfort at different speeds (International Union of Railways, 2008)

3.2.1 Track quality index

Condition of track geometry over a certain distance is characterised numerically as a track quality index (TQI). The index is typically a standard deviation (SD) of track geometry measurements from the design conditions, which is calculated over defined track length, in Europe mostly 200 m (European Standards, 2008). SD index is calculated from (filtered) measurement values for a given track geometrical parameter over a track segment, as formulated in Eqn. (3.1). The higher the SD index is, the worse the track segment is.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \sum_{i=1}^{n}\frac{x_i}{n}\right)^2} \tag{3.1}$$

where σ_i is the standard deviation of measurements of the *i*-th track geometry parameter in mm, x_i is the (filtered) measurement value in mm of the *i*-th parameter at the *j*-th track point in the track segment, and n_s the number of track points (measuring points) in the track segment.

The direct SD approach (as used by some European, Asian, and Middle East countries） is recommended by the ORE (Office of Research and Experiments) based in France (Office for Research and Experiments, 1981). Differently, RIMs apply the Eqn. (3.1) to different lengths of track segments. For example, the SD of the geometry parameters including profile, alignment, cross level and gauge is calculated for a 1,000 m section using 18.9 m mid-chord offsets. Table 3.2 shows how the SD value can be used to rate a railway track (Office for Research and Experiments, 1981).

Table 3.2 Track condition based on standard deviation

Track condition	SD value
Very good	SD <1
Good	1<SD<2
Average	2<SD<4
Poor	SD>4

A single parameter but more complex TQI was developed in the USA. Amtrak, America in a year of 1998 proposed track roughness index which is an average of squared measurement values for a given track geometrical parameter over a track segment (Ebersöhn and Selig, 1994). Similar to the SD approach, the roughness index is calculated for profile, alignment, cross level, and gauge.

The Federal Railroad Administration (FRA) use the length of space curve generated by track geometry measurement systems on a foot-by-foot basis to produce TQI (FRA, 2005). The length of a space curve is estimated as the sum of straight distances between adjacent data points (FRA, 2005).

$$L_S = \sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2} \tag{3.2}$$

where L_S = traced length of space curve, Δx=sampling space; and Δy = difference of two consecutive measures. The formulation makes it easy to express length-based parameters in different wavelength variations (e.g. 31 ft, 62 ft, or 124 ft) (Lasisi and Attoh-Okine, 2018). A set of TQIs, each for gauge, alignment, surface and cross level, are computed using the following equation:

$$TQIs = \left(\frac{L_S}{L_o} - 1 \right) \times 10^6 \tag{3.3}$$

where L_o is the theoretical length of section. As shown in Fig. 3.8, for a specific track segment length, the rougher the track surface, the longer the space curve will be when stretched to a straight line. The findings in FRA (2005) show that the FRA TQIs were well correlated against the Federal Track Safety Standards.

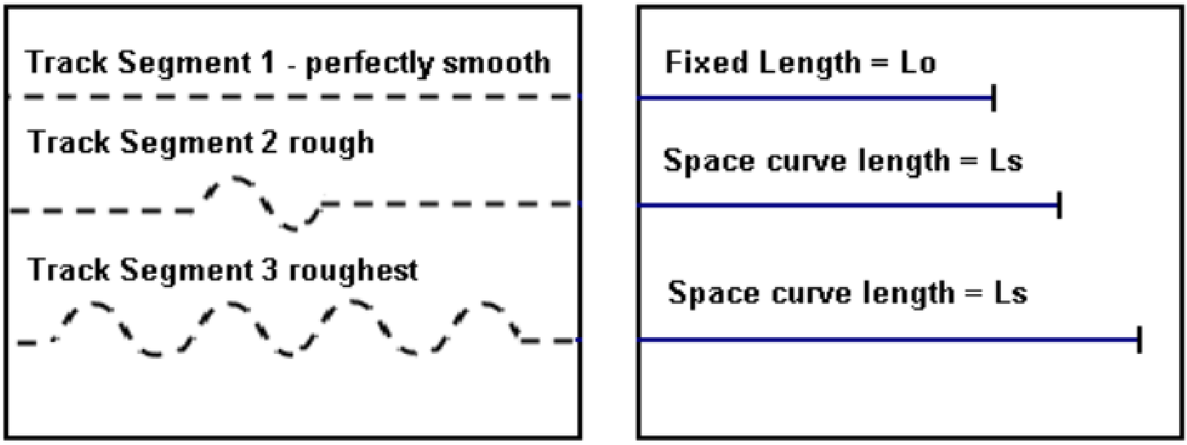


Fig. 3.8 Theoretical definition for fixed length (L_o) and trace length (L_S)

For Swedish railway network, the *Q*-value and *K*-value are two most important TQIs to describe condition of the track. These indices are calculated based of the average longitudinal profiles for the left and right rails σ_H, σ_S and the comfort limits for a given track class. σ_s is the sum of standard deviation of the cant error $cant_{err}$ and the lateral position error of the high rail SH_{err} (refer to Fig. 3.9 and Eqn. (3.4)). In work by Berawi (2013), this parameter is defined as the standard deviation for interaction (calculated as a combined effect from super elevation irregularity and side position of the rail). Both σ_H and σ_s are calculated from short-wavelength signals.

$$\sigma_S = \sigma_{cant_{error}} + \sigma_{SH_{error}} \tag{3.4}$$

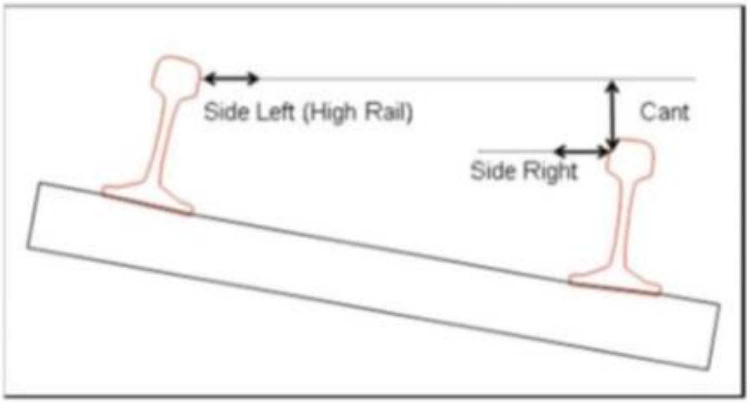


Fig. 3.9 Calculation of σ_S (Source: (Khouy, 2013))

The Q-value, therefore, is calculated per kilometre using formula in the following equation (BVF 587.02, 1997):

$$Q = 150 - 100\left[\frac{\sigma_H}{\sigma_{H_{\lim}}} + 2\frac{\sigma_S}{\sigma_{S_{\lim}}}\right]/3 \tag{3.5}$$

where $\sigma_{S_{\lim}}$ and $\sigma_{H_{\lim}}$ are the comfort limit for the σ_s value and the comfort limit for the σ_H value, respectively. Both limits are defined for different track classes see Table 3.2. The *Q*-value is represented as a percentage. The lower value of the state condition indicates that the train may sway and be perceived uncomfortable by the passengers and vice versa.

The other index, the *K*-value, is the ratio of the total length of the track with deviations below comfort limits Σ^l and the total length of the track L_{tr}. This index is used to obtain an overall picture of the track condition over a long distance and is calculated by the Eqn.(3.6);

$$K = \frac{\Sigma^l}{L_{tr}} \times 100\%\,. \tag{3.6}$$

3.3 Deterioration of track geometry quality

Periodic implementation of an automatic track inspection by means of a track recording car generates data with measurements of track geometry parameters for an equally-distance track points, e.g. 0.25 meter in a distance. After several track measurements from the last restoration point, a track engineer could observe monotonic changes in the value of TQI over time or a specified measureable quantity e.g. accumulated tonnage. This mechanism fits the definition of deterioration-the change of a component's original appearance as a result of both internal and external influences (Kumar and Gardoni, 2013).

Despite track deterioration is slow but failing to plan and deliver an efficient and effective inspection and maintenance strategy could result in high-cost consequences emerging from unplanned track maintenance (Network Rail Limited, 2016). Moreover, track maintenance costs accounting for over half of railway maintenance budgets (Johansson and Nilsson, 2004). A proactive approach to mitigate risk of unplanned track maintenance is through modelling an evolution of track geometry deterioration (Zhang, Andrews and Wang, 2013).

Not only dealing with unplanned track maintenance, track deterioration models have been widely applied as decision support tools at the design stage of track geometry inspection planning. In general, the models assist RIM in forecasting when a track condition falls (passes) below (above) the acceptable level of quality and thus require maintenance action. A key benefit of this particular method of fusion is the deposit of a significant reduction in the number of (predictable and/or unnecessary) inspections, which leads to cost-effectiveness in maintaining railway infrastructures (Stenström *et al.*, 2015; Dindar *et al.*, 2016). Comparison analysis, as reported in the financial reports of RICs such as in (Network Rail, 2008), provides real business evidence that supports the theoretical inference.

As TQIs which are the processed track measurement data satisfy the characteristics of deterioration data, thus, universal deterioration modelling methods, for example a statistical method, can be employed to construct a track deterioration model.

3.3.1 A generic formula of deterioration model

A track geometry deterioration model can be used to estimate the degree of changes in an index quality of a track segment over accumulated tonnage, particularly, when designing an inspection schedule. A fixed length segmentation, normally 100 m or 200 m long segment is typically applied to a tangent track. More segmentation algorithms that have been introduced to track maintenance management can be found in (Guler, 2012; Soleimanmeigouni, Ahmadi and Kumar, 2016). Given a collection of track segments that satisfy following assumptions:

i. track segments are measured in a particular homogenous environment (e.g. the same track measurement vehicle operates at the same time period), and
ii. the measurement (or inspection) times are pre-specified, the same across all the participated segments, and may or may not be equally spaced in time,

thus, it is possible to model the deterioration of the individual track segments using the same functional form and differences between individual segments using random effects. Let Q_{ij} be the measured deterioration of the *i*-th track segment each having *j* measurements, the generic deterioration model can be expressed as (Lu and Meeker, 1993):

$$Q_{ij} = D_{ij} = D\left(t_{ij}; \boldsymbol{\phi}, \boldsymbol{X}_i\right) + \varepsilon_{ij} \tag{3.7}$$

where $D\left(t_{ij}; \boldsymbol{\eta}_i, \boldsymbol{\vartheta}_i\right)$ represents the actual deterioration path of track segment *i* at time t_{ij} $\boldsymbol{\phi} = \left(\phi_1; \phi_2; \ldots, \phi_{|\boldsymbol{\phi}|}\right)$ is a vector of fixed effects that describes population characteristics (they remain constant for all segments); $\boldsymbol{X}_i = \left(X_{1i}, X_{2i}, \ldots, X_{ki}\right)$ is a vector of the *i*-th segment random-effect parameters that representing individual segment's characteristics; *k* is the number of random effect parameter. The measurement error ε_{ij}'s are assumed to be independent and identically distributed with zero mean and constant variance σ_ε^2, $\left(i = 1, \ldots, n;\ j = 1, \ldots, m_i\right)$ where *n* is the number of participated track segments and m_i is the total number of measurements on the *i*-th segment. It is assumed that $\boldsymbol{X}_i \left(i = 1, \ldots, n\right)$ is independent of each other and follow a multivariate distribution function $G(\cdot)$ which may depend on some unknown parameters that must be estimated from deterioration data (Lu and Meeker, 1993). In this section, a track deterioration model is built upon an index of track condition. Hence, any TQI where some are presented in Section 3.2.1 can be used representing a dependent variable of Eqn. (3.7).

In work by Bae, Kuo and Kvam (2007), the Eqn. (3.7) can be expressed as follows:

$$Q_{ij} = D_{ij} = D\left(t_{ij}; \boldsymbol{\phi}, \boldsymbol{X}_i\right) + \varepsilon_{ij} = \eta\left(t_{ij}; \boldsymbol{\phi}\right) + \boldsymbol{X}_i + \varepsilon_{ij} \tag{3.8}$$

where $\eta\left(t_{ij}, \boldsymbol{\phi}\right)$ is a deterministic mean deterioration path with fixed effect parameters $\boldsymbol{\phi}$ for time $t \geq 0$. The focus is on η being monotonic since most deterioration measurements have this quality.

3.3.2 A linear deterioration model

In the context of track-geometry maintenance, a deterioration model has been developed empirically under a different degree of polynomial; however, a linear model has also been of interest to researchers for a number of years (Chang, Liu and Li, 2010). The trade-off between complexity and readable features is a fundamental issue when presenting a deterioration model for decision-making use. In fact, the simplicity of a linear deterioration model results in significant reductions in computational complexity considering the immense size of a railway network. The linear deterioration model is expressed in the following form:

$$Q_{ij} = D_{ij} = D\left(t_{ij}; \phi_i; X_i\right) + \varepsilon_{ij} = \phi_i + X_i t_{ij} + \varepsilon_{ij} \tag{3.9}$$

where ϕ_i and X_i denote the initial state after tamping (or renewal) and rate of deterioration of the linear model for track segment i. The linear models assume that $P(\varphi_i) \leq 0$ is negligible for each segment in order to avoid the certain probability of getting non-feasible deterioration rate (Cheng and Peng, 2015).

A number of studies have applied a linear model for track deterioration. For example, Andrade and Teixeira (2015) constructed the linear model to predict on track geometrical characteristics; standard deviation of longitudinal level and alignment. Their findings show that the deterioration paths before and after tamping will follow two lines with different slopes and intercepts. In addition, the spatial dependency of model parameters is considered in consecutive track sections. They also applied Conditional Autoregressive model to capture the possible spatial dependence. A similar work can be found in Felipe and Teixeira (2016). The authors stressed the need for an assessment of model prediction error before determining a tamping schedule. That study highlighted that errors associates with the deterioration model have meaningful influences on the reliability of the selected track quality indices when searching for optimal track maintenance strategy. Additionally, Chang et al. (2010) utilized a multi-stage linear model with data consisting of 200-m track sections totalling 184 km from the Beijing-Jiulong Railway Line.

3.3.3 Deterioration analysis

Upon availability of deterioration data, several 'classical' methods such as full maximum likelihood estimation, non-linear mixed effects model and two-stage method can be applied to get the estimates of both $\boldsymbol{\phi}$ and $\boldsymbol{X}$. The abovementioned methods are worthwhile to discover for parameters estimation in a case when model parameters appear nonlinearly in the deterioration path. For a simple path (assuming we have a basis or some knowledge to justify the claim/selection), an analytical method which relying on an interaction between a deterioration process and failure time prediction provides a solution to the problem of parameter estimation (Lu and Meeker, 1993).

Under time-varying working loads and dynamic operating environments, the asset under consideration deteriorates, hence, its reliability decreases. When a deterioration path reaches a specified deterioration threshold D_f, thus, the asset meets the definition of 'failure'. Let time to failure $T(t)$ of a deteriorating asset is assumed to be continuous with distribution function $F_T(t)$ and is given as

$$\begin{aligned} F_T(t) &= P(T \le t) = P\left[D(t;\boldsymbol{\phi},\mathbf{X}) \ge D_f\right] \\ &= P\left[\mathbf{X} \ge D_f - \eta(\mathrm{t};\boldsymbol{\phi})\right] = 1 - G_{\boldsymbol{X}}\left(D_f - \eta(\mathrm{t};\boldsymbol{\phi})\right) \end{aligned} \tag{3.10}$$

when the deterioration paths are monotonically increasing with time or

$$\begin{aligned} F_T(t) &= P(T \le t) = P\left[D(t;\boldsymbol{\phi},\boldsymbol{X}) \le D_f\right] \\ &= P\left[\boldsymbol{X} \le D_f - \eta(t;\boldsymbol{\phi})\right] = G_X\left(D_f - \eta(t;\boldsymbol{\phi})\right) \end{aligned} \tag{3.11}$$

when the deterioration paths are monotonically decreasing with time (Freitas *et al.*, 2010).

Note that, a proper failure distribution is determined only if $G_X\left(D_f-\eta\left(0-;\boldsymbol{\phi}\right)\right)=1$ and $G_X\left(D_f-\eta\left(+\infty;\boldsymbol{\phi}\right)\right)=0$. For the decreasing deterioration path, the formulation requires that $G_{\boldsymbol{X}}\left(D_f-\eta(0-;\boldsymbol{\phi})\right)=0$ and $G_{\boldsymbol{X}}\left(D_f-\eta(+\infty;\boldsymbol{\phi})\right)=1$. If $\eta(t;\boldsymbol{\phi})$ has finite asymptotes, for example, the additive deterioration model will not necessarily produce a proper lifetime distribution function. Along with these constraints on *G*, we assume *G* (and hence *F*) is twice differentiable on $(0,\infty)$. In the occasion the distribution function can be expressed in a closed-form (analytical) so as to avoid numerical integration, the transformation process to obtain an estimator for $\boldsymbol{\phi}$ and $\boldsymbol{X}$ becomes less computational intense i.e. it involves a simple algebraic manipulation (Meeker, Escobar and Lu, 1998). Moreover, a simple closed-form expression for failure distribution is more useful for maintenance optimisation, since one can easily obtain the corresponding reliability function and failure rate function.

Consider an actual deterioration path of a particular track segment given in the Eqn. (3.9) where the initial amount of deterioration ϕ is fixed for all segments and the random parameter X that varies from segment to segment is Weibull distributed with cumulative distribution function (CDF) is given by,

$$G_X\left(X\right)=P\left(X\le x\right)=1-\exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] \tag{3.12}$$

with α and β is the time-dependent scale and shape parameter of the distribution, respectively. The CDF of *T* is then given as follows:

$$
\begin{aligned}
F_T(t) &= P(T \le t) \\
&= P\left(\frac{D_f - \phi}{X} \le t\right) = P\left(X \ge \frac{D_f - \phi}{t}\right) = 1 - G_X \\
&= \exp\left[-\left(\frac{\frac{D_f - \phi}{t}}{\alpha}\right)^{\beta}\right] = \exp\left[-\left(\frac{(D_f - \phi)/\alpha}{t}\right)^{\beta}\right], t > 0
\end{aligned}
\tag{3.13}
$$

This is a CDF of reciprocal (inverse) of Weibull because $1/T$ follows a Weibull distribution. At this point, if the distribution fitted failure data, thus, an estimator for the rate of deterioration can be determined easily.

Similarly, if $X \sim \log N(\mu_X, \sigma_X)$ then the CDF of failure time is given by

$$
\begin{aligned}
F_T(t) = P\left(X \ge \frac{D_f - \phi}{t}\right) &= 1 - \Phi\left(\frac{\log\left(\frac{D_f - \phi}{t}\right) - \mu_X}{\sigma_X}\right) \\
&= 1 - \Phi\left(\frac{\log(D_f - \phi) - \log(t) - \mu_X}{\sigma_X}\right) \\
&= 1 - \Phi\left(\frac{-\log(t) + \log(D_f - \phi) - \mu_X}{\sigma_X}\right) \\
&= \Phi\left(\frac{\log(t) - \left[\log(D_f - \phi) - \mu_X\right]}{\sigma_X}\right), t > 0
\end{aligned}
\tag{3.14}
$$

where $\Phi(.)$ is the standard normal distribution function. This shows that T has a lognormal distribution and $\mu_T = \log(D_f - \phi) - u_X$ and $\sigma_T = \sigma_B$, respectively, the mean and standard deviation of $\log(T)$. If X is assumed to be a random effect that follows exponential distribution having $G_X(x) = 1 - \exp(-\lambda x)$, then CDF of failure time F_T is

$$F_T(t) = P\left(X \geq \frac{D_f - \phi}{t}\right)$$
$$= 1 - \left[1 - \exp\left(-\lambda\left(\frac{D_f - \phi}{t}\right)\right)\right] = \exp\left(-\frac{\lambda\left(D_f - \phi\right)}{t}\right), t > 0. \tag{3.15}$$

Suppose that a deterioration path in Eqn. (3.8) is written in the form of $D_{ij} = \phi \exp(Xt); \phi > 0, t \geq 0$, and then the corresponding failure distribution function would be similar to Eqn. (3.15) i.e. $\log\left(D_f / \phi\right)$ instead of $D_f - \phi$.

3.3.4 Updating prior models using inspection results for uncertainty propagation assessment

In the absence of uniform level of parametric uncertainty in the track deterioration linear model, vast of the model outputs are limited to an extrapolation and estimation of gradual deterioration, which leads to a steady dependency on periodic inspections. For the last decades, inspection costs still occupy a substantial percentage of a railway infrastructure company's budget. Unfortunately, a recent survey exposes that, in overall, up to 80% of results of inspections decisions in a railway sector are as expected i.e. no defects detected (Stenström *et al.*, 2015). Despite carrying out inspection helps to mitigate a risk of unplanned maintenance but this statistics is a clear evidence of inspection inefficiency; an excessive number of inspections. This situation, however, progressively encourages researchers to improve their understanding of unsettled elements in track deterioration models, particularly unexplained variance (uncertainty). Through updating belief about prior specifications of degree of uncertainty in track model parameters, the current inspection design would gain an improvement in terms of the value of investment (not necessary concerning cost reduction) for next maintenance cycles.

Gligorijevic et al. (2016) stated that an inspection interval is determined by properly estimating the uncertainty propagation in the model under study. By performing uncertainty propagation, researchers would witness a decreasing trend in the reliability of model prediction caused by the effects of noisiness in input data when predicting further in the future. Andrade and Teixeira (2011) emphasises the importance of proper uncertainty quantification for degradation model parameters when estimating the maintenance life-cycle cost. The authors demonstrated an important correlation between the spatial nature of tracks, such as switch and bridge, and the deterioration rate in the standard deviation of longitudinal-levelling defects. A discussion about the potential effects of improper uncertainty quantification in determining maintenance interval is presented in (Bell and Percy, 2012). In that paper, the authors study the joint application of a Bayesian approach and Markov Chain Monte Carlo methods to reduce the likelihood of delivering false conclusions regarding uncertainty modelling. However, the applicability of the proposed method for a system that adopts components classification, such as railway lines, is not yet proven.

In general, an evaluation of uncertainty propagation links to updating a rate of deterioration in asset condition. This is particularly linked to the re-evaluation of risk levels and these have a direct impact on risk reduction (Zhang and Mahadevan, 2003; Ishak, Dindar and Kaewunruen, 2016). He et al. (2013) materialize this inference with a development of analytical framework for track geo-defect repairs where a geometry degradation model together with survival model was incorporated into the optimization formulation. Roughly, about 20% of the total repair cost can be saved from the proposed model. However, this claim may not be achieved in reality as RIM tends to group various types of track repair works to minimize track possessions. In respect to risk of train derailment, it has imposed significant challenges on RIMs considering the relative size of financial loss in assets (including public property), damages and service disruptions (Nissen, 2009b).

3.3.5 Bayesian inference for updating prior deterioration model

In order to carry out uncertainty propagation, use of probabilistic representation is common to represent both aleatory and epistemic uncertainty. According to Bedford (2008) and Revie et al. (2010), the fundamental problem of probabilistic representation lies in the selection of prior probability distribution, where in most situations, a parameter of interest is quantified with a poor distribution, causing inaccuracy in the prediction results, forecasting, or inference. This shortcoming can be addressed using the Bayesian approach, which uses new data to update prior distribution. The Bayesian approach provides a theoretical inference framework for updating prior beliefs about uncertain quantities once additional information becomes available (if the decision maker can make observations) from the tests and analyses conducted during the development program. An early work on uncertainty assessment using the Bayesian approach has been reported since early 1970 (Randell et al., 2010)**.** Until now, a wide range of extensions has been developed (see review in Lu and Madanat 1994, Zhang and Mahadevan 2003), and most of the works were developed under a probabilistic Bayesian framework.

When a full detailed probabilistic analysis is too costly to perform, and the belief in parameters of interest is partially elicited, the benefits of conventional Bayesian method is shadowed by the high volume of computational and elicitation effort. This element is among the reasons why corporates unwilling to invest on soft computing innovations, which links us to pragmatism terms (Panda and Gupta, 2014). In this situation, approximations to the traditional Bayesian analyses, known as Bayes linear analyses, have been proposed as a logical and justifiable framework to express and review on the beliefs about the recognised uncertain quantities. Unlike the conventional Bayesian method--which heavily depends on fully-specified probability distributions--the Bayes linear method linearly adjusted the prior beliefs about these uncertain quantities based upon the theory of Bayes linear statistics (Goldstein and Wooff, 2007). Instead of using probability as a basis (proxy), Bayes linear method uses the first- and second-order moments to model beliefs for the quantities of interest. This means that decision maker's degree of uncertainty of a quantity of interest e.g. a model parameter is represented by variance.

3.3.6 Bayes linear method

Bayes linear method provides a linear structure of belief specifications which allows users to add new elements to the model relatively easily. In fact, users get flexibility to combine lines of evidence of varying quality from many disparate sources of information when assessing uncertainty about elements of quantity of interest, for example, a rate of change of track geometry deterioration model. Interestingly, adjustments on model specifications are tractable under Bayes linear framework where in some cases it can be performed instantaneously, in particular, when multidimensional space needs to be adjusted. Longer computational time is probably taken when using conventional Bayesian approach.

The term 'linear' in Bayes linear method defines a linear relationship between vector ***B*** and ***D*** in $\boldsymbol{D} = \alpha \boldsymbol{B} + \boldsymbol{R}$ where ***R*** represents the unexplained variance between ***B*** and ***D***. The vector ***B*** and ***D*** denotes a belief structure representing uncertain quantities of interest, B_i and is some vector of quantities that might improve decision maker's prior assessment of ***B***, respectively. First- and second-order moment of ***B***, denoted by, $E(\boldsymbol{B})$ and $var(\boldsymbol{B})$ will be adjusted using elicitation and observed values of ***D***. Prior to the adjustments, decision maker should elicit $E(\boldsymbol{D})$ and $var(\boldsymbol{D})$, and as well constructs a covariance matrix $\text{cov}(\boldsymbol{B},\boldsymbol{D})$ which address the degrees of relationship between ***B*** and ***D***. Note that, the matrix must satisfy characteristics of non-negative definite matrix. Following formula in (Goldstein and Wooff, 2007), the collection ***B***, respectively, has an adjusted expectation and variance matrix:

$$E_{\boldsymbol{D}}(\boldsymbol{B}) = E(\boldsymbol{B}) + \text{cov}(\boldsymbol{B},\boldsymbol{D})\,\text{var}^{\Psi}(\boldsymbol{D})(\boldsymbol{D} - \boldsymbol{E}(\boldsymbol{D})) \tag{3.16}$$

$$var_{\boldsymbol{D}}(\boldsymbol{B}) = var(\boldsymbol{B}) + \text{cov}(\boldsymbol{B},\boldsymbol{D})\,\text{var}^{\Psi}(\boldsymbol{D})\text{cov}(\boldsymbol{D},\boldsymbol{B}) \tag{3.17}$$

where $\text{var}^{\Psi}(\boldsymbol{D})$ is the Moore-Penrose generalized inverse. In case of $\text{var}(\boldsymbol{D})$ is non-singular then $\text{var}^{\Psi}(\boldsymbol{D})$ is simply the usual matrix inverse i.e. $\text{var}^{\Psi}(\boldsymbol{D}) = \text{var}^{-1}(\boldsymbol{D})$.

Consider the collection $\boldsymbol{B} = (\beta_{\boldsymbol{o}}, \beta_1)$ is an interest of one's study. Given observations on a collection of observable quantities $\boldsymbol{D} = (D_1, D_2, \ldots, D_{\boldsymbol{m}})$, prior belief about a vector ***B*** updates via the adjusted expectation, $E_{\boldsymbol{D}}(\boldsymbol{B})$. By calculating the size of adjustment over ***B*** given by observed values of *D* using an Eqn. (3.18), it is able to quantify how deviation of the adjusted expectation is from the prior expectation. Application of a similar principle then occurs to calculate an adjustment over ***B*** given by a portion of ***D***. For an individual assessment, the size of partial adjustment may have referred to and derived from the Eqn. (3.19).

$$Size_D(\boldsymbol{B}) = [E_D(\boldsymbol{B}) - E(\boldsymbol{B})]^T\ \text{var}^{\Psi}(\boldsymbol{B})[E_D(\boldsymbol{B}) - E(\boldsymbol{B})] \tag{3.18}$$

$$Size_{[F/D]}(\boldsymbol{B}) = [E_{F\cup D}(\boldsymbol{B}) - E_D(\boldsymbol{B})]^T\ \text{var}^{\Psi}(\boldsymbol{B})[E_{F\cup D}(\boldsymbol{B}) - E_D(\boldsymbol{B})] \tag{3.19}$$

A partial bearing for the partial adjustment, denoted by $Z_{[F/D]}(\boldsymbol{B})$ expresses both the direction and the magnitude of the changes over ***B*** when we additionally adjust ***B*** by ***F*** given a preceding adjustment by ***D***, through the relation

$$\text{cov}_D\left(B_i, \boldsymbol{Z}_{F/D}(\boldsymbol{B})\right) = E_{D\cup F}(B_i) - E_D(B_i); \forall B_i \in \boldsymbol{B}. \tag{3.20}$$

CHAPTER 4 RESEARCH METHODOLOGY

4.1 Introduction

Comparing to the past decades where the number and sources of disruptions in railway operations are small and limited, embedding a track inspection plan with a resilience model to exhibit a resilience function is an exception. Nowadays, the scenario has changed greatly; not only a disruption has an increasing trend in the number of occurrence but it also occurred in new various ways (Osman *et al.*, 2016). Recent developments in respect to railway operations enable organisations to rethink the resilience of track inspection plan. One of potential strategies in promoting resilience function is turning a resilience model a must-have component to complement conventional methodologies for planning track inspections.

As aforementioned, this research aims to develop a risk-resilience response and recovery model for a disrupted planned railway track inspection plan. The proposed model takes the form of risk-resilience strategy, while having been developed under the disruption management framework. Fig. 4.1 shows the sequence of the research activities that has been performed in this research. The research begins by conducting a comprehensive literature survey to address the research aim and objectives as described in Section 1.3. A theoretical analysis at the end of the literature review gives a clear path to identifying significant gaps in current practice and to complete the research objectives and answer the research questions. The main body of this research is divided into two phases. The first phase is to develop a spatial-temporal prediction model under a method of fusion of NARX and neural network. The model generates 'missing' track measurements data to respond on the occurrence of disruptive events in a track inspection plan. According to the framework of resilience concept the prediction model will be known as a response action. Investigations were then continued further by framing an appropriate response strategy for managing low probability risk. This is important as disruptive events are classified as low probability events where its consequences are dynamically changed over time. In order to overcome limitations of traditional risk evaluation when dealing with low probability events the concept of resilience has been adopted in this thesis. The resilience concept offers a transparent methodology to evaluate consequences of disruptive events and to value of a response model at different phases of disruption management. Moreover, operationally resilient track inspection must have the ability to rapidly respond, recover from disruptions and exploit opportunities from the anticipated events effectively.

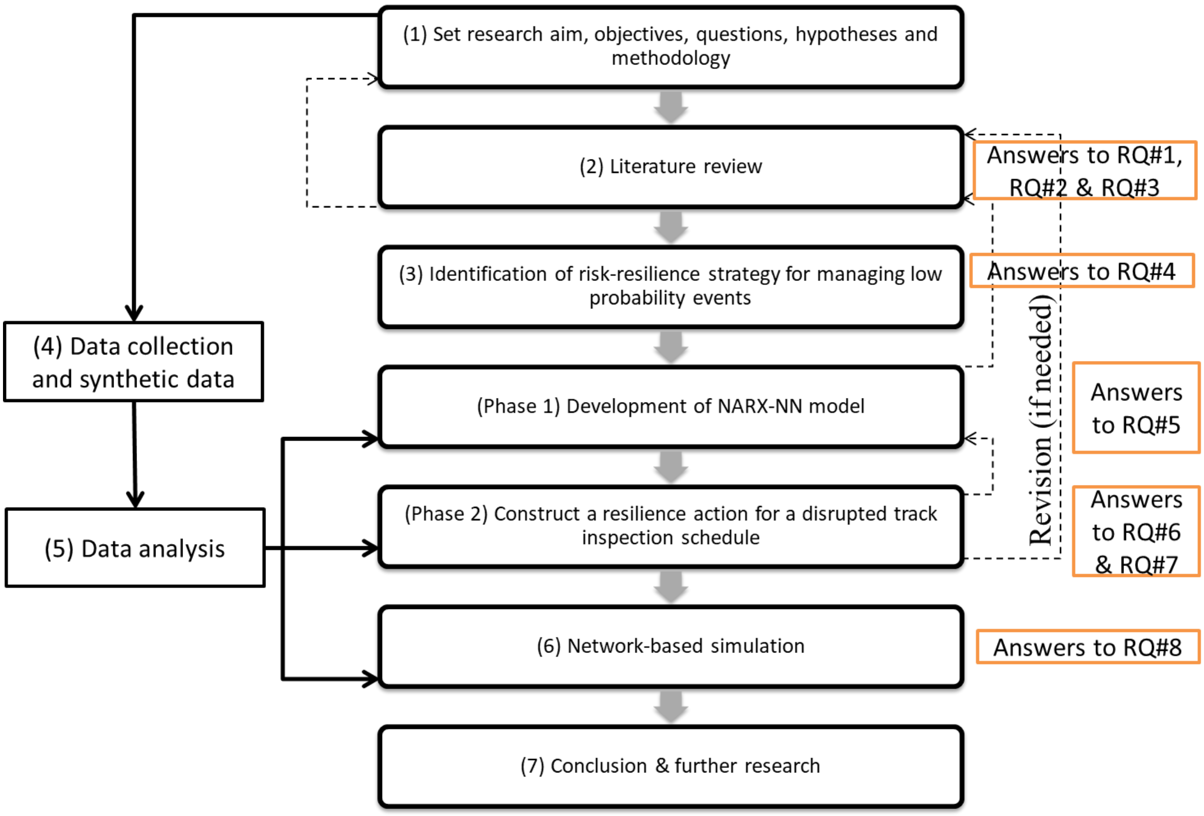


Fig. 4.1 Research methodology

The development of prediction model was then validated in a recovery phase of resilience model. Rescheduling framework was introduced to demonstrate performance of the predicted results in the context of recoverability capability. Various problem scenarios have been setup to address any (serious) limitations in the rescheduling framework.

4.2 Developing the iNARXNN model

The prediction model forms a core focus of this study since it acts as a response to absorb the performance losses in the inspection due to negative impacts from disruptive events. In an effort to lowering the cost of implementing the response action the prediction model operates on the secondary data only which means the data acquisition cost is virtually free. In other words, the prediction model itself generate value-added to the primary data. Hence, an analysis process for secondary data was adopted. Basically, when existing data (termed primary data) are reused (for example, posting research questions of the data that were different from the original question), then the data serves secondary purposes. There are six main steps involved in the data analysis pipeline, briefly stated as follows (Hox and Boeijie, 2005; Richmond, 2006):

i. Design research question(s): Constructs quality research question(s) is important to make the best use of (limited) resources on data collection and analytical tools. Questions should be measurable, clear and concise. Design the questions to either qualify or disqualify potential solutions to the specific problem or opportunity.
ii. Identify secondary data set: A desktop review on the topic of interest is necessary to determine whether there are existing data that would be perfectly reusable in a new research setting, therefore helping researcher to answer research question(s) more thoroughly (and easily). In any case, it is a researcher’s responsibility to ensure that a secondary data set is a good fit for his or her research question.

iii. Evaluate dataset: It is researchers' responsibility to carefully evaluate the quality of existing data before use it to address a new line up of research question. Basic but important information about the primary data that should be at least secured are purpose of the study, operationalization and details on data collection (who, when and where, entities being studied and sampling criteria).

iv. Preparing data for analysis: The data that are collected must be processed or organized for analysis. This process includes structuring the data as required for the relevant analysis tools.

v. Data analysis: Data that are processed, organized and cleaned would be ready for the analysis. During this step, one or more set of results would be obtained from data collected depending on the selection made on analytical procedure or methods.

vi. Results and interpretation: Report results of the data analysis in a logical sequence, that is, allow readers to transit from one finding to another smoothly. Any biased comments should be avoided. Following this, interpret the results to find how it relates to research problems and research questions.

An illustration of six steps in the data analysis process is given in Fig. 4.2, where some steps are interactive.

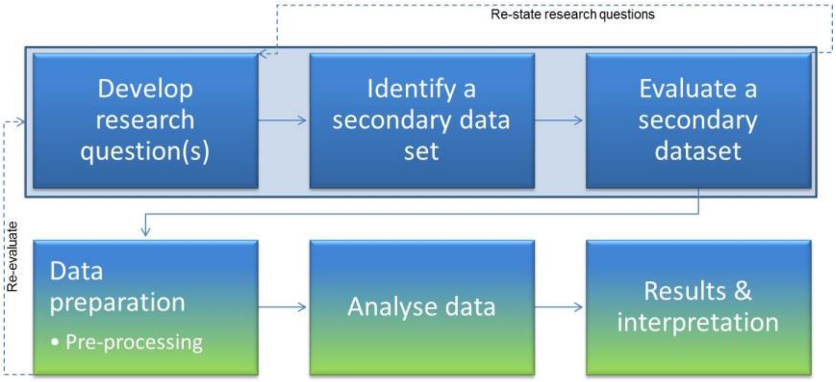


Fig. 4.2 Transforming primary data to (new) information through secondary data analysis

4.3 Developing the recovery model

The investigation was then continued further in the phase of resilience action. At this phase, iNARXNN results were analysed to provide necessary inputs for a track inspection reschedule. The key indicator in this analysis is several track segments might need a new inspection time, i.e. shorten or extend an initial interval. To be reminded, this phase will incur some costs, e.g. rescheduling, new track possession, etc., that probably overshadow benefits of the resilience action. Hence, the inspection reschedule was then formulated as a single constrained optimisation problem which aiming to maximise a specified resilience metric.

Rescheduling can be viewed as an act of remapping the existing schedule due to unavoidable reasons. This relationship creates a positive link to adopt a structure of scheduling model to formulate and solve a rescheduling problem and can be performed subject to an outcome of step (2). In addition, a careful understanding about rescheduling framework is necessary. Vieira et al. (2003) have done an excellent job in developing the rescheduling framework, which has been used by many practitioners in various industries, including railway operation management (Acuña-Agost, 2009). According to the framework, a good justification of environment, strategy and policy of rescheduling helps practitioners be more specific and bold when selecting a method.

Rescheduling is an additional investment by organisations, e.g. railway firms, to limit the impacts of disruptions on the schedule while maintaining resource allocation and avoiding inter-demand conflicts as well as deadlocks. To maximise the benefits of the investment, a rescheduling strategy should be appropriately selected. A predictive-reactive strategy is commonly adopted for static scheduling. Generally, the impact caused by disruptions is evaluated and used as a basis to decide what kinds of rescheduling solution is required. Alternative solutions are necessary if the first result is unacceptable. Unlike a dynamic strategy for dynamic schedules, the latter strategy can only be implemented if a rescheduling policy is provided. Whether periodic, event-driven, or hybrid policy, each defines how and when an original schedule will be updated.

Each of potential solutions from the rescheduling model will be assigned with a value of resilience metric. The metric has been developed on the appraisal of benefits and costs components as recommended by railway management (Rail Safety and Standards Board, 2014). For both cost and benefit components, there will be internal and external contributors. Difficulties arise in several directions, for example, the need for a reliable model to quantify computational time taken, method/algorithm requirement and environmental impacts in term of monetary value as a cost measure. Those features may at the same time inherit individual complexity from their own respective natures. A solution to that particular issue is a generalised resilience cost that assumes different elements measured in a different scale unit that can be combined into one index by performing time/carbon emission/safety/risk to money conversion (Lesley, 2009). A weighting system is applied in the calculation whereby preferences of decision-maker(s) navigate weight values assignment.

In the event the benefit of rescheduling is underperformed, the user is advised to revisit the rescheduling stage before exiting the process. Making adjustments to the rescheduling method, for example, might increase the magnitude of internal benefit as well as overall i.e. sensitivity analysis (Kaewunruen and Remennikov, 2006). If the cost component is still superior after such adjustments, then absolutely no new schedule is needed for the next inspection, i.e. keep going with the initial schedule.

CHAPTER 5 DATA ANALYSIS

Data analysis results presented in this chapter will serve as prior knowledge in i) the development of a response model, and ii) the execution of resilience action for a disrupted planned railway track inspection plan presented in Chapters 6 and 7, respectively.

Fig. 5.1 illustrates a general design of the information flow of an inspection-to-maintenance diagram. A series of on-site inspections is assigned systematically across the railway network at different frequencies, subject to the accumulated traffic tonnage and speed category. The volume of inspection data in the repository increases as more inspections are completed. The information chain begins with a track recording of car samples and process track geometrical parameters in real-time, followed by computerized data processing and analysis, with the track quality index as the output. The quality index will be compared with a set of maintenance limits to define a suitable maintenance strategy to restore the quality of the inspected track. Here, track engineers justify their decision on this particular matter by presenting a risk value associated with the functional failure of a track section between two consecutive inspections. In parallel, a track geometry deterioration model is widely used to estimate the degree of changes in track condition with respect to time, loading, reliability, etc. Both the underestimation of the deterioration rate of track condition and the inability to capture a sudden shift event in a deterioration path are attributed to unplanned maintenance. In the event of disruptions where the on-site track measurements are temporarily unavailable, the risk-resilience model proposed in this study will operate to yield results that generally accomplished with primary data under disruption-free conditions.

The proposed response model entails two components working in a series; an iNARXNN, alongside a variance-based measure for predicting ‘missing’ track measurement data and an output valuation. Results of data analysis will be heavily used at the stage of post-iNARXX in particular for determining closed-form distributions for a deterioration rate of track quality.

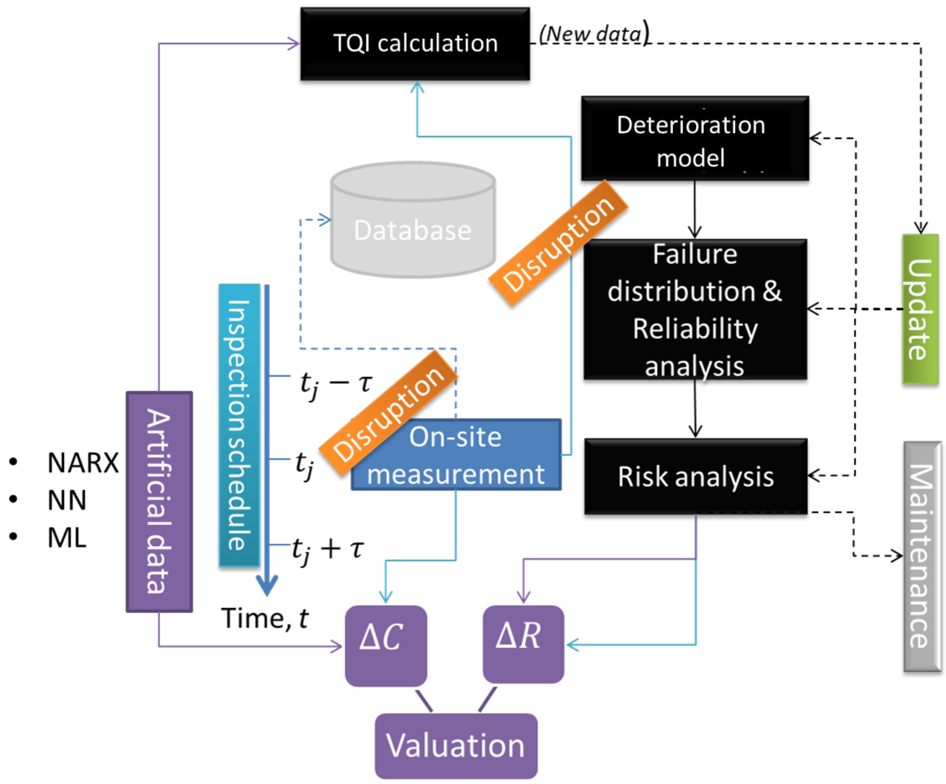


Fig. 5.1 Updating the prior belief of deterioration parameters of a track quality model based on track measurements

When a disruption occurs in the components of track inspection the predicted data is proposed to play a (temporary) role of track measurement in the process of belief updating on the level of track safety, as illustrated in Fig. 5.1. This assignment is termed as a resilience action which may incur additional investments to RIM to recover the disrupted track inspection. Therefore, the risk-aversion based model is formulated in order to quantitatively evaluate the benefits and costs of accepting the iNARXNN's outputs as a recovery agent. Again, results of data analysis will elicit the value of rescheduling that RIMs are willing to accept to execute a resilience action.

5.1 Tamping record

Tamping maintenance records may provide the best possible source to sample track failure (or survival) data. These records can be used to describe tamping effects on track reliability (or survivability) either over the course of one or several failure cycles. The *Tamping Date* attribute in the record defines an exact failure time if the targeted track has (or has nearly) reached a state of 'soft' failure on the tamping day. The 'soft' failure term is used because trains are still operating on defective tracks but at lower speed. However, performance of the track is degrading over remaining availability time. Of course, loose (but not excessively loose) definition of the failure would be imposed. This rule can be easily tested by evaluating a measure of track quality just prior to a tamping operation. This research advocates the concept presented in (Khouy *et al.*, 2013). If it is not the case (i.e. the track has yet to reach a state of failure), the tamping date is then expressed as a censoring point; the corresponding failure data are considered right censored. In any case, the presence of right-censored data in the failure time dataset can be easily dealt with using statistical techniques e.g. Kaplan-Meier, at which an additional assumption can be expected.

In this thesis, the tamping record was employed to address two research questions 1) Are there any variations in the (first) survivor time on track segments after preventive tamping? and 2) Would adjacent track segments have a different type of failure model? Answers to these questions will serve as prior knowledge about a track geometry-related failure distribution and repair model. It is necessary in a response model to develop a disrupted inspection schedule. For that purpose, we executed the abovementioned process, with a summary of all steps of the process being presented in Table 5.1.

Table 5.1 Answers to all steps in the secondary data analysis process.

Steps for secondary data analysis	Problem 1: Tamping effect variations	Problem 2: Repair model identification
Develop your research question	Are there any variations in the (first) survivor time on track segments after preventive tamping?	Would adjacent track segments have a different type of failure model?
Identify a secondary data set	Tamping maintenance record is archived in RIM's database with a codename BIS. The record registers tamping dates at which targeted track segments received tamping maintenance. An inter-arrival time between tamping can be easily calculated which are then transformed into traffic tonnage unit.	
Evaluate a secondary dataset		
(a) What was the aim of the original study?	The tamping record is part of the RIM's standard operating procedure and may be analysed for business performance appraisal.	
(b) Who has collected the data?	Track division of Trafikverket.	
(c) Which measures were employed?	Thirteen attributes of tamping maintenance.	
(d) When was the data collected?	From 2001-11 to 2018-04	
(e) What methodology was used to collect the data?	Data management procedure written for BIS.	
(f) Making a final evaluation	Sufficiently developed data set	

Data preparation	V={Segment ID, time to failure, equivalent accumulated tonnage}. Apply standard pre-processing tools where it is necessarily needed.	
Data analysis	Survival analysis	Repairable system model

5.1.1 Data visualisation

Data visualisation is a preliminary step in a data-driven approach. This step is powerful because it provides a representation in a visual context by making explicit the trends and patterns inherent in the data. Such pattern and trends may not be explicit in text-based data.

The original format of the tamping record is shown in Fig. 5.2. Each tamping operation was registered with 13 attributes of maintenance data; *Plstr* (location track code from-to), *UNE* (track category (uphill (U), downhill (N), single track (E)), *spr* (track type), *Bandel* (track line code), *KmFran* (kilometre marking from the starting point), *MeterFran* (meter marking from the starting point), *KmTill* (kilometre marking to the end point of the link), *MeterTill* (meter marking to the end point of the link), *spm* (length of track section in meter), *sid* (lateral), *Objekttyp* (activity type), *Objekt* (tamping date) and *objnr* (track number). The tamping record belongs to the Kilsmo-Palsboda track section, as illustrated in Fig. 5.3. The track section is 11.22 kilometre long spans over a rail line 416 which connecting Katrinholmes central and Hallsberg station. Out of 13 attributes, tamping date and track location were heavily used in data visualisation.

593x13 table

	1 Plstr	2 UNE	3 spr	4 Bandel	5 KmFran	6 MeterFran	7 KmTill	8 MeterTill	9 spm	10 sid	11 Objekttyp	12 Objekt	13 objnr
1	K-Bt	N	NaN	416	135	664	140	940	5.2688e+03	""	Spårriktning	2003-05	7039
2	K-Bt	N	2	416	135	664	142	258	6.5826e+03	""	Spårriktning	2013-09	50461
3	K-Bt	N	NaN	416	135	664	135	700	36	""	Spårriktning	2001-11	3918
4	K-Bt	N	NaN	416	135	665	135	700	35	""	Spårriktning	2005-08	14447
5	K-Bt	N	NaN	416	135	700	141	400	5694	""	Spårriktning	2005-04	13913

Fig. 5.2 Original data format for track tamping record

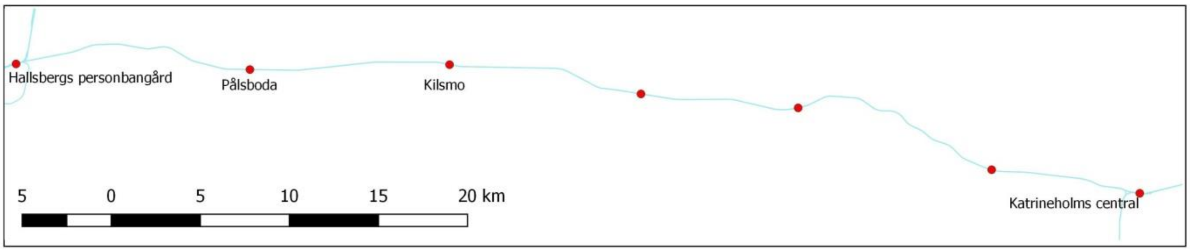


Fig. 5.3 Location of the analysed track section; one part of Line 416

5.1.2 Graph network

A graph network in Fig. 5.4 displays the variability in the number of tamping operations for track segments in the analysed track section over a 200-month period. Interestingly, several track segments receive only one time of tamping during the period. Along the 11.22 kilometres of an analysed track section, three track segments (labelled as segment no. 6, 15 and 25) received a maximum of ten tamping works. The segment labelled no. 6 is a curved track. The other two segments are located near the curve tracks. This observation relates to the fact that a train performs significant breaking/accelerating when entering/leaving a curve to achieve a desirable vehicle speed for minimising jerk effects. Nevertheless, further investigation (e.g. seeking expert opinion and/or examining additional record(s)) is required to understand technical explanations regarding why the segments have received more tamping than other tangent track segments. Soft-soil and/or drainage on the particular track segment may be the factor involved because a similar event is explained in (Ribeiro, 2014).

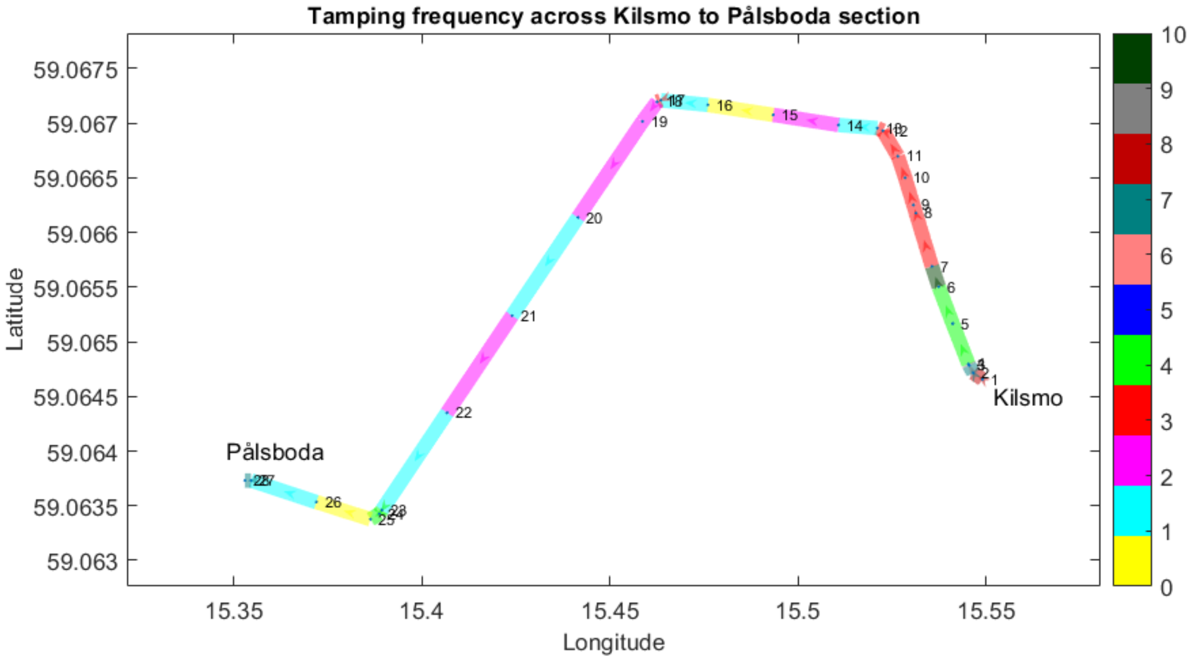


Fig. 5.4 Graph network representation provides an initial indicator to variations in tamping effect.

5.1.3 Tamping map

A tamping map (see Fig. 5.5) is a graphical tool applied to tamping records to discover underlying patterns from the record. A horizontal axis of the map represents a sequence of track points along the analysed track section in a vehicle direction. Tamping dates are arranged in ascending order from top to bottom along the vertical axis. A shaded block (rectangular) on the map indicates the location of a sequence of track points (a track segment) that receives tamping maintenance on the particular tamping date *t*. For referencing purpose, a tag in the form of $\{s_o, s_1, t\}$ is applied to any track segment on a tamping map where s_o and s_1 represent its start and end track points, respectively. The third component of the tag becomes an array of values if a track segment receives more than one tamping over the follow-up period. Tamping works on curved tracks are differentiated from a tangent track using a different shaded colour (e.g. yellow instead of light blue for tangent tracks).

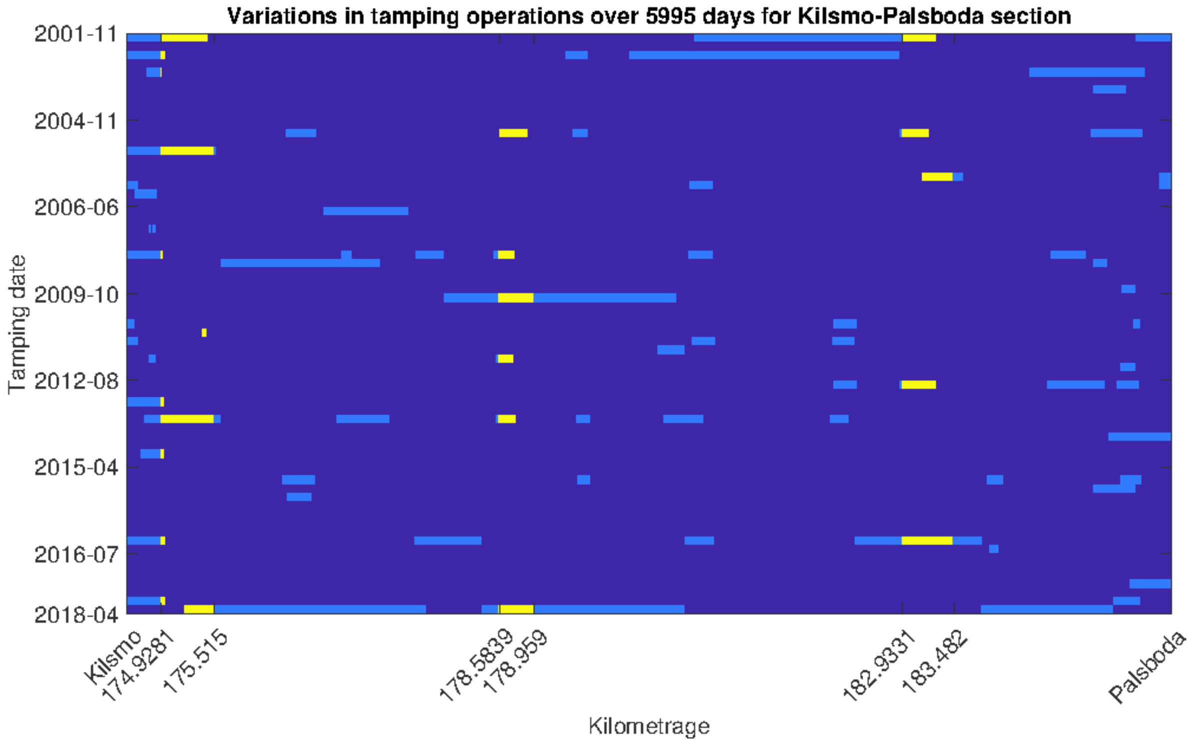


Fig. 5.5 Visual representation of tamping record

5.2 Analysis of tamping effectiveness

Are there any variations in the tamping effectiveness, i.e. the time elapsed since the last preventive tamping, on track segments after preventive tamping? Consider the elapsed time after condition restoration can be treated in the same way as survival data, and then survival analysis can be employed. Survival analysis also called time-to-event analysis is an analysis of survival data using a set of statistical techniques. The term survival data refer to the time from a well-defined time origin until an event of interest occurs. Whatever the event is, it should be properly defined and preferably observable. In the context of rail/track maintenance, it may be that the event is the time until the train wheel has worn out or the time until a track receives the second tamping and so on. Such data which not only involve whether or not an event occurred, but also when that event occurred, can thus also be referred to as time-to-event, survival time or failure time data. These terms are used synonymously in this section.

The elements required to be set prior to the execution of survival analysis (as illustrated in Fig. 5.6) are as follows:

i. Describing the event and time to event

The tamping operation is the event of interest in which the time to the event is not simply a tamping date but is defined by a time-intensity function. The function is linear and calculates an accumulated tonnage between two tamping operations applied to a subject of an event — a segment of tangent track span of 25 meters. This is a minimum length for tamping maintenance to effectively restore a track from short wavelength (3–25 m) track defects.

ii. Time of origin

Equally important, the time of origin should be clearly specified to ensure that all subjects are as much as possible equivalent. The ‘time origin’ is the point at which each subject begins, being at risk for the event of interest. For example, if the time of event of a track segment is being studied, the time origin could be chosen to be the time point of tamping maintenance applied to all subjects concurrently. A subsequent time origin is declared when a tamping machine pauses (or stops) its operation when it moves from one tamping location to another. In other words, only a series of track segments being tamped under continuous operation of a tamping machine share the same time origin.

A careful determination of the time origin is required to avoid observations on the subjects of the event contaminated by unknown sources of variation which may mislead the analysis findings. The action in fact honours an assumption that observations on the subject of an event of interest are identically distributed in preceding a statistical technique.

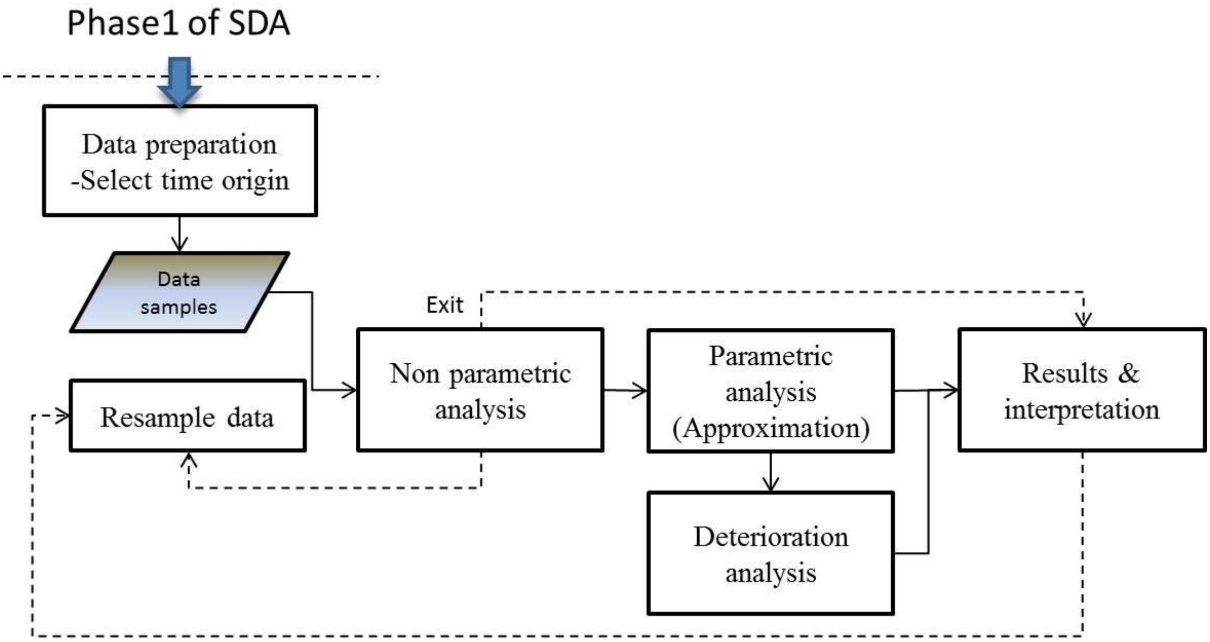


Fig. 5.6 Embedded survival analysis steps to a generic data analysis process

A tamping maintenance decision is primarily subject to longitudinal-level defects. This defect is due to vertical wheel load, where the resulting effects propagate only within a certain distance (e.g. a chord length). Theoretically speaking, two track points separated by 25 meters in distance are highly likely to be independent. This is a basis to claim that an observation on each subject in the survival analysis is independent to others. For now, the independently and identically distributed assumptions are satisfied at such levels of evidence.

The number of samples that will be used to extract information about reliability/survival function through a survival analysis is unknown and is only available to this study when (potential) time origins have been determined from a tamping record. Basically, the sample size needs to be sufficient to achieve a specified probability of a Type I error (or a significance level of approximately 5%) of the (hypothesis) test and power of the test (1-β). The power of the test is directly related to the number of observed events which arise because more subjects usually means more events; therefore, potential time origins could be at tamping date(s) registered for a long length of tamping (for example, most likely more than a 1-kilometre track).

5.2.1 Data preparation

A summary of the track length registered for tamping maintenance in the analysed track section for the past 16 years from November 2001 to April 2018 is illustrated in Fig. 5.7. The figure exhibits a clear gap (separation) in the tamping length for a tangent track. Tamping operations on the right side of the separation point (i.e. 1,000 meters) may be related to planned and preventive maintenance. Additional information from the infrastructure manager could verify this classification. No suggestion regarding the classification for curve track can be drawn out from Fig. 5.7. In any case, we should refer to the tamping maintenance decision framework, with its basic description being found in (Khouy *et al.*, 2013). At this point, we assume that the separation point could be a good starting point for identifying secondary datasets for survival analysis. Of course, the value of the separation point will be re-examined if further evidence is required to answer the research question.

There are eight tamping records registering a tamping length over 1,000 meters; basic information about these records is provided in Table 5.2. Three of these records associated with the tamping date of 201804 were excluded from the time origin selection. This decision can be easily understood because no subjects can be sampled from each of the records (i.e. all subject ends as right-censored data). For the remaining records in Table 5.2, their associated tamping date defines the time of origin. Hence, there will be five datasets for survival analysis. The time to event of each subject for each dataset is calculated by multiplying the calendar days between the time of origin and the event, with an average daily traffic load (in MGT). The average daily traffic for the Kilsmo-Palsboda track section is estimated to be 0.11MGT.

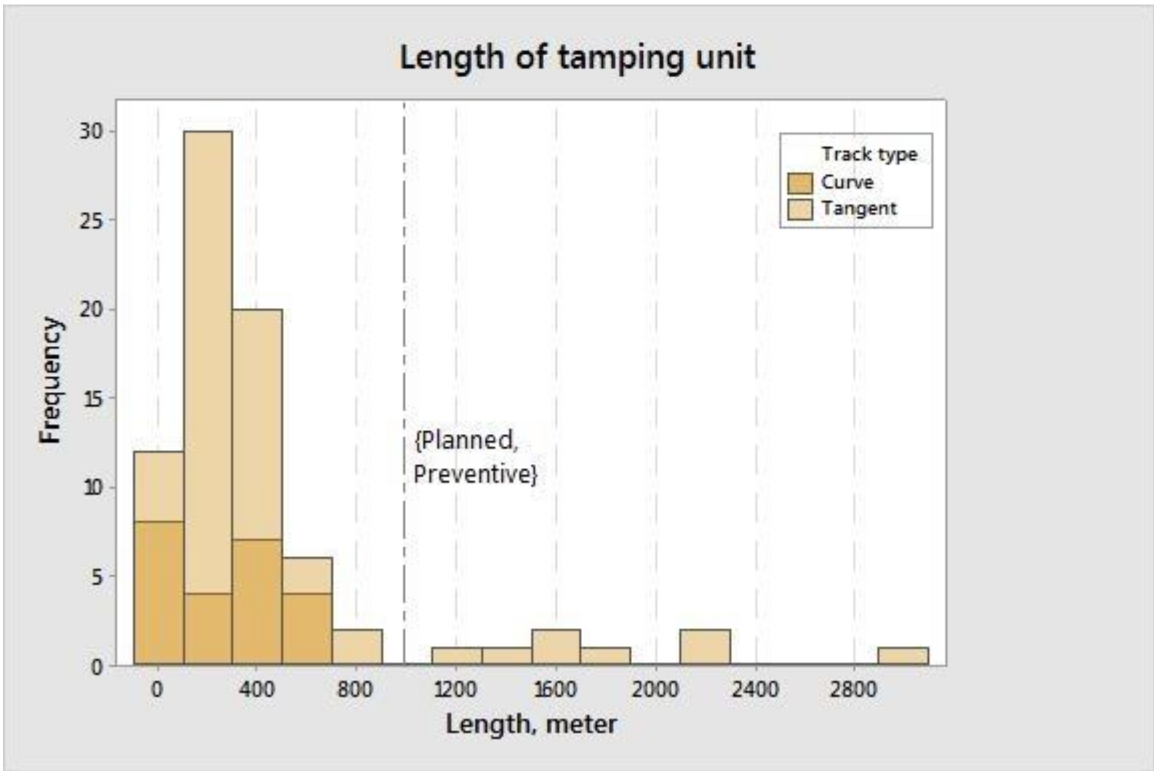


Fig. 5.7 Summary of a tamping length over 16 year's observation

Table 5.2 Profile of planned and preventive tamping works

Tamping date	KM start	KM end	Length (meter)
2001-11	180.688	182.930	2242
2003-05	179.987	182.897	2910
2004-04	184.301	185.549	1248
2009-07	175.590	177.299	1709
2009-11	178.956	180.498	1542
2018-04	175.505	177.799	2294

2018-04	178.956	180.588	1632
2018-04	183.778	185.199	1421

Resulting samples from each dataset are illustrated in an individual value plot in Fig. 5.8. Here, no identical pattern of tamping effectives appears across six datasets. Nevertheless, no further analysis had been assigned to a dataset prepared from a tamping date of 2001-11 due to almost no variation in samples (i.e. only two values of survival times).

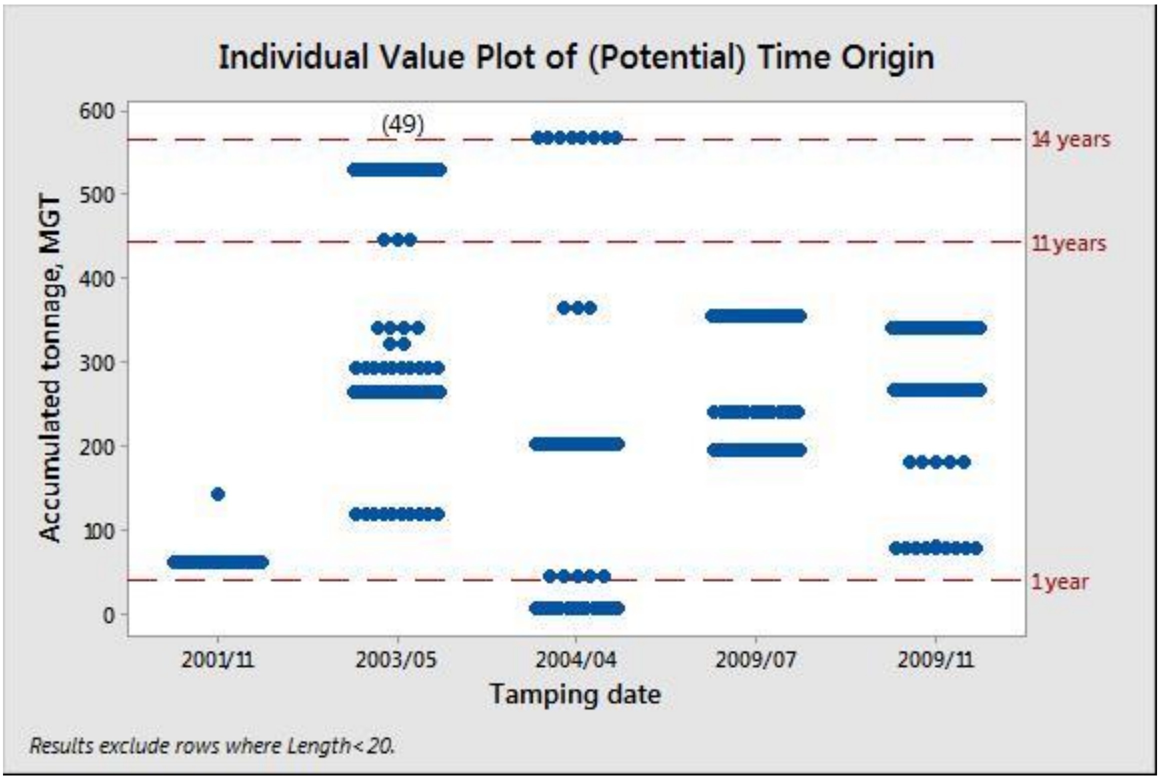


Fig. 5.8 Potential candidates of time origin for survival analysis

5.2.2 Non-parametric function

The Kaplan-Meier survival curve of tamping-related survival data for four datasets is shown in Fig. 5.9. The survival curves were obtained using MINITAB 18. The only censored data used in the KM survival estimate was from the 2003-05 dataset. Out of 117 subjects, 49 subjects — from KM180+919 to KM182+148 — had not (yet) experienced the event by the last observed survival time (approximately at 570 MGT), which is then assigned as a censoring point for the 2003-05 dataset. Event time of the subjects is only known somewhere beyond the right side of the censoring point.

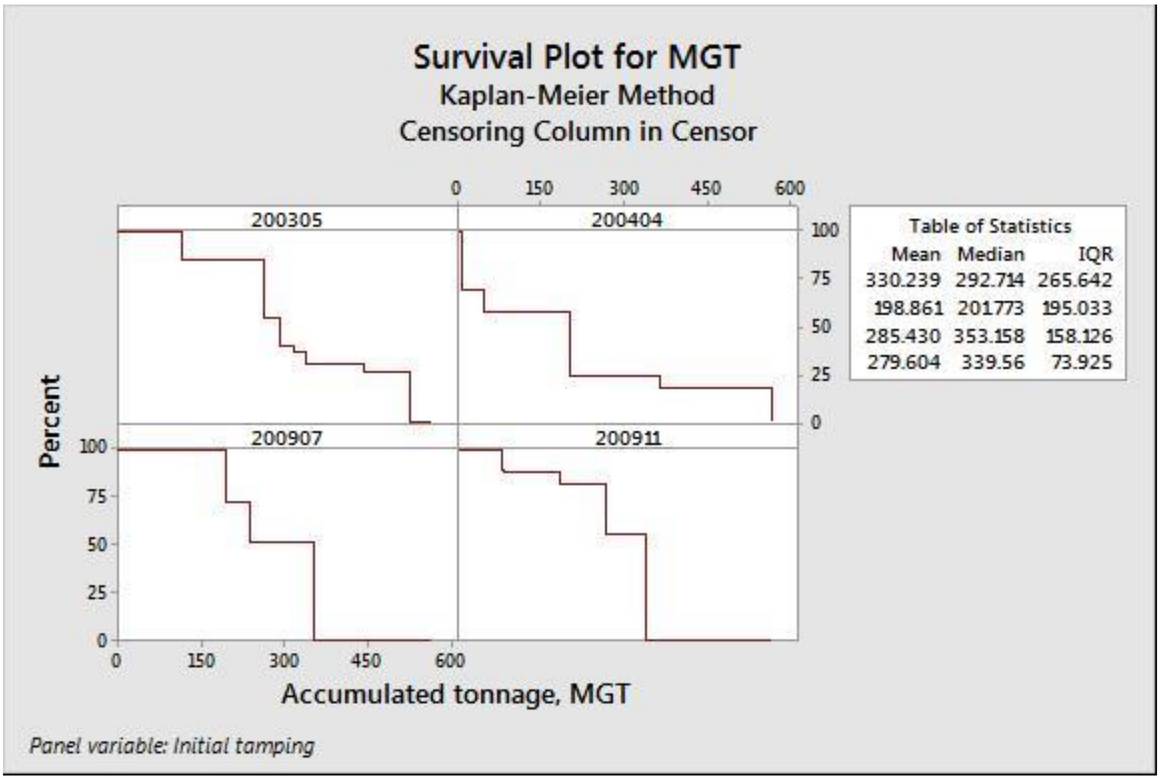


Fig. 5.9 Survivor curves

In all cases except in the 2004-04 dataset, the participating subjects were 100% operationally reliable from their time origin to at least their accumulated tonnage reached 70 MGT (about 1.75 years). The steep decline which is apparent in the early service time after maintenance work appears in the 2004-04 dataset may indicate that several track segments were ineffectively maintained. A poor calibration in a tamping machine is likely a leading cause of why the affected tracks segments {184.990, 185.350} received a subsequent tamping earlier than other subjects (intra and inter case). Meanwhile, for both tamping operations in the year 2009, 50% of subjects remain in a ‘non-failure’ state by 330 MGT (equivalent to 11 years) post tamping. An exact value of mean and median of survival times for each case are shown in the ‘Table of Statistics’ on the upper right corner of Fig. 5.9.

5.2.3 An approximate parametric distribution

On some occasions (e.g. in deterioration analysis), researchers make an assumption on a functional form of the distribution of the survivor curve. Results of non-parametric analysis (e.g. smoothed hazard function) are basically referred for first information when assuming the shape the distribution takes. This is due to the fact that all parametric distribution has a corresponding hazard function. The normal, exponential, gamma, Gompertz and Weibull are often used distribution function in a survival (and reliability) analysis. Obviously, the characteristics of the variable of interest appear in a format that is readable regardless of the parametric function. For example, the failure risk is said to be constant at λ over time if we assume the survival data follow an exponential distribution.

The survivor (tamping effectiveness) function reports the probability of surviving (the time elapsed from the last preventive tamping) beyond time t. The function is simply the reverse cumulative failure function, which focuses on having an event before t. For each dataset in this study, the cumulative failure is plotted as a function of accumulated tonnage: a Weibull distributed failure pattern which its value of shape parameter greater than two yields an *S*-shaped line curve, as shown in Fig. 5.10. In a case of the shape parameter is less than one, a failure rate monotonically increases with time. Table 5.3 shows estimation values of the parameters of distribution.

An analysis on the 200907- and 200911-dataset suggest that 63.2% of track segment on the analysed section will meet a condition of preventive tamping after the accumulated traffic load is in the range of 307 to 313 MGT since the last tamping. Surprisingly, the Weibull’s characteristic life for the 200404-dataset reveals that the same percentage of the population of track segments would receive a tamping maintenance as early as 170 MGT in the accumulated tonnage i.e. an almost one-half faster than the previous value. This could manifest there is a significant difference in the quality of tamping operation.

By applying the values of α_T and β obtained in Fig. 5.10 to the formulation in (O'Connor, Modarres and Ali Mosleh, 2016), their respective mean time to 'failure' (MTTF) and median time to failure (t_{50}) can be calculated, see Table 5.3. For example, the 200404-dataset estimated that a track segment will receive a preventive tamping after its average accumulated traffic load reaches 204 MGT. A higher level of MTTF in MGT (add 130 MGT to the previous value) is however quantified from the 200305-dataset. Not much difference in the MTTF level between the remaining two datasets. On the other hand, a pattern on the t_{50} for a track segment using the available datasets is slightly different from the MTTF case. This indicates that Weibull function for a random variable T is a reasonable approximation.

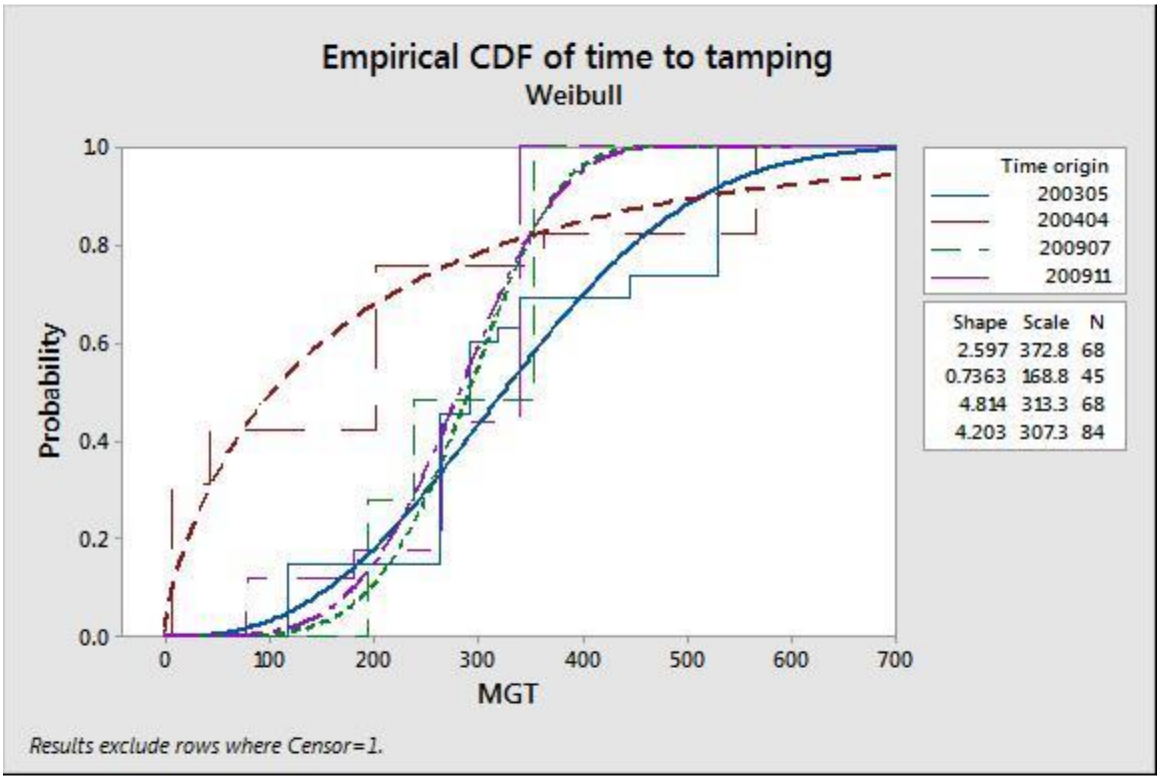


Fig. 5.10 A Weibull distribution to approximately describe observed event data

Table 5.3 Characteristics of Weibull distributions for tamping effectiveness

Dataset	Sample size	α_T (MGT)	β	MTTF	t_{50}	Variance $\left(10^4\right)$
200305	68	372.8	2.597	331.11	323.73	1.875
200907	68	313.3	4.814	287.03	290.33	0.463
200911	84	307.3	4.203	279.33	281.64	0.561
200404	45	168.8	0.7363	204.12	102.61	7.946

As explained in the Section 3.3.3, the parameter values of failure distribution can be used to estimate the corresponding (closed-form) distribution parameters for a deterioration rate of track segment. Results of Fig. 5.10 suggest that a deterioration rate of 20-meter track segment X_{20m} has an inverse Weibull distribution with the scale parameter is

$\alpha_{drt} = \frac{D_{IL} - \phi}{\alpha_F} > 0$ and the shape parameter is $\beta_{drt} = \beta$. Table 5.4 provides, for each dataset, the estimates of the deterioration rate parameters α_{drt} and β_{drt}, and the corresponding first and second moment about zero of the distribution.

Table 5.4 Inverse Weibull parameters for the deterioration rate of track condition.

Dataset	α_T (MGT)	β	α_{drt} (mm/MGT)	β_{drt}	$E[X]$	$E[X^2]$ (10^{-5})
200305	372.8	2.597	0.0054	2.597	0.0033	0.661
200907	313.3	4.814	0.0064	4.814	0.0051	2.382
200911	307.3	4.203	0.0065	4.203	0.0050	2.221
200404*	168.8	0.7363	0.0118	0.7363	-	-

* For $0 < \beta_{drt} \leq 1$, the mean and the variance do not exist (Kundu and Howlader, 2010).

5.2.4 Use survivor functions to obtain a distribution of delay time

There are always detectable defects prior to the point of track failure. The sooner track engineers can detect track defects, the longer time they will have to analyse, plan and perform necessary actions upon an inspected track. For deteriorating assets like railway tracks, an analysis of deterioration path is common in practice for determining at least the distribution function of time defects become visible and time-to-failure distribution (Chattopadhyay and Kumar, 2009; He *et al.*, 2013; Andrade and Teixeira, 2016). A deterioration path (or curve) is a typical two-dimensional graphical representation on how the condition of an asset or asset component deteriorates from an initial point of condition to a defect initiation and then continue to a state or condition (denoted by point “F”) where the object of interest is not capable of performing its function at a specified performance level, as intended by an asset owner. Importantly, a defect detection point (denoted by point "P") is located somewhere between a point of defect initiation and a point "F". Fig. 5.11 illustrates a position of point "P" and "F" on a deterioration path where the *x*-axis and *y*-axis represents a time-in-service for an asset and some measure of performance, rate, condition or suitability for purpose, respectively. In the context of track-geo 'soft' failures, the point "P" and point "F" can be represented by the intersection points between the deterioration path and the alert limit, and intervention limit, respectively.

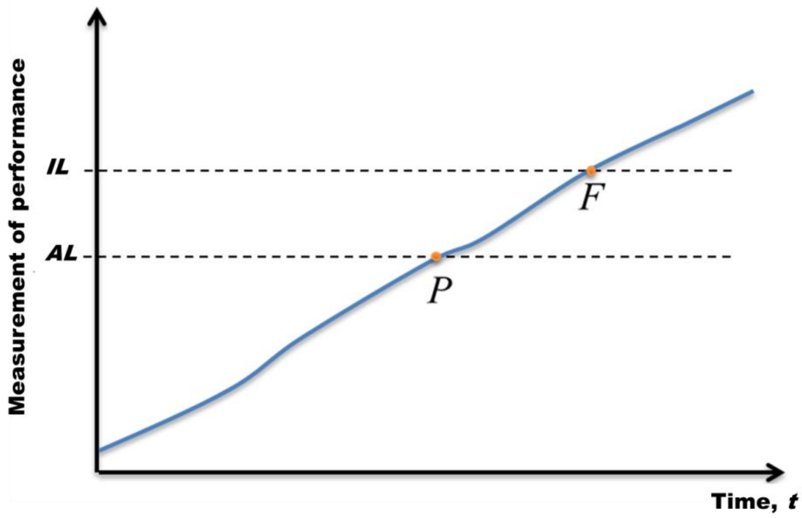


Fig. 5.11 An illustration of the relationship between the deterioration path, points "P" and "F", and thresholds of track geometry defects.

The time period between two events; defects detection and a point where an item is not able to perform its intended function, is termed as delay time. Let a random variable T_d denote the time taken to a point "F" from a point "P". Further, the variable can also be treated as the delay time in the occurrence of failure upon the defect is detected. Hence, T_L is described by the difference of T_F and T_P where T_F and T_P denote the time when the deterioration path reaches a point "P" and "F" from the last restoring point, respectively. The density function of T_L (denoted by $f_{T_L}(u)$) can be then obtained from the convolution of the density functions of T_F and $(-T_P)$. Using the convolution formula the $f_{T_L}(u)$ is

$$f_{T_L}(u) = \int_0^\infty f_{T_F}(t) f_{-T_P}(u-t)\,dt = \int_0^\infty f_{T_F}(t) f_{T_P}(t-u)\,dt. \tag{5.1}$$

Given a distribution function of T_F (which is denoted by $F_{T_F}(t)$) for a 'soft' failure occurs at time *t*, it is straightforward to obtain the distribution function of T_P (which is denoted by $F_{T_P}(t)$) (Freitas *et al.*, 2010). Using information in Fig. 5.10, $T_F \sim Weibull(\alpha_T, \beta)$. Through the deterioration analysis, for a path function $\eta(t) = \phi + Xt$ and a transformation $X(T_F;\phi, D,\eta) = (D_{IL} - \phi)/T_F$, the distribution function of X $G_X(x)$ can be derived analytically from $F_{T_F}(t)$. $G_X(x)$ is given as follows:

$$\begin{aligned} G_X(x) &= P(X \le x) = P\left(\tau(T_F;\phi, D_{IL},\eta) \le x\right) = P\left(\frac{D_{IL} - \phi}{T_F} \le x\right) \\ &= P\left(T_F \ge \frac{D_{IL} - \phi}{x}\right) = 1 - F_{T_F}\left(\frac{D_{IL} - \phi}{x}\right) = \exp\left[-\left(\frac{(D_{IL} - \phi)/x}{\alpha_T}\right)^\beta\right] \\ &= \exp\left[-\left(\frac{(D_{IL} - \phi)/\alpha_T}{x}\right)^\beta\right]. \end{aligned} \tag{5.2}$$

To obtain a distribution function of T_P, let $X = (D_{AL} - \phi)/T_P$, that is, when $X \sim invWeibull(D_{IL} - \phi/\alpha_T, \beta)$,then it is possible to define $F_{T_P}(t)$ by

$$F_{T_P}(t) = P(T_P \le t) = P\left(\frac{D_{AL}-\phi}{X} \le t\right) = P\left(X \ge \frac{D_{AL}-\phi}{t}\right) = 1 - G_X\left(\frac{D_{AL}-\phi}{t}\right)$$

$$= 1 - \exp\left[-\left(\frac{(D_{IL}-\phi)/\alpha_T}{(D_{AL}-\phi)/t}\right)^{\beta}\right] = 1 - \exp\left[-\left(\frac{t}{\frac{(D_{AL}-\phi)}{(D_{IL}-\phi)}\alpha_T}\right)^{\beta}\right]. \tag{5.3}$$

In this case, T_P also follows a Weibull distribution with the same shape parameter as $F_{T_F}(t)$ but its scale parameter is $\alpha_P = \left(\frac{D_{AL}-\phi}{D_{IL}-\phi}\right)\alpha_T > 0$ i.e. $T_P \sim Weibull\left(\frac{D_{AL}-\phi}{D_{IL}-\phi}\alpha_T, \beta\right)$. The corresponding rate of occurrence of defect at time *t* can be obtained from following equation:

$$h_P(t) = \frac{\beta}{\alpha_P}\left(\frac{t}{\alpha_P}\right)^{\beta-1}. \tag{5.4}$$

Using the values of α_T and β obtained in Fig. 5.10, the results of parameter estimates of Weibull distribution of T_P are presented in Table 5.5. Fig. 5.12 shows how the values of α_p and β affect characteristics of the function in Eqn. (5.4).

Table 5.5 Weibull parameters for different delay time models.

Dataset	α_p (MGT)	β
200305	176.6	2.597
200907	148.4	4.814
200911	145.6	4.203
200404	79.9	0.7363

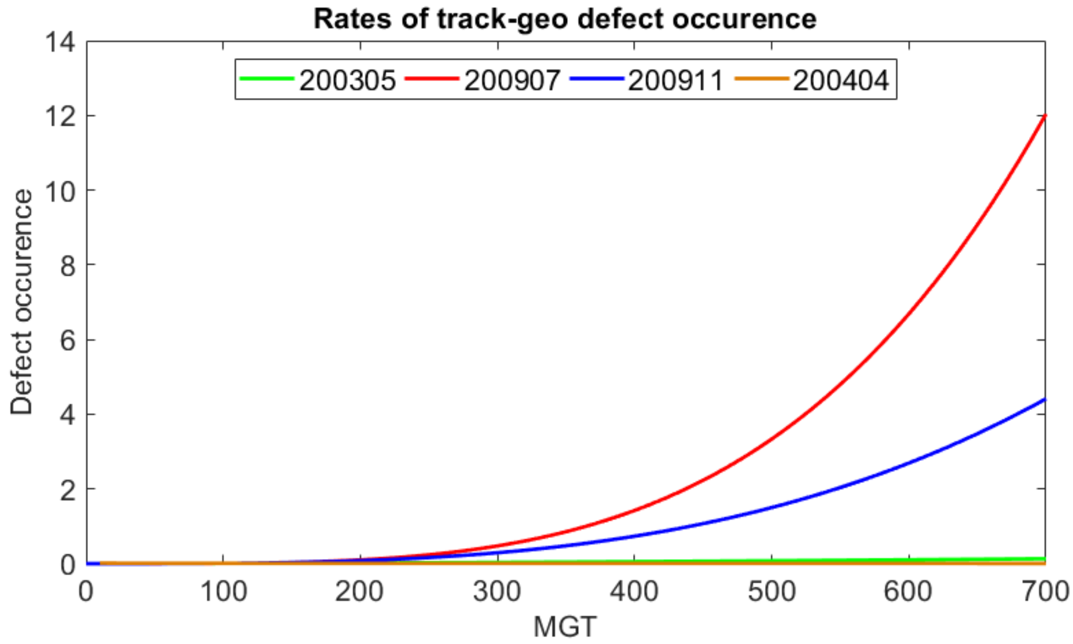


Fig. 5.12 Impacts of Weibull parameters on the occurrence rate of track-geo defects.

5.3 Repair model

If a tamping operation is assumed to be perfect i.e. a complete maintenance is taken, thus, a tamping takes place after the failure of a track segment returns the affected segment to an "as good as new" (AGAN) condition. In the worst scenario, the tamped segment is not in the AGAN condition after the repair, but in the similar condition it was in prior to 'soft' failure. Here, tamping maintenance is assumed to be a minimal repair or "as bad as old" (ABAO). Minimal repair on the track has not changed the failure rate behaviour. However, in the reality of track maintenance, neither repairable model is deemed to be true to every tamping situation. In Khouy et al. (2013) the authors apply the tamping intervention graph (see Fig. 7 in the article) to prepare evidences for the claim. In addition, they also reveal that the majority of tamped track segments were imperfectly maintained (i.e. tamping restores track segments to somewhere between AGAN and ABAO).

Given sequential failure data, a methodology designed by (Coetzee, 1997), as illustrated in Fig. 5.13 was applied to identify an appropriate failure model for track segments in an analysed track section.

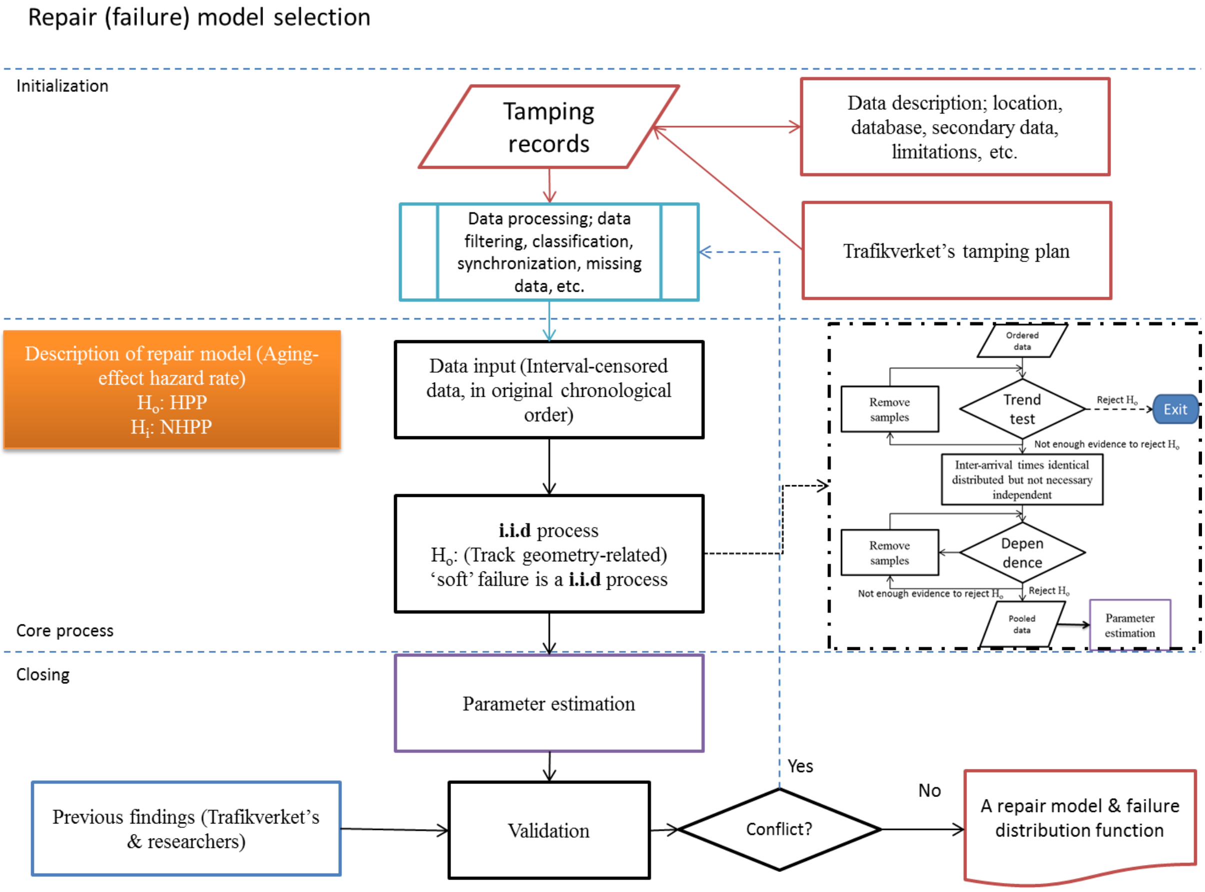


Fig. 5.13 Necessary steps to be undertaken to determine type of failure/repair model.

5.3.1 Results

Decreasing, constant or increasing trend in successive failure rates of track segments can be determined using an event plot. Basically, the plot displays a plot of events (failures and retirements) for the system observed.

From Fig. 5.14, failure data show no trend in failure rates for track segments under cluster 2–10. An assumption that failure data are independently and identically distributed may be imposed (i.e., the time between failures is independent of the age of the track segment). As a result, these track segments were excluded in the next steps of the data analysis. For track segments belongs to cluster 13–16, however, a trend is clearly observed. Reliability of these segments is decreasing over time and it might due to an 'ageing' process taking place. As a result, a NHPP model instead of the renewal process model will be used to model the failure time of the track segments.

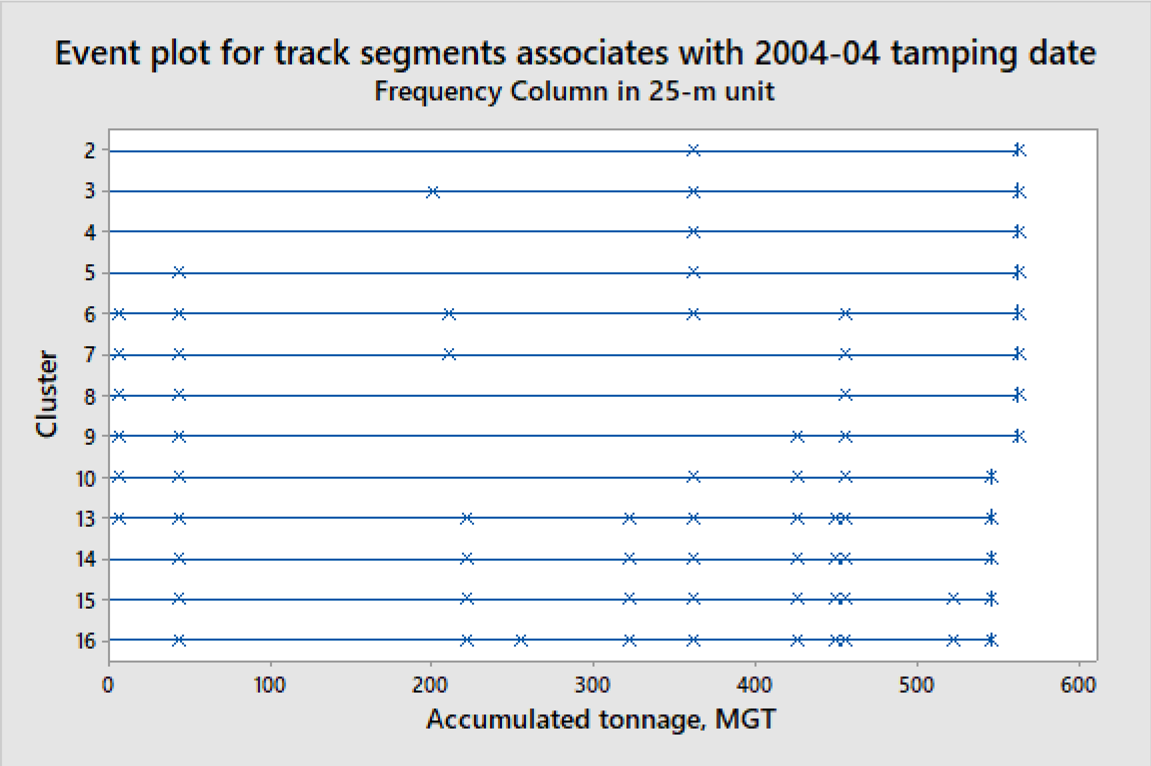


Fig. 5.14 Trend testing using an event plot.

Visual evidence on how well the NHPP model fits the failure data can be gathered from a total-time-on-test (TTT) plot. The TTT plot which is a plot of total time on test statistic against the scaled failure number is useful to identify the presence of non-monotonic trends in failure data.

From Fig. 5.15, an increasing trend in the intensity of failures is observed from cluster 14, 15 and 16. Hence, it is possible to say that track segments in these clusters are deteriorating over time. Also, it might indicates that effects of tamping maintenance on these particular track segments towards ABAO class.

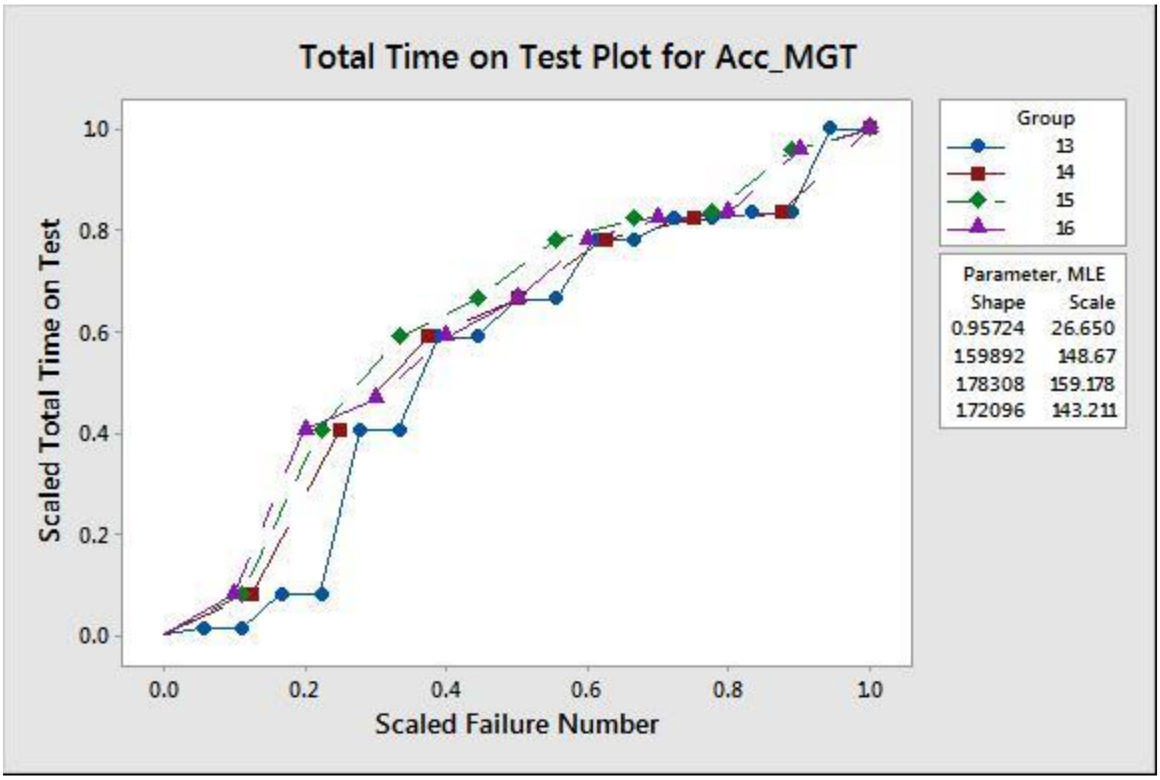


Fig. 5.15 Deteriorating track segments

5.4 Summary

In this chapter, tamping maintenance records obtained from the *Travikverket* were statistically analysed to infer prior knowledge about i) track geometry-related condition deterioration and ii) parameters of failure process of rail tracks that will be used in Chapters 6 and 7, respectively.

The results of the data analysis presented in Section 5.2.3 suggest that a deterioration rate of track quality has an inverse Weibull distribution. This finding will serve as basis in the Bayes linear method to assess uncertainty propagation in the deterioration rate of track quality (details in Section 6.5). This method is embedded in a prediction model iNARXNN for post-processing to check for inconsistencies between predicted and past track measurements. The results of the data analysis are also included in the development of a value model presented in Chapter 7. The denominator in the value model requires an expected number of failures for risk calculation. Results in Section 5.3.1 provide evidence that the expected value of failures should be calculated from the NHPP model, not the HPP model.

CHAPTER 6 A RESPONSE MODEL FOR DISRUPTED INSPECTION PLANS

This chapter explains how a method of fusion involving artificial neural network and failure models (the results of data analysis in Chapter 5) can be formulated as a prediction model to provide a disrupted inspection plan with the capacity to absorb an initial impact of disruptions; paving its way towards resilience. Besides, the step-by-step procedure application of a prediction model is provided to generate 'missing' track measurement based on values of observable track-related quantities of previous times for straight track segments. The prediction models were developed using MATLAB software (version R2010b) to predict the 'missing' track measurements. Results of the prediction model presented in this chapter serve a measure for performance evaluation of iNARXNN in quick response to disruptions in a track inspection plan.

6.1 Introduction

When disruption to the track inspection plan incurs, a track engineer can expect some delays in receiving of recent track measurements (and inspection data) whereby it may affect the planning of maintenance activities. Leaving tracks unattended longer than necessary increases the risk of late defect detection. As a result, unplanned repair instead of a planned function must be performed. Certainly, this has a direct impact on track maintenance expenses (Stenström *et al.*, 2016). In an extreme situation, ineffective track inspection could become a causal factor of train derailment (Rail Accident Investigation Branch, 2014). One possible way to respond to a disruptive event in track inspection is by finding an appropriate response action to disrupted track measurement. The term 'disrupted' is used to emphasise a direct loss attributed to disruption in TIP. As explained in Chapter 1 and 4, an innovative framework of disruption management has been employed to identify an appropriate response action for disrupted TIP.

Here, the motivation of employing the DM framework is that a response action can impose the minimum change possible in all aspects of the affected process or system of interest. A smooth transition from pre- to post-disruption is required because disruption is temporary in nature. To satisfy this requirement, a track engineer must know and preserve as much as possible the integrity of the ‘disrupted’ component or unit of the process (Jespersen-Groth *et al.*, 2009). In this study, track measurement data serves as representative information of ‘disrupted’ component.

A railway track is an example of a complex and continuous asset; thus, its length is a critical parameter in the track maintenance equation. This description means that from a statistical point of view, track measurements are spatially correlated within a certain interval. This dependency could introduce a bias in an analysis of data model identification. Similarly, periodic track measurements provide a platform to trace an evolution of track geometry changes over tonnage (or time). Taking the results of statistical analysis of track measurement data into consideration, both temporal and spatial effects should be considered when deciding upon an appropriate response action for disrupted track measurements. In line with the requirements above, an application of artificial intelligence to derive adaptive responses to a disrupted TIP appears to be promising. This research therefore proposes a novel artificial intelligence framework to predict track measurements. Briefly stated, the new method first constructs a statistical model for track measurement data before being used to generate the missing track measurement data.

6.2 Description of data generation model

The main goal of developing a prediction model is to provide an approximation to the actual physical process or system and predict its outputs (Menezes and Barreto, 2008). When a thorough nonlinear relationship between input covariates and the model output of the process is not necessary, statistical models are more effective for the creation of prediction models. This method uses a simplified mathematical function to describe dynamic interactions in the input/output variables (Pisoni *et al.*, 2009). One of the most convenient model structures for prediction purposes is the nonlinear autoregressive model with exogenous variables (NARX). In this model, a nonlinear mapping function defines the complex relationship between the targeted (output) and external (input) variable. The function takes past values of the input and output variables to generate the current value of the output variable(s). In relation to the method used to obtain the mapping function, the iNARXNN model could have different shapes of function. Among various methods such as Box-Jenkins model, neural networks have been proven to be universal function approximators, that is, it can approximate complex functions with arbitrary accuracy conditioned by the number of training epochs and quality of data (Hornik, Stinchcombe and White, 1989).

Neural networks have become popular in simulation and prediction for two reasons; they handle the nonlinear relationship between input-output variables implicitly (i.e. without requiring an in-depth knowledge or making assumptions about the problem under investigation), and they generalise well against unseen data. However, the beneficial features of a neural network would disappear if it is trained using a small set of training data. A lack of data can occur in the case of disrupted track inspection, particularly for track sections associated with low measurement frequency which thereafter are prone to train derailment

(Liu, Saat and Barkan, 2011). Ones should note that only measurements recorded from the last tamping maintenance (i.e. a restoration point) will be used for modelling. Any track measurements gathered outside from the interval is declared obsolete because the state of an analysed track has been reset by maintenance work. This restriction placed on data preparation for the neural network is unique for TIP.

Overestimation is one of the side effects of overfitted data. To mitigate the overestimation risk in prediction models, a neural network utilises Bayesian regularisation. Bayesian regularisation limits the size of the network parameters to make the network produce a smoother response. Apart from its strong generalisation properties, a regularised neural network also takes less computational time for model training. The model validation process, which is scaled as $\mathrm{O}\left(N^2\right)$ where N is the number of data points, is no longer necessary (Burden and Winkler, 2008). The time-saving benefit appears in the case of large quantities of data.

To date, several predictive models have been proposed (Liu, Xu and Wang, 2010; Xu *et al.*, 2011, 2015), but all models are not designed with respect to the unique nature of disruptions to track inspection. To that end, an integrated formulation of a Bayesian regularised neural network and the NARX model best fits to generate AI-based data for track measurements in the presence of unexpected events.

6.2.1 Formulation of a prediction model

During a track geometry measurement, the TRC travels at a track speed of between 70 and $120\,\mathrm{km/h}$ to measure seven track geometrical parameters at equally-spaced track points (positions) over a single track code. One track code spans K kilometres of track and is thus divided into several non-overlapping track segments. For a track segment *S* with a length of m metres, there can thus be a sequence of measurement points, denoted as $x_i^{S_j}$ where $i=1,2,\ldots,l_p\left(=m/r\right)$ and $j=1,2,\ldots,l_s\left(=K/m\right)$. Here, r denotes the distance (in meters) between two adjacent measurement points, l_p the number measurement points along a track segment and l_s the number of track segments over a track section. When the n^{th} track measurement is carried out in track segment S_j , each $x_i^{S_j}$ for $\forall i \in S_j$ is assigned with an amount of deviation of *G* track geometrical parameters. Here, it is expressed as $y_k^n\left(x_i^{S_j}\right)$ wherein $k=1,2,\ldots,G$.

When a disruption incurs, for example in the TRC, before the $\left(n+1\right)^{th}$ track measurement, AI-based data is generated for $y_k\left(x_i^{S_j}\right)$. Data generation is an immediate response to avoid time delay when supplying recent track measurements for an assessment of track condition. This action, in turn, reduces the probability of dealing with track geometry-related unplanned maintenance due to late defect detection on the affected track code. For data generation purpose, an identified track geometrical parameter is defined as an autoregressive variable in a NARX model and is denoted as $\tilde{y}_k^{n+1}\left(x_i^{S_j}\right)$. The proposed NARX model also takes $\left(p+1\right)$ external variables $u_l\left(x_i^{S_j}\right)$ for $l=1,\ldots,p,p+1$ as model inputs. Therefore, the corresponding NARX can be expressed as

$$\tilde{y}_k^{n+1}\left(x_i^{S_j}\right)= f\left(\begin{array}{l}\tilde{y}_k^{n+1}\left(x_{i-1}^{S_j}\right),\tilde{y}_k^{n+1}\left(x_{i-2}^{S_j}\right),\ldots,\tilde{y}_k^{n+1}\left(x_{i-d_1}^{S_j}\right),u_1\left(x_{i-1}^{S_j}\right),u_1\left(x_{i-2}^{S_j}\right),\ldots,u_1\left(x_{i-d_2}^{S_j}\right),\\ u_p\left(x_{i-1}^{S_j}\right),u_p\left(x_{i-2}^{S_j}\right),\ldots,u_p\left(x_{i-d2}^{S_j}\right),u_{p+1}\left(x_{i-1}^{S_j}\right),u_{p+1}\left(x_{i-2}^{S_j}\right),\ldots,u_{p+1}\left(x_{i-d_2}^{S_j}\right)\end{array}\right) \tag{6.1}$$

where $f(\cdot)$ is an unknown nonlinear mapping function of $d = d_1 + (p+1)d_2'$s previous known outputs. Each of the p external variables is defined by $y_{h\neq k}^{n+1}\left(x_i^{S_j}\right); h \in \{1, 2, \ldots, G\}$. At this point, track geometry parameter h is assumed to be irregular, and in terms of the targeted track geometry, k is non-independent. The trade-off analysis between model complexity and model performance at the end of the model development process is used to justify these assumptions.

Generally, any external variable can be excluded from the NARX model if its contribution to the model performance comes at the expense of reducing model simplicity. On the other hand, the term '1' is added to account for the current size of change in the specified track geometrical parameter. The amount creates the difference between $y_k^n\left(x_i^{S_j}\right)$ and $y_k^{n-1}\left(x_i^{S_j}\right)$.

6.2.2 Input-output variable(s)

The selected track geometrical parameter is defined as an autoregressive variable, while the other parameters act as external inputs in the NARX model. Here, a longitudinal track level is appointed as the disrupted track measurement. Longitudinal levelling is among many track geometrical parameters measured by a track recording car and has been acknowledged as the leading indicator for track tamping maintenance decisions (Khouy, 2013). This establishment is due to the fact that in tangent tracks track irregularities in the vertical direction evolve faster than other track geometry defects. Note that the longitudinal level is assessed for both left and right rails and this requirement has imposed two nodes in the output layer of our NARXNN model.

The primary motivation for the use of external data series in the proposed data generation model is stemmed from unparalleled characteristics of the response action we seek for a disrupted track measurement. Data integration was performed to the past and present measurement values of other track geometrical parameters, e.g., alignment and gauge, together with the past values of the research object to forecast the size the research object would evolve in both spatial and time for a short period of service time. A fusion method of integrated NARX and neural network (denoted as iNARXNN) is expected to reveal nonlinear relations between these geometrical parameters over a span track length whose variations in track stiffness are undisclosed.

A rationale behind the use of track alignment in the iNARXNN model is its influences over vertical track forces as well as the longitudinal level (Karis, 2009). A longitudinal track level and track alignment have been combined in various ways to manage track geometry problems (Haigermoser *et al.*, 2015). For example, Soleimanmeigouni et al. (2018) established such level of relationship between these parameters in the long-term track deterioration process. A similar relationship might exist in the short-term prediction model, but direct use of this relationship might be inappropriate due to a different set of model constraints. A key distinction between short and long term track prediction model is the validity of track measurements (or track quality indices) on the aspect of track deterioration where models in the latter category describes the gradual deterioration of track whereas a sudden shift/abrupt changes in track deterioration is the former model's focus (Xu *et al.*, 2015; Osman and Kaewunruen, 2018). For short-term prediction model, the tamping effect is only valid to the first track measurement after a maintenance work covering the analysed track segment. Nevertheless, nonlinear interactions between track alignment and longitudinal level would be worthwhile to be investigated on the context of disruption management. Similar to longitudinal level, a deviation of track alignment is also measured and analysed for left and right rails simultaneously. Apart from track geometries, track gauge measurement is used as an exogenous variable in our iNARXNN model. Variations in track geometries and gauge can lead to large lateral wheel and axle forces, resulting in derailment or damage to the structure of the track (Choi *et al.*, 2013).

As explained earlier, the prediction model in this study is built after the completion of n successive track measurement runs. This mechanism means that n previous track measurements are accessible. Hence, a series of deterioration rates of a longitudinal level is defined as the final external variable for the iNARXNN model. The rate of deterioration of longitudinal level is simply the difference of longitudinal level at two inspection times divided by an inspection interval τ i.e. $\left(y_k^n - y_k^{n-1}\right)/\tau$.

To unify with the neural network's terminology, this research uses input/output data series term to denote track geometrical parameter over an analysed track segment applied in the data generation model.

6.2.3 Application of neural network on the NARX model

A neural network is a computational paradigm inspired by the structure of biological neural networks and their way of encoding and solving problems. The fundamental NN model is adjusted according to a user’s motive and resources available and is trained systematically with filtered input-output data to solve the problem under investigation.

In this study, a NN is applied to obtain the nonlinear function $f(\cdot)$ in Eqn. (6.1). This requires an appropriate network configuration in line with the model description. The NARXNN architecture in Fig. 6.1 shows an input layer including q recurrent unit with p+1 neurons, a hidden layer and an output layer with q neurons describing the dynamic relationship between input and output variables in Eqn.(6.1). The lines represent weighted connections, and the squares represent bias thresholding nodes. A two-line circle in the hidden layer is the lag/delay element for an input variable. Here, the exact value of both d_1 and d_2 is unknown and will be determined based on model training result.

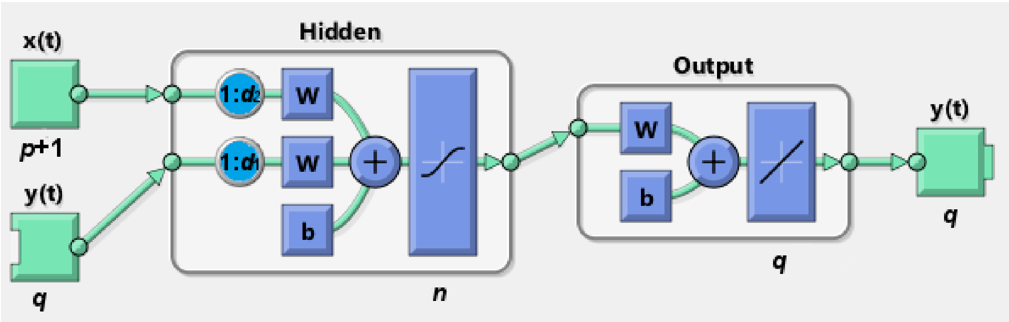


Fig. 6.1 An overview of NARXNN architecture

Prior to the NN training, a configuration of the network must be attained. The number of nodes in the input and output layer, respectively, reflects the number of input and output variable(s) of the system of interest. In case of a NARXNN model, a user also has to initiate the value of a delay unit in model output and external series. Apart from that, the number of hidden nodes is generally unknown and would be iteratively adjusted during the network training.

6.2.4 Mitigation of risk of overfitting when using NN

When an NN is trained using a small training data set, the resulting NN-based model may perform poorly in any instances not included in the training set. The NN is likely overfitted to the training data, significantly reducing the generalisability of the model. Also, the network model tends to generate large errors when new instances are presented. Applying this limitation to data generation, the proposed NARXNN might have a limited capacity to identify sudden shifts in the evolution of track condition over accumulated tonnage. To address this issue and improve the model's generalisation, a regularisation method was adopted in the NN model training. The central idea of the method was to use prior knowledge about the model of interest (e.g., NARX model) in the learning process of NN to decrease the model's complexity (e.g., it has too many nodes in hidden layers and network parameters) and to decrease the model's variance (Burden and Winkler, 2008). In the case of the NARX model, users always provide a number of external inputs (e.g., data series) that can be used to assign an initial value for the input and output delays of the neural network structure. Since the size of the data for each external input was also available, the limitations of computational resources for a model to achieve a desired level of performance (e.g., the prediction error) could be addressed at an early stage of the training network. By incorporating prior knowledge into the neural networks, it allows users to simultaneously achieve a prediction model with low variance and low bias. Overall less complex models typically give an excellent performance with unseen data (test set). Aside from demonstrating good generalisation properties, training the neural network using the regularisation method is less time consuming, as a model validation step is unnecessary (Burden and Winkler, 2008; Yue, Songzheng and Tianshi, 2011). With a validation process scaled as $\mathrm{O}\left(N^2\right)$ where N is the size of data, the benefits of the regularisation method become apparent in cases of large data sets (Burden and Winkler, 2008).

To obtain the best combination of weighting and bias values, producing a prediction model that generalises well, the network is trained on the selected architecture and training set to optimise the following performance function:

$$wMSE = \gamma E_W + (1-\gamma) E_D \tag{6.2}$$

where γ is the regularization parameter. E_W and E_D represent the mean of the sum of squares of the network weights and biases and the mean sum of squared of network errors, respectively. In the stated performance function that is a typical error function with an additional term, the training forces the NN to utilise small weights and biases, i.e., low E_W value. Burden an Winkler (2008) and Piotrowski and Napiorkowski (2013) stated that the successful application of NN to a problem requires a proper choice of learning algorithm; thus, the Bayesian regularisation technique is recommended if the data is small and prone to overfitting. Several assumptions must be made before any training when the Bayesian regularisation is used. The network weights and training data are considered random variables with a Gaussian prior distribution. The prior probability over the weight is then updated by incorporating training data according to Bayes' rule. The optimal weights of the network are then determined when the posterior probability is maximised; this is equivalent to minimising the error function in Eqn. (6.2).

Though NN can approximate any real-valued continuous functions its success is highly dependent upon a proper selection of NN's components such as data preparation and pre-processing, network architecture, model assumptions and constraints, objective function and post-processing steps. In a particular case of disrupted TIP, this study has highlighted the importance of good NN configuration to model dynamics links between track geometrical parameters.

6.3 Network training

6.3.1 Track measurements (Raw data)

Track measurements from track inspection cars for a 1-kilometer long track section somewhere in the Kilsmo-Palsboda line were used to demonstrate how an AI-based response model works. The measurements were provided by Trafikverket in a *MAT-File* format. There are seven data structures representing track geometrical parameters namely *z_left1* (longitudinal level of left rail), *z_right1* (longitudinal level of right rail), *y_left1* (alignment of left rail), *y_right1* (alignment of right rail), *cl* (cross level), *twist3m* (3-meter twist) and *G* (gauge). Notice that besides track geometrical parameters, the data file has also geographic information related to each sample (labelled as *xcoord*) and provides information about track curvature (labelled as *curv*). Due to data privacy this study cannot disclose the location of inspected track segments. Therefore, values of *xcoord* were transformed into a series of integer; starts with 1 and ends at $l_p\left(=1000/r\right)$. Also, values in *curv* were processed to filter out all non-tangents track segments. Fig. 6.2 shows two pairs of longitudinal level measurements recorded at two consecutive time points (inspection dates) where each pair of data series belongs to the same rail side of an analysed track section.

For a one-kilometre length of track section there will be a sequence of 4000 track (measurement) points $x_i; i=1,2,\ldots,4000$ if a distance between two adjacent points is (or approximately) $r=0.25\text{ m}$ i.e. four sample points per metre. The sampling interval depends on the sampling rate employed by a track recording car during track measurement. Considering that seven (or more) track geometry measurements can be stored in an array for each x_i in every inspection run, a data cube representation was used on the collected measurements to better visualise spatial and temporal variations of track measurements along the track. A data cube for the collected data is shown in Fig. 6.3. The figure is 3-dimensional representation with each cell $\left(x_i,\mathbf{y},\mathbf{d}\right)$ of the cube representing a combination of values from track position, track geometry measurements and inspection dates.

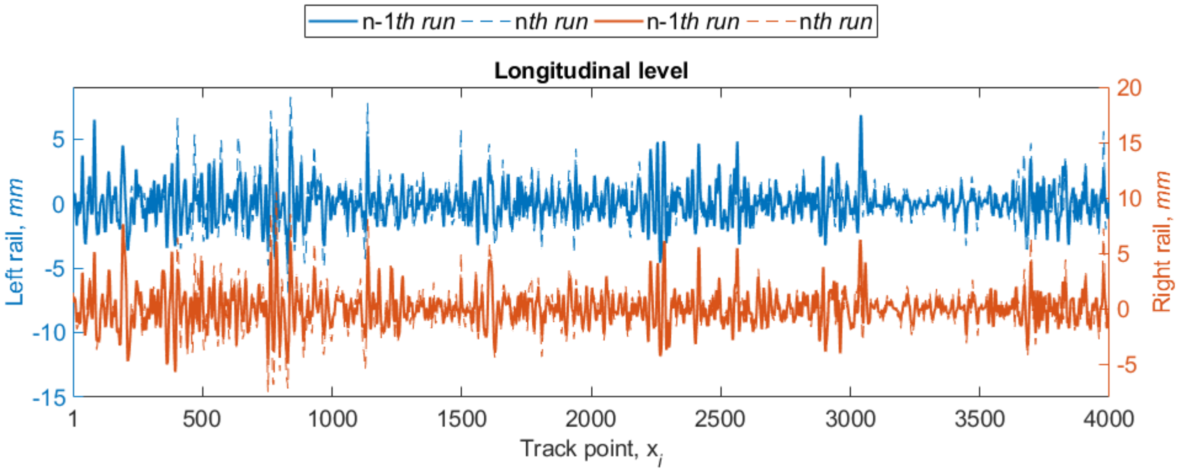


Fig. 6.2 Two consecutive runs of track geometry measurement over the 1-kilometre long track. Longitudinal level is measured at every 0.25 meters and this generates 4000 track points as labelled in the track point-axis.

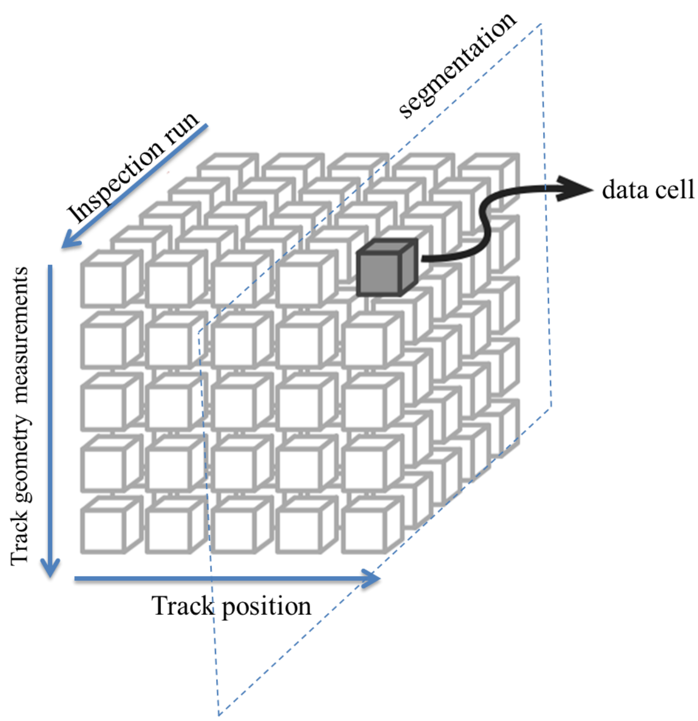


Fig. 6.3 Data cube representation for a collection of track measurements

6.3.2 Data preparation

Sixty segments of 50-meter long track were prepared to demonstrate how NARX-NN networks should be trained and further to determine their best network configuration. The track segments are attributed to the output of the 30%-overlapping segmentation algorithm applied on the data cube (displayed in Fig. 6.3). Notations $S_i; i = 1, 2, \ldots, l_s\ (= 60)$ are used to differentiate the analysed segments. Using the descriptions of a track segment in Section 6.2.1 where $l_p = 200$ for every track segment, the sample size for learning the iNARXNN model given in Eqn. (6.1) was 200. Each sample stored 15 arrays of values, where each array consisted the measurement data of seven track geometrical parameters. The 200 samples were divided into 80% for the training and 20% for the testing of the iNARXNN.

Since a neural network is driven by data, this study applied machine learning tools, such as clustering in the proposed model, to generate a quality training-testing data set, i.e., the track segments that best captured the local characteristics of the track longitudinal level at the selected track section. By identifying the outlying track segments with respect to the entire set of the 60 track segments, only a few iNARXNN models need to be developed to predict the "missing" longitudinal level over a track section. Additionally, the predicted values of the longitudinal level for other track segments that did not participate in the learning process of the iNARXNN could be generated from one of the developed models. This strategy manifests the potential benefit of promoting a high degree of generalisation in the iNARXNN model, as presented in Section 6.2.4. It also uses less computational resources than a full-60 iNARXNN model; thus, one track segment means using one prediction model.

Selected track segments were expected to exhibit high level of dissimilarity, which can be observed by evaluating the statistical feature measures extracted from the track measurements data. This study used root mean square (RMS) and the measured values' kurtosis of the left and right rails' longitudinal level to extract the numerical characteristics of each track segment. These statistical features are widely used statistical indicators in time (and spatial) domain characteristics analyses (Leturiondo *et al.*, 2015). While kurtosis is useful to detect peakedness in measurement data, RMS is particularly useful to measure the magnitude of measurement data. For a track segment S_j consists of $\mathbf{x} = (x_1, x_2, \ldots x_n)$ where n is the number of measurement points, the RMS and kurtosis of the k-th track measurements for S_j were calculated with Eqn. (6.3) and Eqn. (6.4), respectively.

$$RMS\left(y_k\left(\mathbf{x}^{S_j}\right)\right) = \sqrt{\frac{\sum_{i=1}^{n}\left(y_k\left(x_i\right)\right)^2}{n}} \tag{6.3}$$

$$Ku\left(y_k\left(\mathbf{x}^{S_j}\right)\right) = \frac{1}{n\left(RMS\left(y_k\left(\mathbf{x}^{S_j}\right)\right)\right)^4}\sum_{i=1}^{n}\left(y_k\left(x_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(y_k\left(x_i\right)\right)^2\right)^4 \tag{6.4}$$

A scatter plot of the RMS and kurtosis of the longitudinal level for both the left and right rails is shown in Fig. 6.4. The dots in the plot correspond to all pairs of $\left(RMS\left(y_k\left(\mathbf{x}^{S_j}\right)\right), Ku\left(y_k\left(\mathbf{x}^{S_j}\right)\right)\right)$ for $k \in \{6,7\}$ and for $j = 1, 2, \ldots, n$. As shown in Fig. 6.4, the 60 track segments could be potentially partitioned into two clusters; thus, the within-group-object similarity is minimised and the between-group-object dissimilarity is maximised. To statistically assign each track segment to a prospect cluster, the k-means algorithm, where $k = 2$ was applied on the scatter plot in Fig. 6.4. The algorithm found the centres of k (in this case, the value was two) clusters and iteratively assigned track segments based on a selected dissimilarity measure, which aimed to have an objective function. Given the two sets of longitudinal measurements, $y_k\left(\mathbf{x}^{S_A}\right)$ and $y_k\left(\mathbf{x}^{S_B}\right)$, $A, B \in \{1, 2, \ldots, 60\}$, the dissimilarity between the track segments denoted by $Diss(,)$ was defined as:

$$Diss\left(y_k\left(\mathbf{x}^{S_A}\right), y_k\left(\mathbf{x}^{S_B}\right)\right) = dist_{Euclidean}\left(y_k\left(\mathbf{x}^{S_A}\right), y_k\left(\mathbf{x}^{S_B}\right)\right) \tag{6.5}$$

where $dist_{Euclidean}(\cdot)$ was the Euclidean norm. The Euclidean distance between track segments was calculated using Eqn. (6.6). In Fig. 6.4, while the lines show cluster boundaries, an 'X' marker labelled a cluster centre for two clusters that resulted from *k*-means clustering using built-in functions in Matlab2018. Table 6.1is a summary of the clustering result.

Because track segments S_{29} and S_{51} are the closest track segment to each cluster centroid they were selected for the iNARXNN learning process. The area in the square in Fig. 6.5 illustrates the difference in local characteristics of longitudinal levels between S_{29} and S_{51}.

$$\begin{aligned} &dist_{Euclidean}\left(y_k\left(\mathbf{x}^{S_A}\right), y_k\left(\mathbf{x}^{S_B}\right)\right) \\ &= \sqrt{\left(RMS\left(y_k\left(\mathbf{x}^{S_A}\right)\right) - RMS\left(y_k\left(\mathbf{x}^{S_B}\right)\right)\right)^2 + \left(Ku\left(y_k\left(\mathbf{x}^{S_A}\right)\right) - Ku\left(y_k\left(\mathbf{x}^{S_B}\right)\right)\right)^2} \end{aligned} \tag{6.6}$$

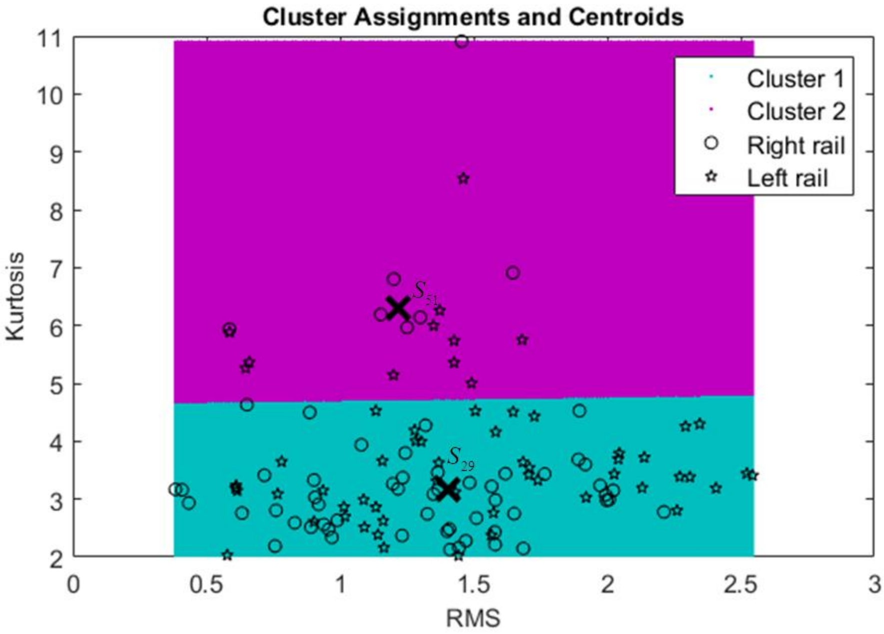


Fig. 6.4 Application of distance-based clustering at the pre-processing step

Table 6.1 Clustering results

Properties	**Cluster 1**	**Cluster 2**
Centroid (RMS, Ku)	(1.4024, 3.1773)	(1.2168, 6.2864)
Number of members	53	7
The nearest member (track segment) to the centroid	S_{29}	S_{51}
Distance between centroids	5.3714	

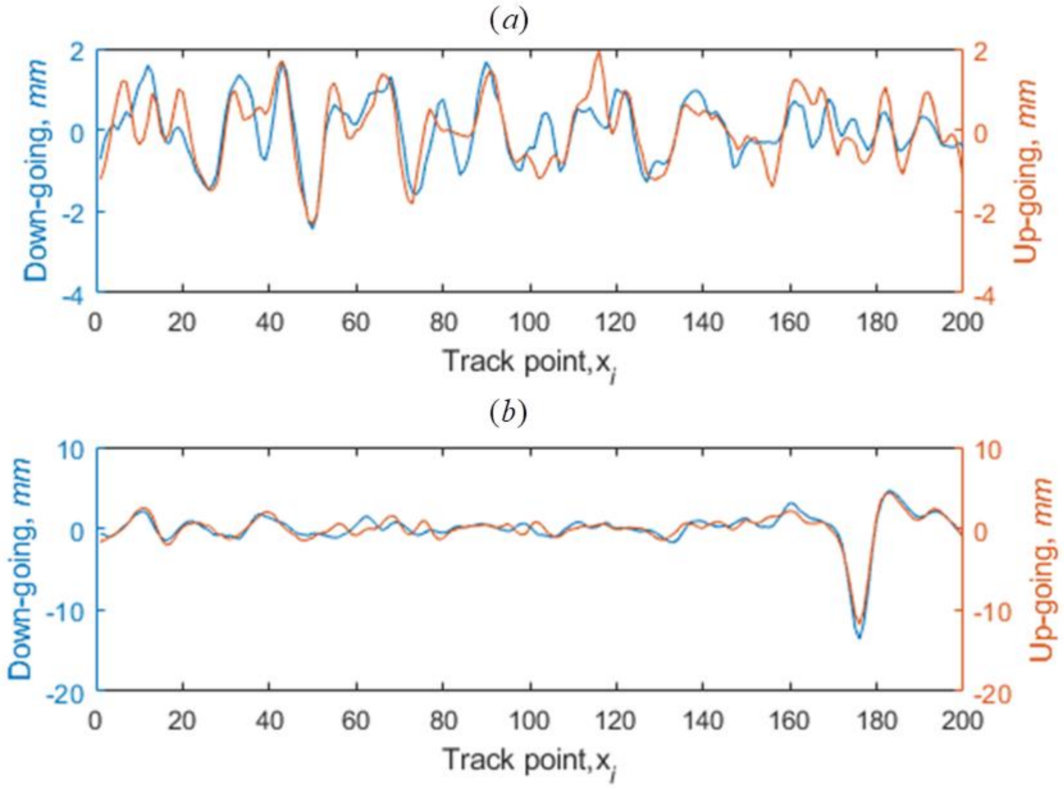


Fig. 6.5 Longitudinal level for both left and right rails; (a) track segment S_{29}, and (b) track segment S_{51}

6.3.3 An identification of affordable network

Training NN is a data-intensive task and there always possibility NN overfit data, which cause, for example, the resulted prediction model does a poor job predicting new data. In the event where time, budget or technical constraints limit the size of training data, a promising way to mitigate the risk of over fitting in NN-based function approximation is by reducing the capacity (complexity) of a NN i.e. avoided oversized network. As stated in (Ref), the capacity of a NN to absorb information is limited by its number of parameters. Thus, training an oversized NN from insufficient data making the model exhibits a poor generalization. Well-tuned parameters make a good balance between training performance and generalization capability and then inhibit the effect of over fitting. A simple decision rule based on a degree of freedom in a NN model is to keep the number of parameters in a network less than the size of training set (Lawrence, Giles and Tsoi, 1997). A mapping technique can be applied on the training set to determine feasible selection(s) of network topology. The notation iNARXNN(d_1, d_2, n) represents a network topology for our iNARXNN model will be used in the remaining texts. An example of mapping output is given in Fig. 6.6 where an integer in each cell represents the number of network parameter in a iNARXNN. Based on the decision rule, only iNARXNN models on the left side of the boundary line (marked with a bold black line) in Fig. 6.6 will be trained to find the best network.

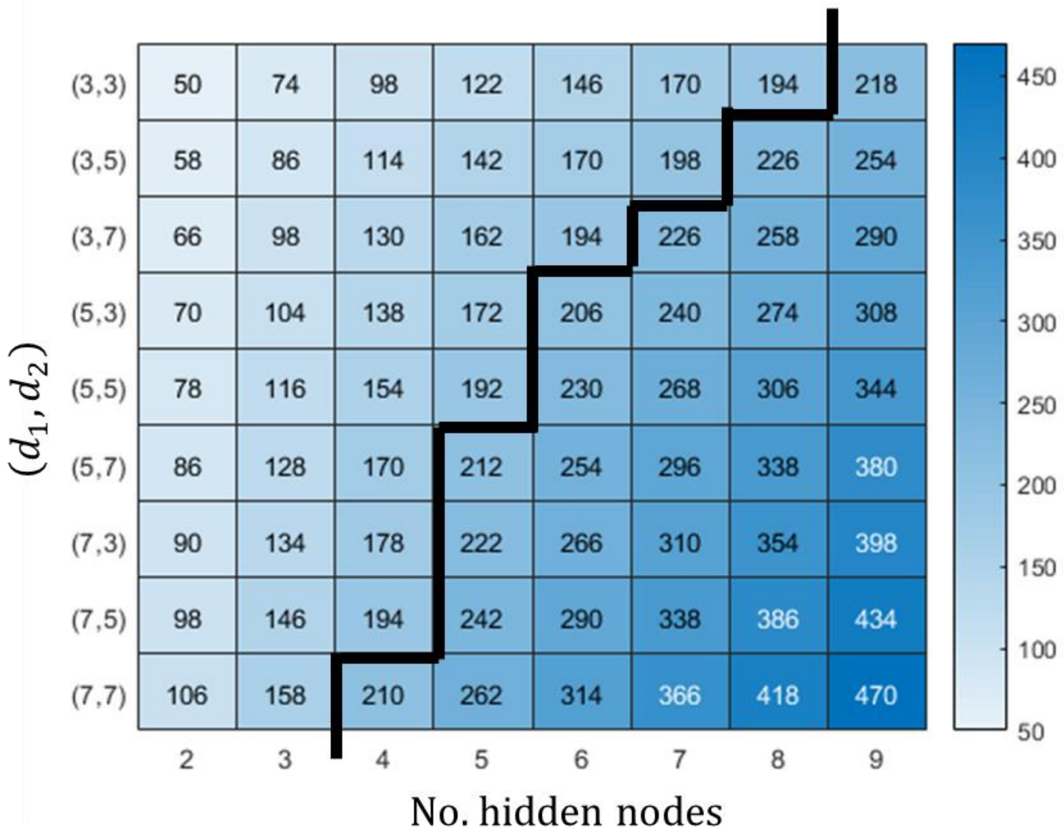


Fig. 6.6 Degree of freedom of iNARXNN for a designated boundary condition imposed on delay units; d_1 and d_2. The number of hidden nodes of the network is restricted to nine corresponds to N=200.

6.3.4 Best performance iNARXNN networks

Selected networks from the previously filtering step were trained with a collection of data series associated with S_{29} and S_{51}. During the training stage, each model received 30 repetitions of simulations with random initialization (Iyer and Rhinehart, 1999).

The average values of the weighted MSE (*w*MSE) over 30 samples are recorded in Fig. 6.7 for all networks. First of all, for each network, as the number of hidden nodes increases the *w*MSE start to decrease, i.e., increases in model performance. The evolution of *w*MSE over *x*-axis occurs steadily in each network. This observation encourages us to further test networks with a large number of hidden nodes.

On the other hand, regardless of the number of hidden nodes configured in the iNARXNN (d_1, d_2, n), we can observe a clear pattern about the effect of the parameter d_2 in the performance of the network. For instance, for networks under iNARXNN$(d_1 = 3, d_2, n)$, the *w*MSE is substantially increased when a network in this group takes additional delay units in the external variables i.e. $d_2 > d_1$. The same pattern also appears for NARX-NN$(5, d_2, n)$. However, iNARXNN$(7, d_2, n)$ experiences an increment in the *w*MSE when $d_2 < d_1$. These two opposite observations highlight the existence of the mixed effects of external variables on the performance of our iNARXNN model. In other words, information between adjacent track points is not necessarily meaningful to accurately predict 'missing' longitudinal level.

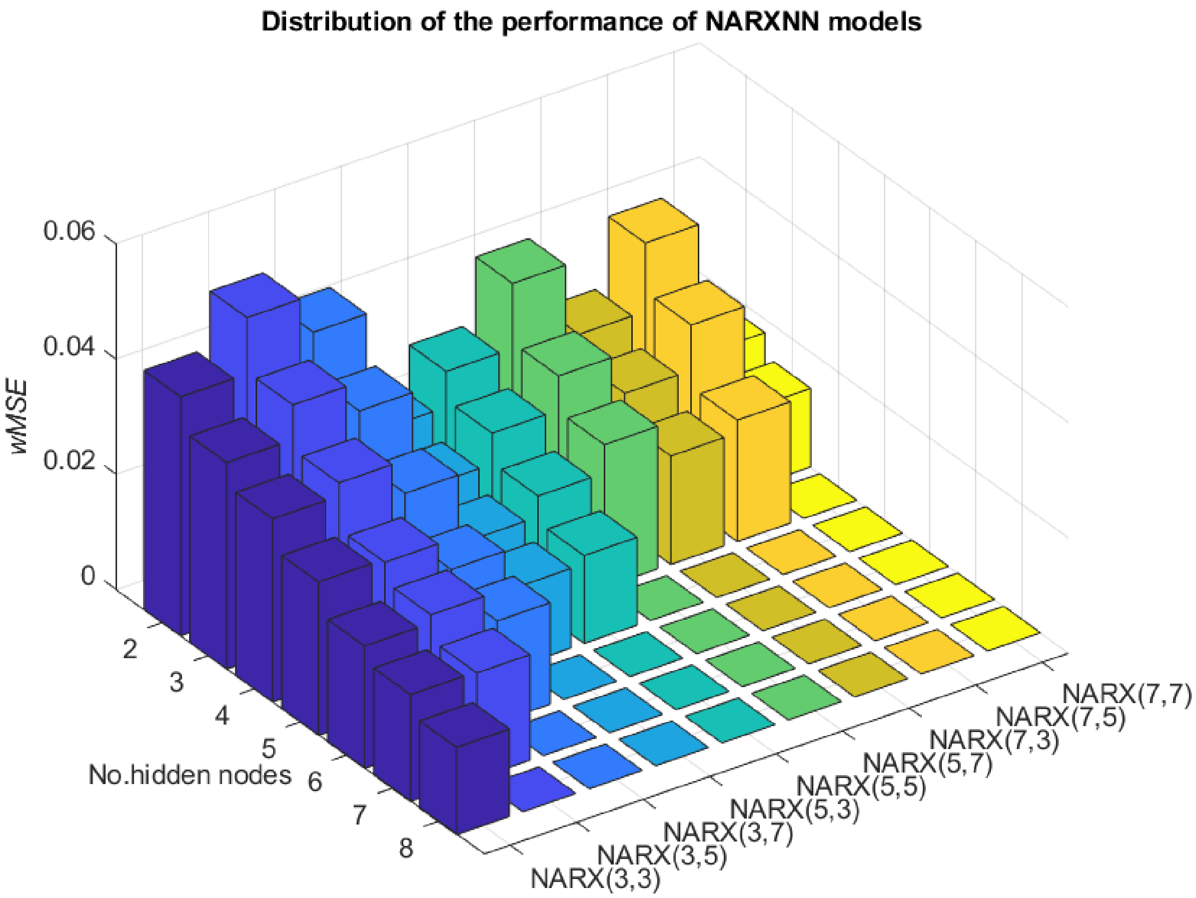


Fig. 6.7 Mean values of neural networks training results for a range selection of network topology

An application of clustering technique on trained iNARXNN networks (see Fig. 6.7 for example) can reveal such categories (or classes). This kind of information eases the process of prediction model selection. Fig. 6.8 shows clustering results on all networks where *k*-means technique was used for clustering which was performed subject to a pairwise value of (network size, *w*MSE). Any iNARXNN models with a number of parameters less than 150 can be considered as a simple prediction model.

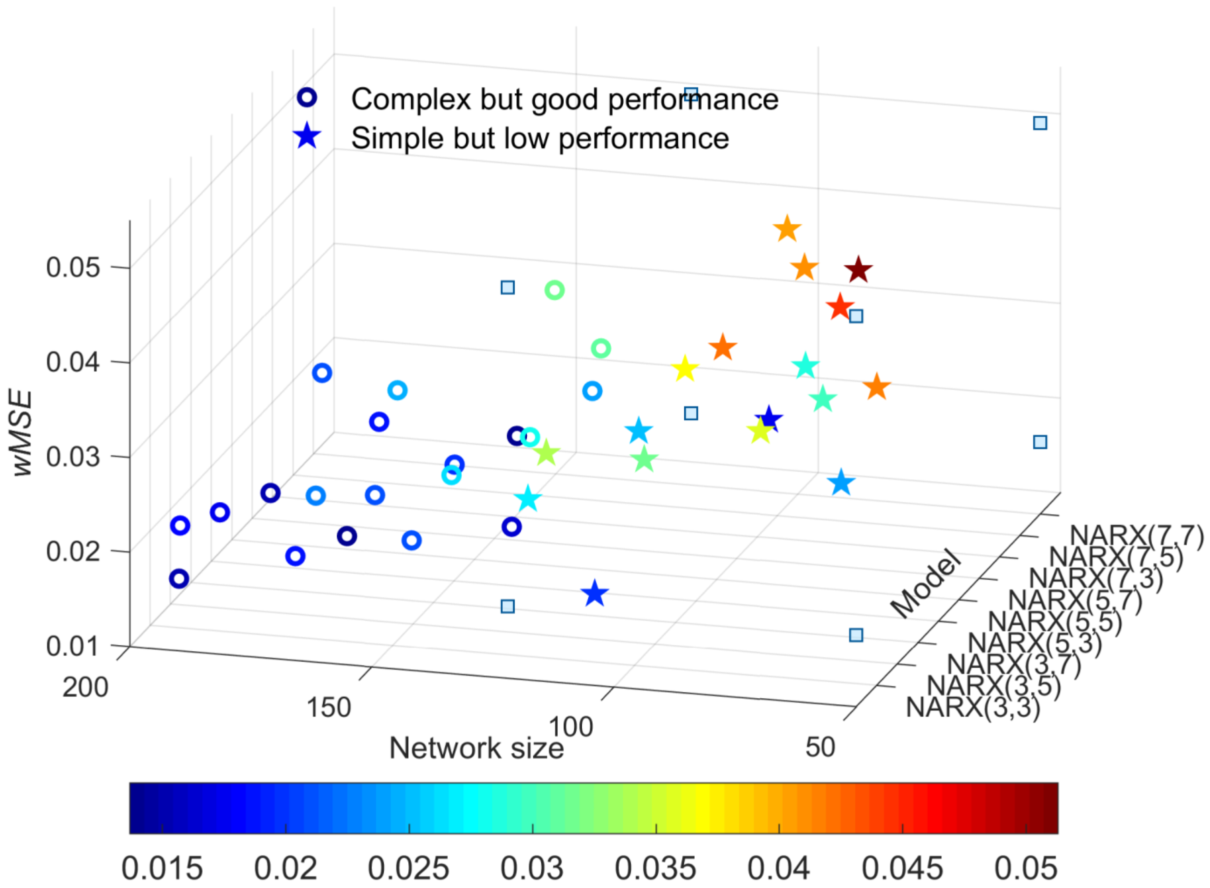


Fig. 6.8 The appearance of k=2 clusters of iNARXNN models for S_{29} case. Resulting clusters indicate that any networks with a number of parameters less than 150 should be considered as a simple prediction model.

6.4 Independent tests

Upon a completion of training stage, selected iNARXNN model(s) can be independently tested. Prior to the independent tests, the test data series formed by the n+1*th* track measurements are split into two parts. The first part contained track measurements data from the first d_3 track points; this was reserved for network initialization. The value of d_3 will vary from one network to another, as it takes the largest of d_1 and d_2 within the network under investigation. The remaining part of the test data series is later compared with the response from the network outputs.

6.4.1 Correlation plots

Fig. 6.9 (a) shows the predicted value of longitudinal level for x_i in the range of $[1681,1880]$ for three tested networks. Observations from the correlation plots in Fig. 6.9 (b)-(g) show that the three models predicted longitudinal level correctly based on the high correlation between predicted and target values.

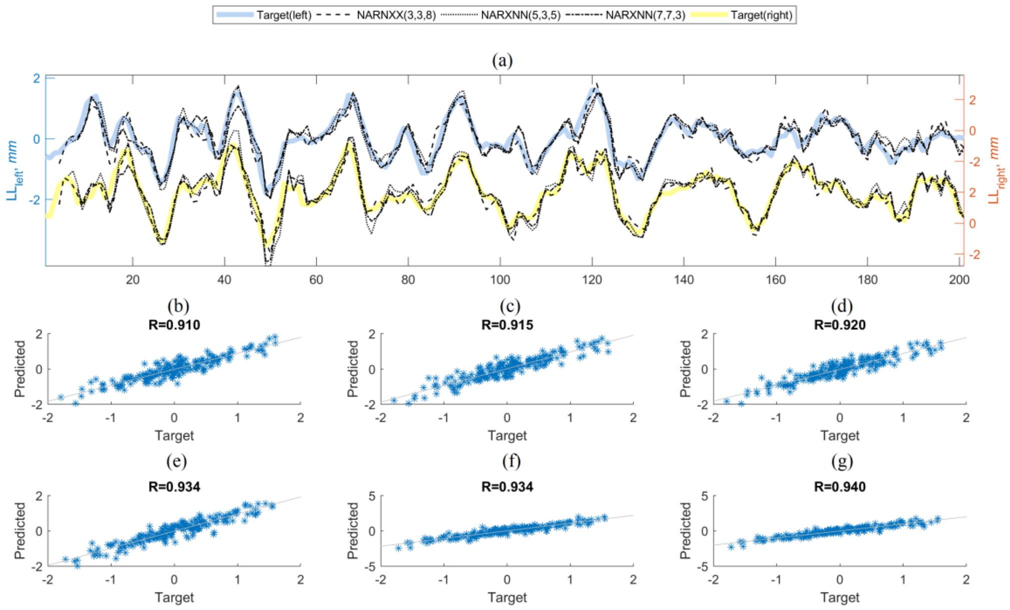


Fig. 6.9 Predicted and target values of longitudinal track level of test data series S_{29} for the left and right rail are superimposed in the plot (a). Overall prediction accuracy of different networks can be compared from the scatter plot (b)-(d) and (e)-(g) for left and right rail of S_{29}, respectively.

6.4.2 PSD comparison

An analysis of track measurement in the wavelength (or frequency) domain is useful when information about the shape of track defects and their wavelength content is needed (Haigermoser *et al.*, 2015). The information is further processed to describe track quality condition (Berawi, 2013). When considering spectral analyses, the PSD of track measurement is calculated where the corresponding power spectrums graph (see Fig. 6.10 for example) provide quality indicators to an analysed track segment. Here, the PSD was used to examine the relation between the artificial and actual longitudinal level data.

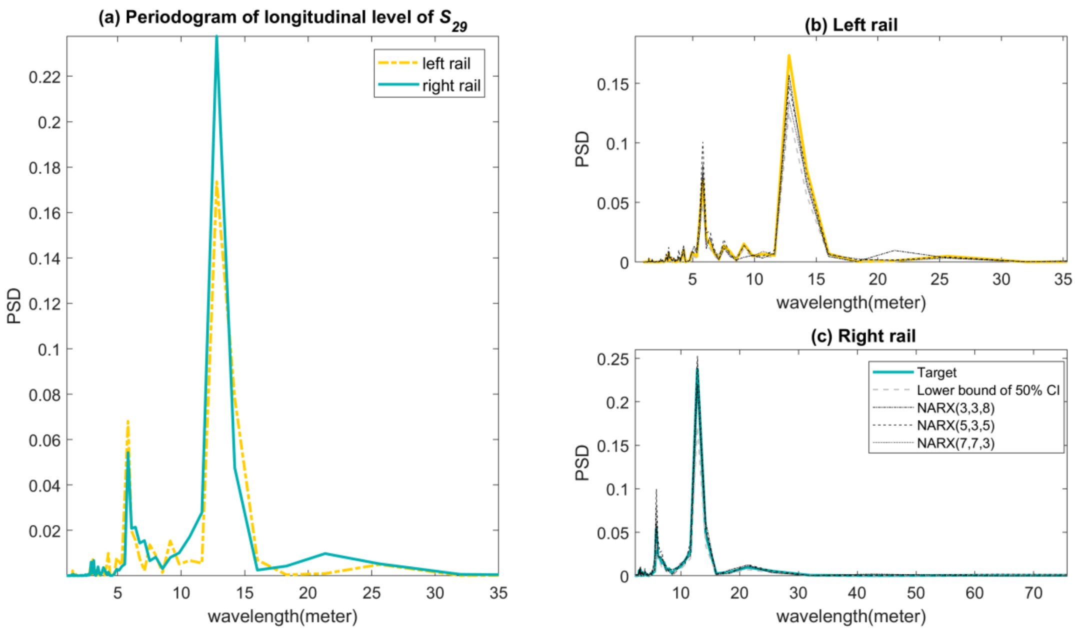


Fig. 6.10 PSD comparisons between a) the left and right rail, and artificial and actual longitudinal track level of b) left rail, and c) right rail for S_{29}.

6.5 Post-processing

As discussed in sub-sections 3.3.5-3.3.6 monitoring uncertainty propagation can reveal a sign of transition of unexpected changes in the value of the selected parameter (a deterioration rate, for example) upon an arrival of new data. By detecting the transition points (if exist) it creates valuable information (e.g. intrinsic value of previous inspection decisions) for a track engineer to better plan the next inspections. As a result, an area of track sections corresponds to the transition points will receive extra attention from a track engineer for further track examination. This specific advantage of uncertainty propagation analysis has been transferred to the development of iNARXNN model but with different tone of settings. It will be placed at post-processing stage of neural network to detect if any segments of the predicted track measurements has under-and/or-over estimated a TQI value. This is necessary to ensure outputs of iNARXNN do not mislead a track engineer when assessing track condition (see Section 3.2). Importantly, results of the post-processing contain no redundant information to what has been retrieved from the correlation analysis and PSD comparison.

For the post-processing stage in the iNARXNN model, the Bayes linear method has been proposed to assess uncertainty propagation in a rate of change of TQI in a deterioration model. Two quality measures associated with Bayes linear analysis, namely, partial size and bearing adjustment of expectation of prior belief, iteratively display how parametric uncertainty propagates from one inspection decision/result to the next. As will be shown later, graphical representation of these measures exhibits a transition point in the level of uncertainty i.e. a splitting of TQIs into two categories: gradual and sudden changes in TQI. In the light of iNARXNN, revisit the main body of iNARXNN is suggested for any TQIs produced from iNARXNN outputs under the latter category.

The Bayes linear method discussed in Section 3.3.6 evaluated the value of applying information obtained from the prediction of longitudinal levels in respect to an effort of strengthening prior belief about a deterioration rate of TQI that characterised the track geometry condition over time. This action could be seen as a short-term initiative i.e., taking place within an inspection interval, to averse from confronting with a sudden shift in track deterioration. For this purpose, a series of $Q_i = \{q_{i1}, q_{i2}, \ldots, q_{iM}\}$; $i = 1, 2, \ldots, N$ was obtained using Eqn. (3.1) where q_{it} is the TQI of a track segment and *i* corresponds to *t*-th predicted longitudinal levels for $t = 1, 2, \ldots, M$ with M being the number of acquired track measurements. For track recording car a machine travelling over a track section without looping, every Q_i from the same section will have the same value of *M*.

The initial condition after tamping and the deterioration rate of TQI of the track segment, denoted as β_o and β_1, respectively, made a collection of the quantity of interest $\mathbf{B} = (\beta_o, \beta_1)$ that this study wished to use to make inferences. Table 6.2 shows the initial belief about *B*, which is based on the results of the data analysis presented in Section 5.2.3. Goldstein and Wooff (2007) offered practical recommendations for belief elicitation in a case where users had little idea of where the true **B** lay over the given planning horizon *T*. Given the values of a subset $\mathbf{D} = (D_1, D_2, \ldots, D_{M-1})$ of $\mathbf{B}$, the prior belief about **B** is updated via the adjusted expectation $E_{\boldsymbol{D}}(\boldsymbol{B})$. Each quantity $D_j; j = 1, \ldots, M-1$ consisted of a finite number of deterioration rates of TQI obtained from a track section, from which S_{29} was extracted. To update the prior belief of **B**, Bayes linear method also required prior first- and second-moment in every D_j. A careful examination required a sequence of statistical tests to avoid the findings becoming irrelevant. An alternative hypothesis would be that each D_j deviates from normality i.e. a deterioration rate of TQI follows the inverse Weibull distribution.

In order to test an assumption of normality, in every D_j, the Anderson-Darling (AD) statistical test was performed. In brief, the AD test has a high power to reject the null hypothesis H_o which entails the definition that follows:

H_o: *The quantity* $D_j = (d_{1j}, d_{2j}, \ldots, d_{mj}); m \leq l_s$ *is a random sample from a normal distribution if the p-value associated with the AD test is not less than the selected significant level.*

Using the AD test the normality hypothesis had been rejected at the significance level $\alpha = 0.05$. Subsequently, the Weibull-type probability paper procedure which has been proposed by Khan and Pasha (2009) was performed to seek evidences to infer that each D_j can be represented with the inverse Weibull distribution. Prior belief about moments in ***D*** is then presented in Table 6.3.

Table 6.2 Prior specifications about B structure

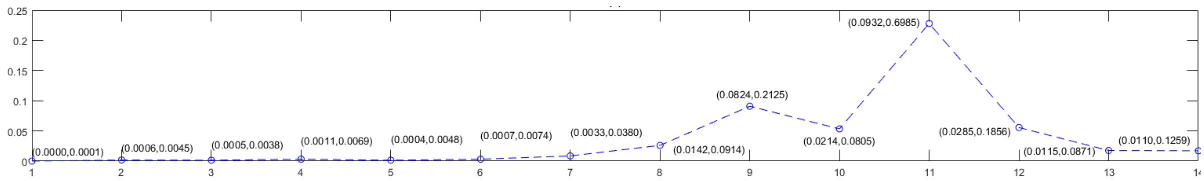
Source (Dataset)	$E(\beta_o)$	$\mathrm{Var}(\beta_o)$	$E(\beta_1)$	$\mathrm{Var}(\beta_1)$
200305	0.1	2	0.0033	0.661
200907	0.1	2	0.0051	2.382
200911	0.1	2	0.0050	2.221

The belief about ***B*** overall updated to an expectation of $E_{\boldsymbol{D}}(\beta_o)$ and $E_{\boldsymbol{D}}(\beta_1)$ with variances of $var_{\boldsymbol{D}}(\beta_{\boldsymbol{o}})$ and $var_{\boldsymbol{D}}(\beta_1)$, respectively. In Fig. 6.11(a), comparing to the maximum value of the size of adjustment, i.e. using the first 11 quantities, a decision of using a full ***D*** has extremely decreased the highest $Size_D(\boldsymbol{B})$ about 95%. However, the $Size_D(\boldsymbol{B})$ associated with full ***D*** has a percentage increment about 360% as compared to a decision using only the first quantity. We see that there is no significant change in the $Size_D(\boldsymbol{B})$ despite extending the initial test to include more quantities (up to six quantities). An average individual adjustment on (intercept, rate), as shown in Fig. 6.11(b), shows that all of the first eight $D_{\boldsymbol{i}}$ fairly have similar information gains. However, there is a clear fluctuation in the size of adjustments when $d_{1,\dots,8\cup j}; \mathrm{j} = 9,\dots,14$ was tested. Among all *j* quantities, the tests showed that D_{11} has adjusted the prior belief the most and followed by D_{12} as the next best informative quantity to use for belief updating. Adding D_{13} into $\boldsymbol{d}_{1,\ldots,12}$ dramatically reduces the $Size_D(\boldsymbol{B})$ but the value is likely unchanged with a participation of D_{14} in tests. Moving to Fig. 6.11(c), testing results show that prior belief updated in a different direction from what it experienced with $\boldsymbol{d}_{\boldsymbol{F}/\boldsymbol{D}}$. In fact, a direction of change can be seen in the negative region of Bearings itself, for example, $d_{D_{11}/D_{10}}, d_{D_{12}/D_{11}}$ and $d_{D_{13}/D_{12}}$. Importantly, evidences from the Fig. 6.11(c) shows that results of iNARXNN are trustable at least for one-time step prediction.

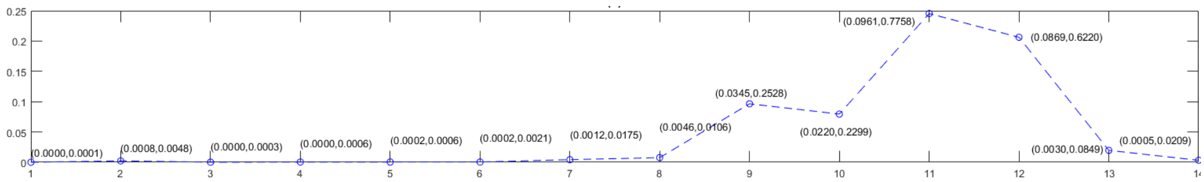
Table 6.3 Prior specifications about D structure

Variable	Prior distribution	Expectation	Variance
D1	InvWeibull	0.0044	0.9521
D2	InvWeibull	0.0072	0.811
D3	InvWeibull	0.0069	1.254
D4	InvWeibull	0.0084	1.588
D5	InvWeibull	0.0071	1.089
D6	InvWeibull	0.0081	1.001
D7	InvWeibull	0.0091	2.100
D8	InvWeibull	0.0082	0.0888
D9	InvWeibull	0.0089	1.077
D10	InvWeibull	0.0099	1.843
D11	InvWeibull	0.0121	2.100
D12	InvWeibull	0.0077	1.331

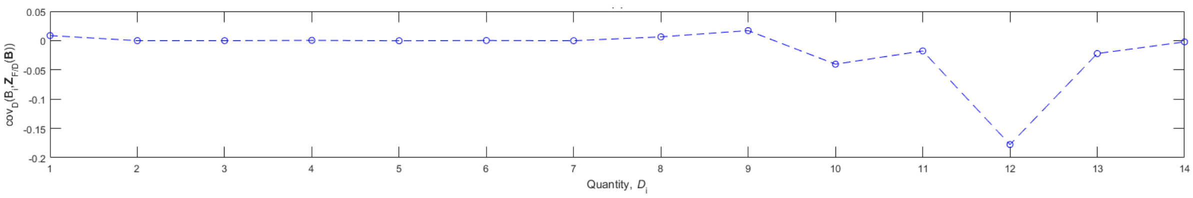
D13	InvWeibull	0.0071	0.887
D14	InvWeibull	0.0070	1.111



(a)



(b)



(c)

Fig. 6.11 Evolution in uncertainty propagation in the belief structure over a defined planning horizon represented in three modes; a) Size of adjustment, b) partial size of adjustment, and c) partial bearing of adjustment.

6.6 Summary

In this chapter, a prediction model iNARXNN was developed to predict 'missing' longitudinal levels for both left and right rails. At a prediction power of roughly 95%, while keeping its spectral features nearly identical to the original one, iNARXNN is a reliable response model to restore resilience to a disrupted track inspection. Between track inspections, a track manager can rely on the predicted measurements, rather than field measurements, for track quality assessment. With the accurate predictions of iNARXX, infrastructure managers will gain an advantage when designing a rescheduling strategy that can reduce the number of inspections, i.e. no on-site inspections. The next chapter provides a resilience model for a disrupted track inspection that assigns values to different rescheduling strategies, including one incorporating iNARXNN.

CHAPTER 7 RESILIENCE ACTION

7.1 Introduction

Limited resources, for example, inspection vehicles and manpower, and time restrictions to assess tracks, have underpinned the need of (the predetermined) inspection schedules to effectively prioritise inspection tasks. In this regard, track inspection is a proactive action to assure safety and operability of the tracks. However, due to the fact that the complexity of the TIS model increases along the amount of uncertainty associated with track inspection and maintenance; thus, not all uncertainties are able to be taken into consideration at the design phase (Snowling, 2000). Well said, the TIS is designed under uncertainty. In the study by Zhang et al. (2000), the authors explained different types of uncertainties that arise in modelling non-destructive inspections, which in principle demands parallel and/or hierarchical complex analysis due to a number of constraints, such as lack of quality data, model characteristics and design, knowledge–interpretation clash and inadequate information. Comprehensive discussion with regard to this issue can be found in (Brugnach *et al.*, 2008). In line with this, performing trading model complexity with uncertainties, especially those with very low information, is necessary and it indirectly creates an opportunity for railway infrastructure managers to effectively manage disruptions.

Realising that most disruptions are unforeseeable, many studies have focused on reducing the consequences of disruptions (Chopra and Sodhi, 2004) and some have employed disruption management as an underlying theory. In regards to disruptions in TIS, an inspection reschedule would be an appropriate resilience action. Therefore, the goal of rescheduling is to minimise impacts from disruptions and assure the survival of the dedicated schedule until the end of targeted period i.e. a preventive maintenance cycle. Besides protecting organization's reputation, the benefits of inspection reschedule is found in the effectiveness of rail maintenance decision, for example, the optimal downtime per unit time will start to decrease if there is a disruption during the day of operations (Zhen *et al.*, 2016).

As an initial inspection decision might be altered, the value of rescheduling should be established in order to guide decision makers. The value in rescheduling is said to be created if benefits exceed total incurred costs. For this reason, the benefit-cost model of planned (and preventive) maintenance is revisited. Benefit-cost analysis is a common practice in every department in any business sector to project company cash flow in future flow when an investment in new or improving the current product (or service) is proposed. To execute the process analysis successfully, expected benefits from and costs applied to the proposal should be properly formulated before the identification of the driving factors that influence the benefit–cost trade-off are carried out. Prior to formulate a measure of value of inspection reschedule, a cost breakdown of track inspection (as well as track maintenance) is presented.

In such cases where value does not exist, the object of interest i.e. the rescheduling process, could demonstrate value added to track inspection scheduling, through an introduction of innovative procedure to manage disrupted schedules. In this chapter, risk and economic elements are studied in parallel to establish resilience metric for the formulation of rescheduling disrupted track inspection.

7.2 Track inspection costs

Inspection is preceding tasks/activities in a condition-based maintenance strategy. It becomes asset owner's responsibility to design an efficient and effective inspection strategy to capture as much as possible the benefits of asset maintenance. Impacts of the proposed design and decision in maintenance, in the long run, directly appear in the operation and support cost, as illustrated in a cost distribution diagram (Gilmore and Valaika, 1992; Obrenovic, Jaeger and Lemmer, 2006), see Fig. 7.1. Apart from inspection, capacity, substance and quality of railway tracks are major parameters that should be stressed on when evaluating track maintenance cost over the remaining lifetime (Zoeteman, 2001; Setsobhonkul, Kaewunruen and Sussman, 2017). Interdependence among these parameters is observable in the maintenance component of the LCC model, which can be systematically drawn from a cost breakdown structure.

The cost breakdown structure is a tool, at the analysis stage of LCC process for identifying all relevant activities that consume the organizational resources or the so-called 'cost elements' with respect to each cost category, e.g. maintenance costs. A precise definition and formulation are important to avoid overlooking the cost elements that significantly affect the total LCC. This recommendation would also considerably accelerate the process of tracking down cost drivers.

A cost driver is basically an activity measure that serves as a basis for the activity's cost allocation. For example, the number of inspections and type of inspection could be a cost driver for a system quality assurance activity, but the cost drivers may be attached to different units of costs. When several cost elements are similar with respect to their cause-and-effect relationship, these elements can be pooled and represented with a homogeneous cost driver(s). At this point, cost allocation should be made based on the degree of correlation between the consumption of the activity and the consumption of the cost driver. The accumulated cost incurred for a specific group of activities is known as the cost activity pool. To gain insight about the deviation in total LCC of the baseline design/system due to the changes in system objectives, a sensitivity analysis on the identified cost pools is recommended.

Fundamentally, there is not much deviation with respect to the maintenance of LCC models in whichever industry that homes the system of interest. The only thing that differentiates the models is the cost breakdown structure depending on managerial viewpoint, which is domain dependent. For railway track maintenance, the breakdown and identified cost drivers shown in Fig. 7.2 are an adaptation of those used in (Nissen, 2009a) with simplification and/or addition from (Thoft-Christensen, 2012; Stenström *et al.*, 2015).

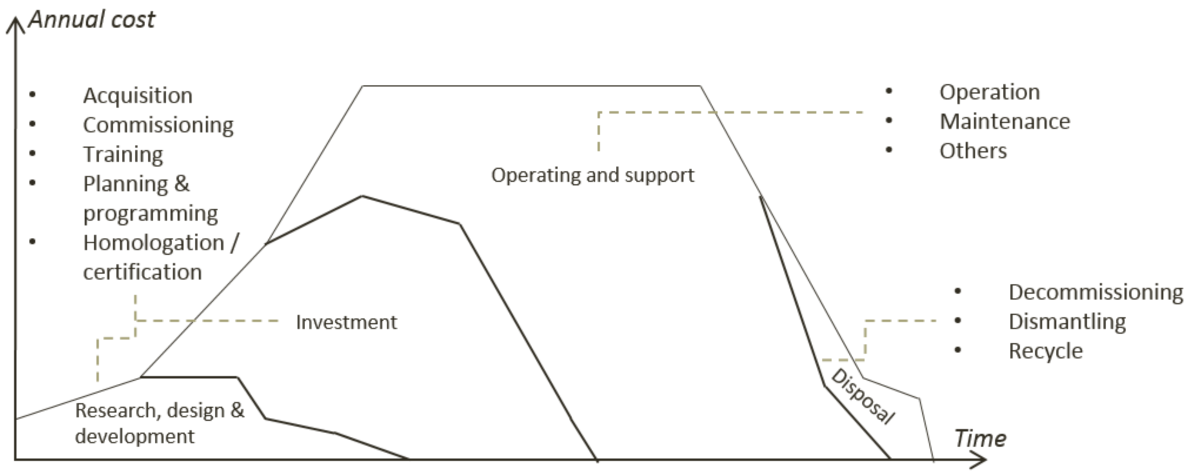


Fig. 7.1 An illustration of notional profile of annual activities expenditures by major phases (cost categories) over a system life-cycle cost. The pattern shown is only for illustrative purpose, the actual curve will vary from one system to another. Adapted from (Obrenovic et al. 2006; Office of the Secretary of Defense 2014).

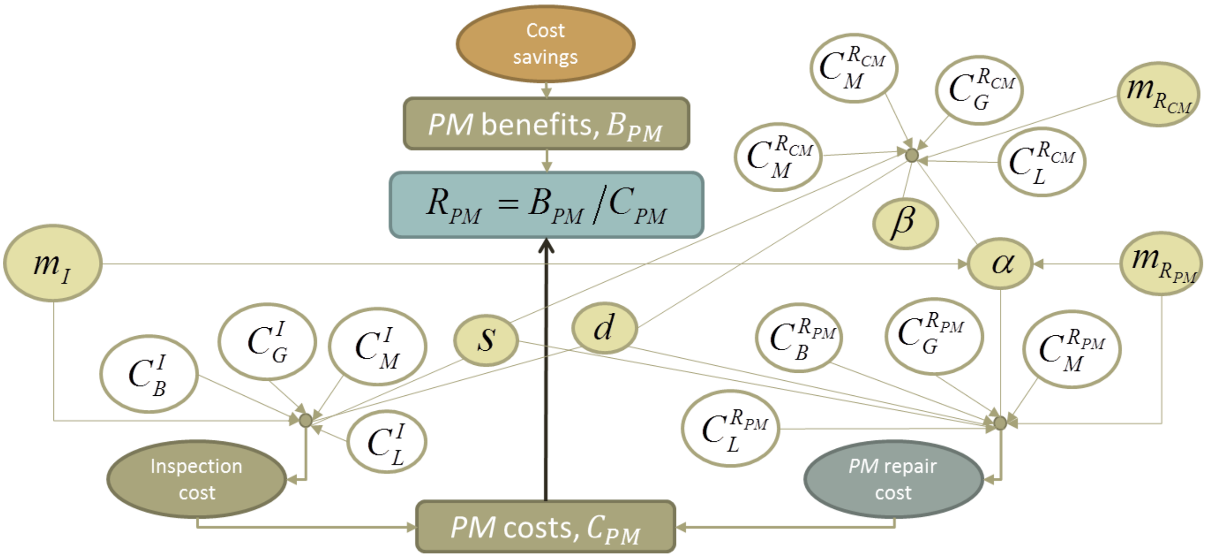


Fig. 7.2 Major cost elements and drivers related with preventive track maintenance. At least one inspection is required otherwise the C_{PM} is solely dependent upon R_{PM}. Note that C_x^y takes $x \in \{B: \text{labour}, G: \text{logistic}, M: \text{machine \& materials}, L: \text{business loss}\}$ and $y \in \{I: \text{inspection}, R_{PM}: \text{PM repair}, R_{CM}: \text{CM repair}\}$.

Let m_I be the minimum number of inspections under a periodic interval, for example, τ. Then, the total cost of track inspection over the PM cycle I_{PM} can be calculated from the generalised equation given by Stenström *et al*. (2015):

$$I_{PM}\left(m_I, \tau, s, d\right) = m_I\left(\tau\right)\left(C_B^I\left(s,d\right) + C_G^I\left(s,d\right) + C_M^I\left(s,d\right) + C_L^I\left(s,d\right)\right) \tag{7.1}$$

where s is the number of personnel on the inspection (or maintenance) team and d is inspection (or maintenance) times, including administrative time, logistic time, active inspection (or repair) time and service/production loss time. Let m_{PM} be the number of potential failure repair and m_{CM} be the number of functional failure repair. Then, both the PM repair cost $C^{R_{PM}}$ and the CM repair cost $C^{R_{CM}}$ can be calculated similarly.

To summarise, conducting short inspection intervals enables more frequent inspections to be carried out, in which up-to-date information about the condition could increase the opportunity of defect detection. Nevertheless, railway infrastructure companies have to assign a minimum value for m_I due to their limited resources and the large network size to be inspected. The ratio between investment benefits and incurrence costs is a primary and widely accepted analysis tool for financial profit assessment. Regarding preventive maintenance, the ratio for each finite PM cycle R_{PM} is calculated by Eqn. (7.2) (Stenström *et al.*, 2015):

$$R_{PM} = \frac{B_{PM}}{C_{PM}} = \frac{\alpha\beta C^{R_{CM}}}{I_{PM} + \alpha C^{R_{CM}}} > 1 \tag{7.2}$$

where α and β represent the probability of defect detection and likelihood of track condition having a sudden shift, respectively. Both parameters have values in the interval $(0,1]$ which leads to $\alpha\beta > 1$.

7.3 Mapping TIS in rescheduling framework

In complex (i.e. interrelationships between several resources exist) situations, applying changes to an initial schedule is not interesting but, in the event of disruption, changes by means of rescheduling are necessary. Fox et al.(2006) points out that rescheduling is a challenging operation as the minimisation of changes in the original schedule and the impact of disruptions are two principle objectives that need to be satisfied. Recently, a new study of rescheduling is highly recommended to adopt a framework designed by (Vieira, Herrmann and Lin, 2003) to obtain an insight about the necessary requirements to mitigate disrupted risks upon a predetermined schedule in a realistic way. Fig. 7.3 shows four dimensions/compositions underlying the framework. Issues and development of each dimension is discussed in the following sub-sections accordingly. The scope of discussion is limited to relevant research materials associated with railways, transportation and maintenance.

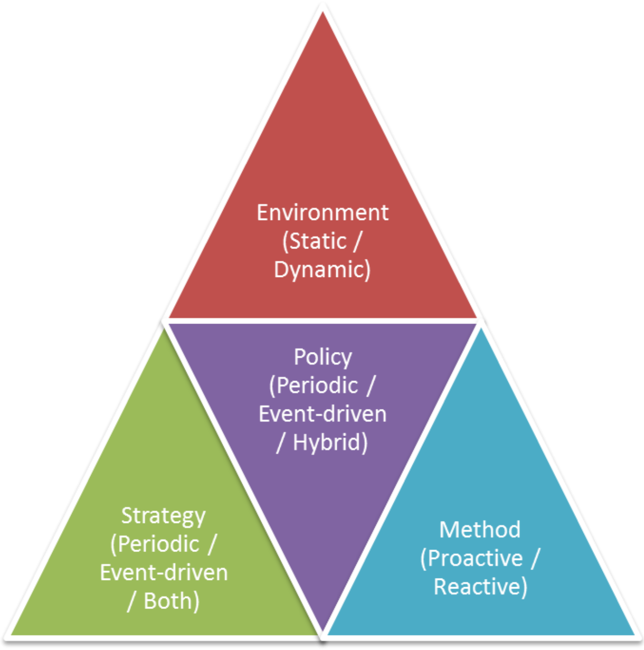


Fig. 7.3 Interdependent components of rescheduling framework

7.3.1 Dimension 1: Environment

The rescheduling environment explains which part of a disrupted schedule needs a modification. It begins by classifying the schedule into static or dynamic. In contrast to a dynamic schedule, a static schedule has a finite number of tasks. In the context of railway operation, and probably in all public transportation business, resources schedules including timetables have a finite size of tasks where their information is certain, such as the number of trips in a week, routes and fares. In fact, most railway-related schedules are decided and planned several months before the operation year starts. However, for some static schedules, even though the number of tasks is known, uncertainty in the schedule parameters might occur. For example, RIM usually knows the number of tracks that need to be inspected, but, an exact inspection frequency of each track is not certain due to random failure occasions (Kashima, 2004).

7.3.2 Dimension 2: Strategy

Rescheduling is a synonym to recovery action and would be performed upon disruption occurrence. In this light, disruption itself is unexpected as a scheduler/planner does not know an exact time when it will occur even if we can study potential sources and consequences of disruption. However, those characteristics are not an obstacle to users who aim to prepare a rescheduling strategy ahead of time. Having the knowledge on what strategy needs to be implemented would ease the process of deciding whether an existing schedule requires rescheduling or not. Moreover, to what extent changes could take place in a schedule is possible in regards to the process of visualisation and estimation.

Furthermore, the preferences of schedulers/planners during the schedule design stage influences his/her approach towards disruptions. In predictive-reactive scheduling, the predetermined schedule is rescheduled in response to disruptions and other changes within the environment. In such cases, the provisional schedule could be fundamentally deviated from the original one, which may cause poor performance owing to affecting other planning activities based on the original schedule. For these reasons, some authors tend to generate a schedule that aims to act robustly in response to a specific disruption (Okoh and Haugen, 2015). Meanwhile, Fang et al. (2015) stated that establishment of a schedule that behaves in a robust way toward different disruptions is hard and becomes harder as time for rescheduling is often limited in railway systems. As stated in Cacchiani et al. (2014), contingency plans are real when handling disruptions to railway timetables and are being practised in the Netherlands. There are one thousand pre-planned periodic timetables for various types of disruption. Preparation of those plans is definitely a loss in a year where no disruption occurs. Moreover, only an experienced scheduler is capable of performing a plan selection. Nevertheless, an application of predictive strategy in an uncertain environment demands extra careful observations when dealing with problem formulation, as resources, e.g. time slots, vehicles, etc., are already over-utilised.

Regardless of the chosen rescheduling strategy, the consequence level of the reschedule factor needs to be measured first. In this light, rescheduling is exempted if the consequence level is small (relative to users definition) or otherwise, the schedulers will move to the second and third steps which is finding a rescheduling solution(s) and revise the solution selection. The existence of many rescheduling methods opens an opportunity to offer several alternatives that can improve the performance of an existing schedule to the decision-makers (Vieira, Herrmann and Lin, 2003). Hence, in case the solution generates unsatisfactory results (user-defined value), the decision makers could apply the second-best alternative or they can repeat the solution step with or without the parameter changing. Consequently, there are several classification schemes introduced to reschedule methods and this study will discuss some relevant classifications in the following section. As such, a rescheduling policy would help users to attain a good justification for selecting particular methods and to revise the solution.

7.3.3 Dimension 3: Policy

An implementation of a predictive-reactive rescheduling strategy is driven by a rescheduling policy, which refers to a type of event that orders rescheduling of an existing schedule. In a periodic situation, use of a single schedule for numbers repeatedly keeps schedulers from spending resources again on schedule generation. To own this opportunity, a scheduler should gather comprehensive information from different aspects related to the system which the schedule is operating in. For example, track inspection schedules always come after an international freight train, passenger train and major possessions schedule are announced. However, as internal and/or external environments of the system probably change due to various reasons, rescheduling the existing schedule before the next period begins could yield stability and reduce schedules performance deterioration. In this situation, a scheduler applies a periodic policy where a fresh system status focusing on irregularities is taken into account when seeking a better schedule.

As discussed earlier, a disruption might attack a schedule during its execution period. Depending on the level of consequence associated with the disruption, a range of rescheduling methods can be used to create or update a schedule. At this point, rescheduling is driven by events. In most situations, those events are not part of the schedulers consideration when creating an initial schedule due to uncertain information about them. In the case of dynamic schedules, the event is not necessarily a disruption. It could be a regular task with incomplete information. A periodic schedule also can be rescheduled due to unscheduled events. Vieira et al. (2003) uses a hybrid policy term when referring to this situation.

7.3.4 Dimension 4: Method

Understanding which type of rescheduling strategy a user can implement when selecting a rescheduling method can be a straightforward action. For a predictive strategy, many methods ranging from evolutionary algorithms to numerical simulations have been adopted to generate a robust schedule in response to a specific disruption. Yang et al. (2014)

proposed a method to create predictive schedules that include inserted buffer time (idle time) as a means to absorb some level of uncertainty without rescheduling. Typically, this method employs a bottleneck algorithm to handle parameter uncertainties and then places idle time to improve schedule robustness. In work by Higgins et al. (1999), an integer programing has been used to create a robust schedule that minimises train delays due to maintenance activities. Louwerse and Huisman (2014) reviewed works of decomposing railway schedules and solved sub-schedules partially. For incomplete sub-schedules, real-time information is used to resolve the incompleteness issue at an appropriate time.

As the level of uncertainty is too high, there is a tendency to perform rescheduling upon arrival of the disruption. Schedulers become reactive and deliver necessary efforts to minimise the size of changes and/or impacts due to a disruption. Occasionally, a scheduler may find an unexpected event triggers small deviations and probably chooses to continue operating the schedule without making any changes. If that is the case, schedulers simply assume that the impacts of minor disruption will be automatically accommodated in a short time and will not affect the remaining operations schedule. Another typical way to repair a schedule from that situation without changing a sequence of remaining operations is by shifting all of the operations altogether to the right on the time axis. For railway applications, shifting the method is almost impossible to be implemented particularly in a timetable because a delay of minutes in all train services would cause chaos in railway stations.

A preferable method of handling disruptions with respect to railway operations is probably a recovery method, also known as a partial reschedule. This method aims to keep a degree of deviation from the existing schedule as low as possible when parts in the schedule that are directly or indirectly affected by disruption are given treatment (rescheduled). In Walker et al. (2005), a cost increment is allowed when doing schedule recovery but it is aimed to keep it small. A vast collection of recovery algorithms can be found in Cacchiani et al. (2014). Branch-and-bound, mixed integer programming, alternative graph model, genetic algorithm, column generation and heuristic algorithms have been applied in various type of rescheduling problem related to railway. Interestingly, the authors highlight a great challenge of performing recovery on an integration phase (i.e. timetable, rolling and crew schedule). Integrated rescheduling of two resources has succeeded in various studies, such as in Fekete et al. (2011), which uses integer programming.

Periodic schedules can have full-scale rescheduling before the beginning of a next schedules period. This situation normally appears when the existing schedule is no longer feasible for the next application due to the system within the schedule changing vastly. At this point, the literature on scheduling methods, for example in Narayanaswami and Rangaraj (2011), can be revisited to foster new schedule development. Conceptually, repair and recovery methods take less computational effort compared to total rescheduling (Fox *et al.*, 2006). In different versions, the use of contingency planning can be seen as an example of total rescheduling. A substantial number of publications regarding contingency plans have been published in the area of resource allocation (Blos and Miyagi, 2015).

7.4 Resilience metric for track inspection reschedule

A track inspection schedule comprises of a finite number of (work) trips which consist of a set of all of the movements an inspection team (man and machine) may make in the network. There may be more than one sequence of movements (corresponding to alternative routes) that can define a path through the network. The dedicated schedule I is stable if no disruptions occur at any time during an observation period. In other words, no change, amendment or delay will be applied to the schedule. Regardless of the disruption events that lead to rescheduling I, the aim is to maximise the value of inspection reschedule V. This can be calculated as follows:

$$V(I,\hat{I}) = -\frac{C_r}{w_b \Delta Risk(I,\hat{I})} \tag{7.3}$$

where C_r is the total cost incurred for rescheduling. A weighting factor $0 < w_b < 1$ is applied to $\Delta Risk\left(I,\hat{I}\right)$ which represents the size of change in the risk score corresponds to I due to an introduction of updated TIP $\hat{I}$. The factor is introduced to avoid overwhelming the risk of missed opportunity.

7.4.1 Costs of rescheduling

In this study, C_r is formulated as the sum of $\Delta C_{ins}\left(I,\hat{I}\right) = C_I - C_{\hat{I}}$ and C_p. The first component is a difference in the cost of single type of track inspection between two TIPs; I and $\hat{I}$. For a track section of length L (in kilometre) which is expected to receive m inspections, the corresponding inspection cost can be calculated from following function:

$$C_I\left(m,c_d\right) = mc_d + \sum_{i=1}^{m} C_i^{poss} + mc_a \tag{7.4}$$

where c_d is a direct cost per inspection. For a vehicle-based inspection such as a track geometry inspection, c_d can be estimated from the cost for operating inspection vehicle in which $c_d = \frac{L}{200} c_h$ where c_h is an operational cost per 200 meter section per vehicle (Soleimanmeigouni *et al.*, 2016).

Track inspections involve a high volume of short-duration tasks (in the range of 1–4 hours), and it is important to perform them systematically and objectively as inspections incur a track possession cost. The base formula to calculate cost of planned track possession is given as follows (Halcrow, 2013):

$$C_i^{poss} = \left(t_{DM} \times MRE \times \eta_i\left(j\right)\right) \times n_f \tag{7.5}$$

where t_{DM} and $\eta_i(j)$ denotes estimated delay minutes per passenger journey caused by track possession (in minutes) and the average number of passenger journeys during time band *j*, respectively. The Marginal Revenue Effect (MRE) is the amount of long term revenue estimated to be lost by a passenger operator from one minute of lateness per passenger journey (Halcrow, 2013). The revenue lost is because a proportion of passengers switch away from travelling by rail because of delays. The MRE by ticket type can be calculated using the following equation (Ford, 2018):

$$MRE_g = \left(LTM_g \times ElasGJT_g \times rev_g\right) \big/ GJT_g \tag{7.6}$$

where LTM_g and $ElasGJT_g$ respectively represents the Late Time Multiplier (LTM) and generalised journey time (GJT) elasticity by ticket type *g*. Basically there are three ticket types i.e. $g = 1, 2, 3$ namely, 1) full (peak time), 2) reduced (off-peak) and 3) seasons. LTM expresses late arrival time in equivalent units of scheduled travel time. For example, passenger rail transportation in the United Kingdom uses a late time multiplier of 6.0 for both airport inbound and outbound journeys. GJT represents the train ‘s quality of service aspects, including the journey time, interchange and frequency. In principle, GJT could be enhanced by reducing in-vehicle time, transfer time, service headway and number of interchanges. Changes in GJT result in a demand response in the rail demand-revenue model by applying a $ElasGJT$ (Wardman and Batley, 2014). The higher the absolute value of $ElasGJT$, the larger the percentage of change in the volume of rail demand. Revenue per passenger journey by ticket type is denoted by rev_g.

The total amount of revenue lost which can be obtained by multiplying all terms inside the bracket in Eqn. (7.6) is discounted at a notification discount factor n_f; an incentivise IM to plan possessions (Halcrow, 2013).

As administration elements have important features in scheduling (and maybe rescheduling) track inspection, the corresponding administrative cost is attached to C_I. For simplicity, c_a is set to be proportional to the total direct cost of inspection; $k\left(mc_d\right)$ where a constant $k>0$, that is, Eqn.(7.4) turns into following:

$$\begin{aligned} C_I\left(m,c_d\right) &= mc_d + \sum_{i=1}^{m} C_i^{poss} + m\left(kc_d\right) \\ &= \left(1+k\right)mc_d + \sum_{i=1}^{m} C_i^{poss} \end{aligned} \tag{7.7}$$

Consider m^* out of m planned inspections in I will be rescheduled over the remaining maintenance cycle, the ΔC_{ins} is calculated as follows:

$$\begin{aligned} \Delta C_{ins}\left(I,\widehat{I}\right) &= C_{I/\underline{I}}\left(m^*,c_d\right) - C_{\widehat{I}}\left(\underline{m}^*,c_d\right) \\ &= \left(1+k\right)\left(m^* - \underline{m}^*\right)c_d + \left[\sum\nolimits_{\forall i\in I/\underline{I}} C_i^{poss} - \sum\nolimits_{\forall i\in \widehat{I}} C_i^{poss}\right] \end{aligned} \tag{7.8}$$

where $\underline{I} \subset I$ be the set of planned and completed inspections and $\underline{m}^*$ denotes the number of inspections in $\widehat{I}$. In the case of $\underline{m}^* = m^*$, ΔC_{ins} is then simply a difference in the total cost of track possessions between $I/\underline{I}$ and $\hat{I}$ and can be calculated from the following:

$$\Delta C_{ins}\left(I,\widehat{I}\right)=\sum\nolimits_{\forall i\in I/\underline{I}} C_i^{poss}-\sum\nolimits_{\forall i\in\widehat{I}} C_i^{poss} \tag{7.9}$$

There are three types of (administration) activity that could be involved to reschedule I ; $W=\{1:\text{adjust},\ 2:\text{cancel},\ 3:\text{add new}\}$. For each activity *j*, $j\in W$ a fraction of administration cost, $w_j k c_d$ will be charged as a penalty cost on I. Let n_j be the count of activity *j*, $j\in W$, required to construct $\hat{I}$, thus, the corresponding total penalty cost can be calculated from $C_p = k\sum\nolimits_{\forall j\in W} n_j w_j c_d$.

Hence,

$$\begin{aligned} C_r &= \Delta C_{ins}\left(I,\widehat{I}\right)+C_p \\ &= k\left(m^*-\underline{m}^*\right)c_d+\left[\sum\nolimits_{\forall i\in I/\underline{I}} C_i^{poss}-\sum\nolimits_{\forall i\in\widehat{I}} C_i^{poss}\right]+kc_d\sum\nolimits_{\forall j\in W} n_j w_j \end{aligned} \tag{7.10}$$

7.4.2 An example

In this example, any reschedule decision is considered economically feasible if the corresponding reschedule cost has a negative value, i.e., $C_r<0$. Two possible scenarios of track inspection reschedules were considered; 1) from the weekend to afternoon off-peak, and 2) from the weekend to weekday morning off-peak. Table 7.1 displays parameters values used to calculate Eqn. (7.11).

For both scenarios, reschedule costs over four different values of adjusted delay minutes t_{id} are displayed in Fig. 7.4. Clearly, the reschedule cost increases as the delay minutes caused by new track possessions increases. Interestingly, the original inspection schedule corresponds to the second scenario experienced a progressive increment in the reschedule cost in comparison with the first scenario of rescheduling. Overall, no feasible solution can be achieved for the first and second scenario when t_{id} rises more than 84 and 0.6 seconds, respectively.

Table 7.1 Assigned values to model parameters in Eqn. (7.11)

Item	**Parameter**	**Value**	**Reference**
Direct inspection cost	L	13.6 *km*	Section 5.1.2
	c_h	£20	(Soleimanmeigouni *et al.*, 2016)
Admin cost	k	0.11	(Stenström *et al.*, 2016)
Reschedule factor	w_2	0.68	(Halcrow, 2013)
	w_3	1.50	
Minutes delay	t_{DM}	2.32	Raildeliveygroup.com
Discount factor	n_f	0.45	(Halcrow, 2013)
MRE	LTM_g	3.0	
	$ElasGJT_g$	-1.10	
	GJT_g	12 minutes	(Wardman and Whelan, 2001)
	rev_g	£5.40	(UK Office of Rail and Road, 2017)
Average passenger journeys	Weekday morning off-peak	1558	(National Statistics 2017)
	Weekday afternoon off-peak	1025	
	Weekend service time	585	

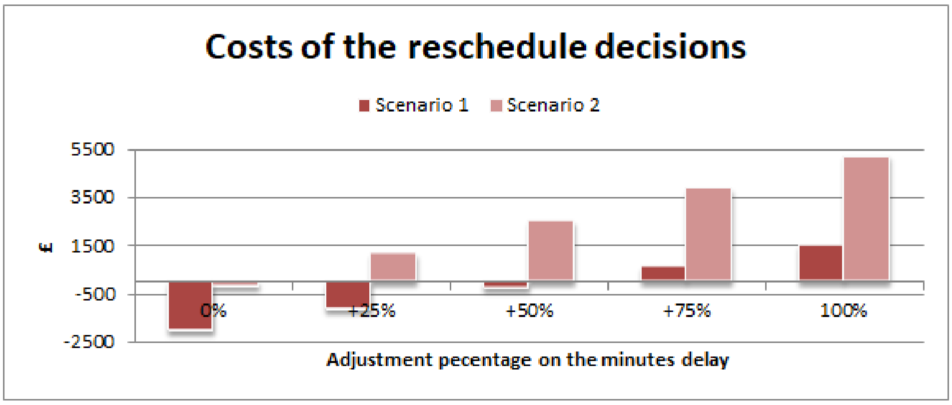


Fig. 7.4 Effects of increasing minutes delay on the reschedule cost.

7.4.3 Risk of opportunity loss

The risk of missing an opportunity to repair a defective track before it fails due to presence of an undetected defect during that period can be calculated as follows:

$$Risk(h_1, h_2) = C_u \sum_{r=1}^{R} S_r N_r(h_1, h_2) \tag{7.11}$$

Eqn. (7.11) is a summation of R risk factors where each factor is a product of the severity of failure, S and the total expected number of 'soft' failure events in the periods (h_1, h_2) N multiplied by the potential financial loss C_u. Values of the pair (S, N) are associated with a type of defect that can lead to a specific track failure mode.

7.4.4 Financial loss

Overlooked or inefficient inspections contribute to the increase in numbers in the percentage of undetected track defects. Some undetected defects may expand or increase to cause the track to 'soft' failure but the chance of this causing an actual train derailment is less likely. Let $X = \{X_d, X_b\}$ in which X_d and X_b, respectively, denote an event of rectifying a detected track defect and 'soft' failure. When X_b is performed as a result of missed opportunities of X_d, a difference in total cost between these two events, denoted by $\Delta C_{repair}(X_d, X_b)$ could represent a potential financial loss to profit attributed to undetected track defects. Based on information of repair cost models in Liu et al. (2014), C_u can be formulated as follows:

$$
\begin{aligned}
C_u &= \Delta C_{repair}(X_d, X_b) \\
&= (C_r^{X_b} - C_r^{X_d}) + \Delta C_{delay}(X_d, X_b) \\
&= (C_r^{X_b} - C_r^{X_d}) + (C_y^{X_b} - C_y^{X_d})
\end{aligned}
\tag{7.12}
$$

where C_r^j denotes a total direct cost of tamping operation per 200-m-long track section and is a summation of labour, administrative, logistic, machine, tools, machinery and materials cost. The superscript in C_r^j, $j \in X$ indicates the type of tamping maintenance. According to Andrade (2016), $\Delta C_{repair}\left(X_d, X_b\right)$ is estimated at €$4000\left(30000\ \text{SEK} \times 0.9\right)$. Similarly, the rule is applied to determine a difference in a delay cost due to temporary speed restrictions (TSR) between X_d and X_b, denoted by $\Delta C_{delay}(X_d, X_b)$. The base formula for calculating corresponding total cost of train delay for any event in X is given as follows (Andrade and Teixeira, 2016):

$$
C_{TSR}^{j} = c_o \sum_{i=1}^{n_{train}} t_{DM}^{i,j} t_{days}^{j} \tag{7.13}
$$

where $\sum_{i=1}^{n_{train}} t_{DM}^{i,j}$ is the total minutes of train delay per day due to TSR imposed for 200-m-long track section. Let t_{DM}^{i,X_d} and t_{DM}^{i,X_b} denote estimated delay minutes per affected train *i* for 200-m-long track section caused by planned and unplanned TSR, respectively. Whereas the value of t_{DM}^{i,X_b} is assumed to be 0.774 minutes = 0.0129 *h* for all train type, the value of t_{DM}^{i,X_d} can be calculated using the following equation (Andrade and Teixeira, 2016):

$$
t_{DM}^{i,X_d} = 0.2 \times 60 \left(\frac{1}{s_i} - \frac{1}{\min\left(s_i, s\left(z_j\right)\right)} \right) \tag{7.14}
$$

where s_i is the maximum permissible speed of train *i* in track section and $s(z_j)$ is the maximum permissible speed given the operation state z_j (i.e., 80, 120, 160 or 230 km/h. In terms of train delay, a unit cost (per minute), c_o is invariant with respect to type of tamping maintenance.

7.4.5 Severity

There are a few recommendations regarding the severity value as summarised by Konur et al. (2014). For example, Zhao et al. (2007) suggested the use of percentage of rail breakage that lead to train derailment. Statistics of broken-rail train derailments provided by Liu et al. (2011) would be a good resource of input for severity estimation. In the event of no changes in the track structure or track operation speed, one can use a constant value (i.e., less than 1) for severity but it might differ as suggested in the previous works (Barkan, Dick and Anderson, 2003; Podofillini, Zio and Vatn, 2006).

7.4.6 Expected number of 'soft' failure event

Section 5.3.1 provides evidences that 'soft' failures of track segments follow a nonhomogeneous Poisson process (NHPP) model. From the property of NHPP follows that the expected number of failures corresponding to defect type *r* in the interval (t_{j-1}, t_j) $E\left[N_f^r\left(t_{j-1}, t_j\right)\right]$ is

$$E\left[N_f^r\left(t_{j-1}, t_j\right)\right] = \int_{t_1}^{t_2} \lambda_f^r(t) dt, t_{j-1} \le t_j \tag{7.15}$$

where $\lambda_f^r(t)$ is the rate of occurrence of failure (ROCOF) corresponding to defect type r at time *t*. Unlike for homogenous Poisson process, the ROCOFs for NHPP are different at different time periods rather that being a constant. Obviously, one must derive the ROCOF for different defect type.

In the case of a track geometry related failure, a track segment meets the condition of a specified failure when track geometry deteriorates over a specified deterioration threshold. An application of several thresholds for track maintenance management is inherent the advantages of the delay time concept. In this concept, the failure process is treated as a two-stage process, with the first stage is being when a detectable defect arises i.e. from a new (or after maintenance work) to an initial point of defect (refereed to a normal operating stage). Upon the arrival of a defect, the object of interest such as a track segment or section spends an additional period of operational time and eventually fails if the defect is not rectified beforehand. In order to avoid late defect detection, therefore, the related risk is mitigated through an application of inspection and/or condition monitoring. In the context of track maintenance, an inspection vehicle is primarily assigned to rail lines to carry out an inspection to determine whether track is defective or not.

Consider that a track segment (or track section), which is expected to receive n_I inspections where the corresponding sequence of inspection times is $\langle t_1, t_2, \ldots, t_{m-1}, t_m \rangle$. In an arbitrary inspection interval $\left(t_{j-1}, t_j\right]; j \leq n_I$ the expected number of defects due to defect type r initiating within the interval can be represented by

$$E\left[N_d^r\left(t_{j-1}, t_j\right)\right] = \int_{t_{j-1}}^{t_j} \lambda_d^r\left(t\right) dt \tag{7.16}$$

where $\lambda_d^r\left(t\right)$ is the rate of occurrence of defect type r (ROCOD) at time t. As cited in Mahmoudi et al. (2017), it now follows from the property of NHPP that

$$\lambda_f^r(t) = \int_{t_{j-1}}^{t_j} \lambda_d^r(u)\, f(t-u)\,du \tag{7.17}$$

where $f(t-u)$ denote the probability density function that a failure occurs a time t arising from a defect at u. Using Eqn. (7.17) in Eqn. (7.15) then,

$$\begin{aligned} E\left[N_f^r(t_1,t_2)\right] &= \int_{t_{j-1}}^{t_j} \lambda_f^r(t)\,dt \\ &= \int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t_u} \lambda_d^r(v)\, f_r(v-u)\,dv\,du \\ &= \int_{t_{j-1}}^{t_j} \lambda_d^r(u)\, F_r(t_j-u)\,du \end{aligned} \tag{7.18}$$

where $F_r(y)$ denote the cumulative distribution function of delay time corresponding to defect type *r*. Eqns. (7.16)-(7.18) are for the perfect inspection i.e. a defect is detected upon a positive inspection at the end of inspection interval. However, not all inspections are perfect.

In the case of failures have occurred in $(t_{j-1},t_j]$ caused by defects that originated at any prior interval $(t_{k-1},t_k]$ $1 \le k \le j-1$ due to imperfect inspections the corresponding expected number of failures over the current interval is given by

$$N_f^r(t_{j-1},t_j) = \sum_{k=1}^{j-1}\prod_{i=k}^{j-1} 1-\beta_r(t_i)\int_{t_{k-1}}^{t_k} \lambda_d^r(u)\left[F_r(t_j-u)-F_r(t_{j-1}-u)\right]du \tag{7.19}$$

where $\beta_r(t);(0<\beta_r<1)$ is the rate function of defect detection. Initially, each defect type is assumed to have a specific rate function. Hence, summing Eqn. (7.18) and Eqn. (7.19) will give the total expected number of failures corresponding to defect type *r* in the interval $(t_{j-1},t_j]$ as follows:

$$N_f^r\left(t_{j-1},t_j\right)=\sum_{k=1}^{j-1}\prod_{i=k}^{j-1}1-\beta_r\left(t_i\right)\int_{t_{k-1}}^{t_k}\lambda_d^r\left(u\right)\left[F_r\left(t_j-u\right)-F_r\left(t_{j-1}-u\right)\right]du+\ldots$$
$$+\int_{t_{j-1}}^{t_j}\lambda_d^r\left(u\right)F_r\left(t_j-u\right)du \tag{7.20}$$

One requires $\beta\left(t\right)$ to operate Eqn. (7.20). As part of non-destructive testing, a track geometry inspection is attached to probability of detection (POD) function. Basically, POD is expressed as a function of SD value. While the defect size in terms of the percentage of railhead area %HA grows exponentially with the accumulated tonnage on track *t*, thus, the $POD\left(t\right)$ function can be written as;

$$POD\left(t\right)=a_1e^{b_1(t)}+a_2e^{b_2(t)} \tag{7.21}$$

where values of model parameter a_1, a_2, b_1 and b_2 may vary by the type of track geometry defect. Since tonnage of traffic accumulated since the last repair (or renewal) is easily estimated, thus, this research has decided to use Eqn. (7.21) to generate value for β for given *t*.

From the property of NHPP follows that the probability of *n* failures occurring in $\left(t_{j-1},t_j\right)$

$$P\left(N_f\left(t_{j-1},t_j\right)=n\right)=\frac{\left(E\left[N_f\left(t_{j-1},t_j\right)\right]\right)^n e^{-E\left[N_f\left(t_{j-1},t_{j-2}\right)\right]}}{n!}. \tag{7.22}$$

If $R\left(t_{j-1},t_j\right)$ is the expected reliability of the system in the interval $\left(t_{j-1},t_j\right)$ it can be represented that

$$R\left(t_{j-1},t_j\right)=P\left[\text{no failures in }\left(t_{j-1},t_j\right)\right]=e^{\left(-E\left[N_f\left(t_{j-1},t_j\right)\right]\right)} \tag{7.23}$$

Therefore, the expected number of 'soft failure' occurring to the track segment corresponding to defect type r during the observation period $\left(t_1, t_{n_I}\right)$ is $\sum_{j=1}^{n_I} N_f^r\left(t_{j-1}, t_j\right)$.

Consequently, the risk of opportunity loss due to undetected defects in the period $\left(t_1, t_{n_I}\right)$ is given by

$$Risk\left(t_1, t_{n_I}\right) = C_u \sum_{r=1}^{R} S_r \sum_{j=1}^{n_I} N_f^r\left(t_{j-1}, t_j\right) \tag{7.24}$$

7.4.7 Numerical examples

Railway tracks are defined on directed graph $G = (V, A)$. The set of vertices $V = S \cup \mathrm{O}$ corresponds to the set of stations $S = \{s_1, \ldots, s_n\}$ and the depot O. The arc set $A = \{a_{12}, \ldots, a_{ij}\}$ where $i \neq j$ includes all variable connections between vertices in Q with $|A|$ represents the total number of arcs. Fig. 7.5 depicts a graph representation of (mini) network which has $n = 7$ nodes and $|A| = 12$ arcs considered in this section. Arcs of the sub-network (excluding from/to a depot) were divided into two sets; Set *A* and Set *B*, based on the inspection requirement for track geometry defects, see Table 7.2. In this example, the effects of track length and track layout were not considered when calculating the value of rescheduling.

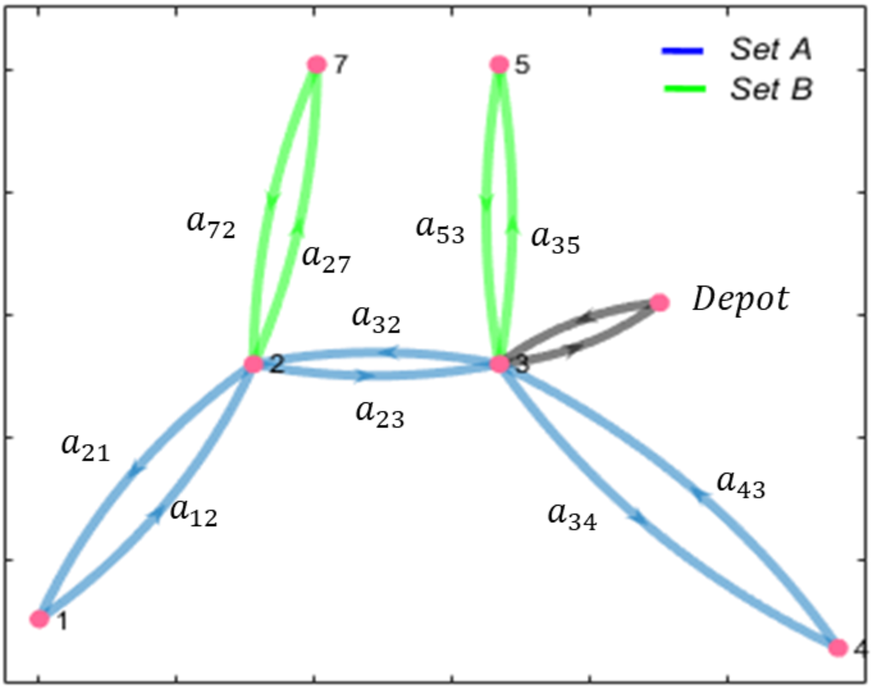


Fig. 7.5 Graph transformation of typical double rail track line. Effects of track geometry and elevation as well as condition of relevant track components are left out from this example.

Table 7.2 Arc classification for *G*.

	Set*A*	Set*B*
Annual accumulated tonnage (in MGT)	15-35	> 35
(Minimum) inspection per year	Twice	Four times
Arc member	$\{a_{12}, a_{21}\} \cup \{a_{23}, a_{32}\} \cup \{a_{34}, a_{43}\}$	$\{a_{35}, a_{53}, a_{27}, a_{72}\}$

In respect to track geometry inspection, a track-recording car carries out geometry measurements by traversing the targeted tracks at a predefined speed. For an individual TRC trip (it will be called trip in the remainder of the text), a set of arcs will be inspected during a track possession interval despite some track sections not yet requiring inspection i.e. the aggregation approach. Every trip begins and ends at a depot. For the network in Fig. 7.5, there are two default trips; $O - SetA - O$ and $O - \{a_{23}, a_{32}\} - SetB - O$. A collection of the N=13 trips was designed to be performed in the observation period $T_M = 2$ years and is treated as I, as shown in Table 7.3.

Table 7.3 Presumed track inspection schedule applied on the network G.

Trip		**Arc**				
No.	Date, d_i	$\{a_{12}, a_{21}\}$	$\{a_{34}, a_{43}\}$	$\{a_{23}, a_{32}\}$	Depot	Set B
$i = 1$	18/10/18[1]	1	1	1	1	0
2	27/12/18	0	0	1	1	1
3	06/01/19	1	1	1	1	0
4	27/03/19	1	1	1	1	0
5	26/05/19	0	0	1	1	1
6	15/06/19	1	1	1	1	0
7	03/09/19	1	1[2]	1	1	0
8	23/10/19	0	0	1	1	1
9	22/11/19	1	1	1	1	0
10	10/02/20	1	1	1	1	0
11	21/03/20	0	0	1	1	1
12	30/04/20	1	1	1	1	0
13	19/07/20	1	1	1	1	1[3]

[1] An observation period starting from 31/07/18 until 19/07/20
[2] Effected arcs due to a disruption
[3] Final inspection is undertaken on all arcs

In the event of no disruptions occur within T_M, risk score of losing an opportunity to repair defective track due to undetected defect for every arc in G under I, is displayed in Fig. 7.6. The risk calculation was performed using Eqn. (7.24) provided in and based on following considerations:

- Condition of each rail section is assumed to be nearly new (resulting from a perfect maintenance) at the beginning of T_M
- No maintenance or repair works are taking place during an observation period.

- A sequence of arcs in any trip has no effect in the risk calculation. This is based on the fact that a trip is completed in less than 24 hours due to working-hour regulations.
- Values of parameters for the probability distribution that represent the occurrence of defect and track-geo failure were extracted from Sections 5.2.3 and 5.2.4.
- The value model described in Section 7.4.1 was implemented in MATLAB (version R2010b).

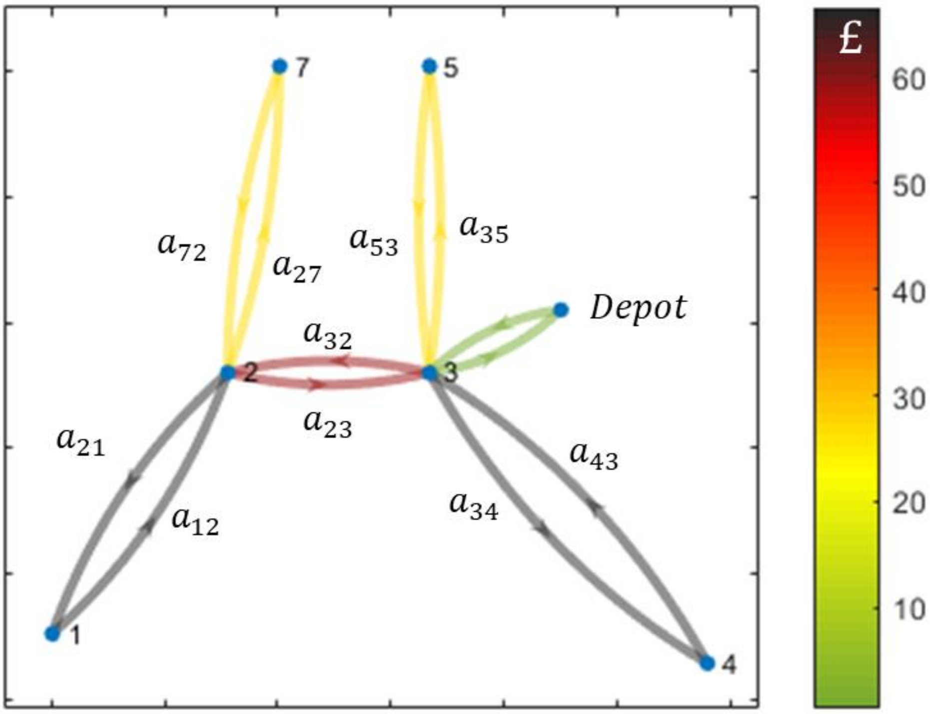


Fig. 7.6 Risk score over *G* under *I*. A more frequent inspection resulted lower score in both a_{23} and a_{32} compared to other arcs in Set *B*.

Consider that the 5th inspection of Set *A* scheduled on (03/09/2019) cannot be performed on both a_{34} and a_{43} due to a disruptive events. This situation implies an adjustment on the corresponding trip which lead to the removal of a_{34} and a_{43} from the trip no.7. To cope with disruptions, there will be at least possible three rescheduling strategies that may be implemented on the affected arcs.

1. *Strategy1*: No action will be performed and wait for the next inspection scheduled for Set*A* on (22/11/2019).
2. *Strategy2*: Insert the affected arcs into the nearest trip which is trip no.8. This strategy allows the affected rail sections to be inspected together with a different set of rail section. As a result, an adjustment is required for trip no.8.
3. *Strategy3*: Shift trip no.9 (the 6th inspection times of Set*A*) to the left on time axis; i.e., closer to trip no.6. This strategy will change the inspection policy from periodic to non-periodic but is only applied to the 6th and 7th inspection of Set*A*. This is strategy associated with the outputs of iNARXNN. A new interval between 1) the 4th and 6th inspection v_{46}, and 2) the 6th and 7th inspection v_{67} of Set*A* can be determined using the ratio rule given as $\frac{v_{46}}{v_{47}} = \frac{\hat{d}_6 - d_4}{d_7 - d_6}$ where $\hat{d}_6$ is a new inspection date of the 6th inspection.
4. For simplicity, $\hat{d}_6$ was rescheduled in the middle between d_4 and d_7. This assignment not only allocates a_{34} and a_{43} with a new inspection interval, but also every arc in Set*A*; $v_{46} = 120$ days instead of 80 days. Importantly, this decision satisfies the condition of inspection interval in Table 7.2. However, other trips,(particularly trip no.8) are not affected by this strategy.

Fig. 7.7 depicts the effects of possible rescheduling strategies in order to cope with disruptions in the RIP in Table 7.3. Values in the figure were computed using Eqn. (7.3). As expected, the risk from a missed opportunity to repair defective rails increases in all proposed strategies. *Strategy2*, among all possible strategies, possesses a little increment in the risk about 2.0 units but it incurs cost, around £ 1.30. Despite penalty administrative fees were also imposed on both *Strategy1* and *Strategy3*, the strategies generate a refund (shown by a negative reschedule cost). This can be explained by the fact that no reduction in the number of inspection was offered in *Strategy2* which could be applied to reduce the penalty cost. Comparing *Strategy3* and *Strategy1*, it appears that the size of refund from trip cancellation has an important role in reducing the penalty administrative fees.

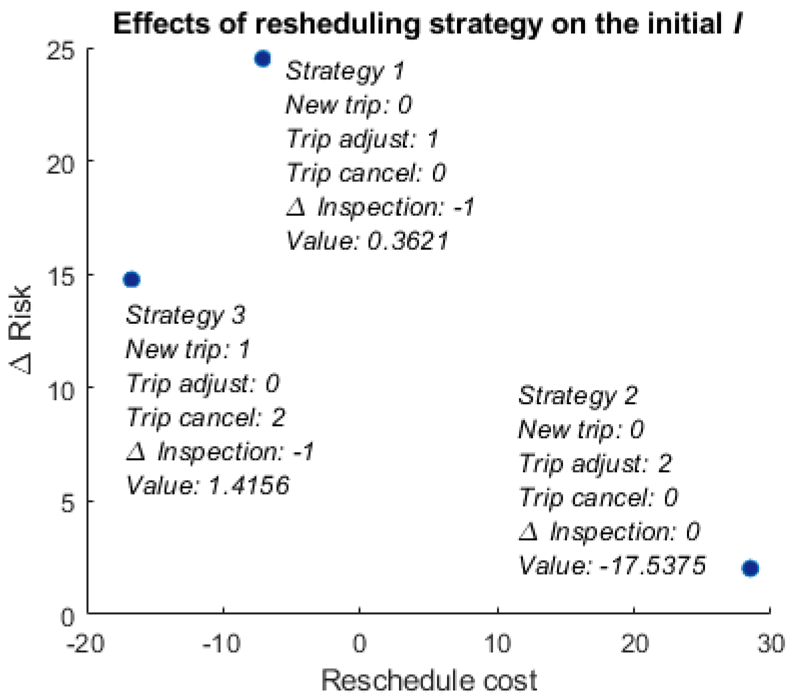


Fig. 7.7 Value of rescheduling from three strategies.

The one with the lowest value of rescheduling is *Strategy2* in which a decision maker has to spend £ 0.4406 for a unit increment in the risk of missed opportunity. Either *Strategy1* or *Strategy3* creates positive decision value. Comparison results show that value of rescheduling is influenced by the number of inspection reduction in *I* and the selection of new (temporary) inspection interval for v_{46} and v_{67}. Keeping the first factor unchanged, sensitivity of the v_{46}/v_{67} ratio to value of rescheduling was investigated. In Fig. 7.8, an evolution of value of rescheduling of *Strategy3* over a range of v_{46}/v_{67} ratio values is explained by a Gaussian function. Clearly, an optimal decision in regard to *Strategy3* is to shift trip no.9 to the left by 20-30 days i.e. 50-60 days after a disruption.

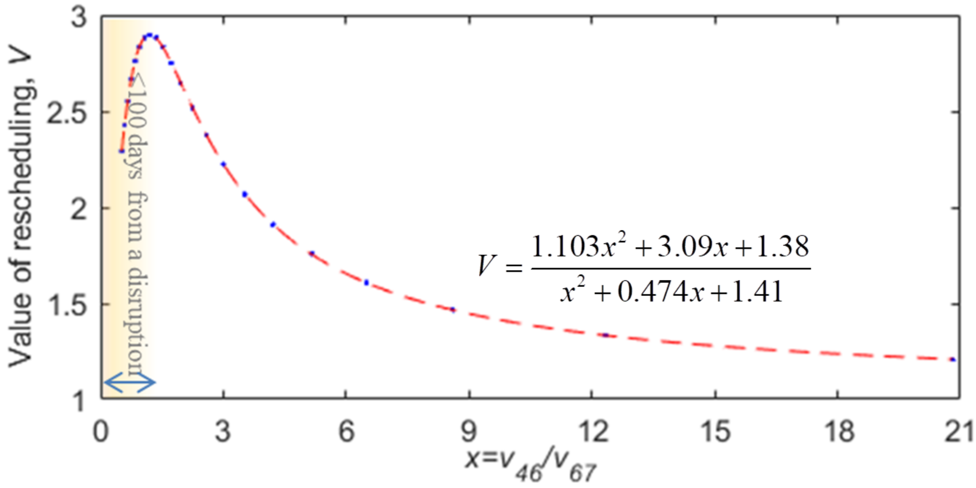


Fig. 7.8 Effects of an interval decision on the value of rescheduling associated with *Strategy3*.

7.5 Summary

In this chapter, a value model defined by the ratio of rescheduling costs to an expected benefit from restoring resilience is presented. While the difference between two inspection costs determines the costs incurred for rescheduling, the economic benefit from rescheduling is proportional to the reduction in the risk of unplanned maintenance. Rescheduling costs increased steadily with the delay minutes in the passenger journey. As expected, the main contribution to rescheduling cost is track possession cost, which is a function of marginal revenue effect.

Different rescheduling strategies to restore resilience to a disrupted track inspection were evaluated in this study using the proposed value model. The results for all are encouraging, but the rescheduling strategy that incorporates predicted track measurements from iNARXX generated the highest value. This finding provides convincing evidence as to the benefits of iNARXNN in practice.

CHAPTER 8 CONCLUSION

When it comes to disruption management, there is not much RIM can do to predict when and/or where disruption will occur. Furthermore, most disruption events are low probability, which may be why many organisations—including those in the railway maintenance sector—make less effort toward disruption management. Crew loss, machine breakdowns, and special inspection requests can potentially disrupt the execution of track inspection in various ways. Similarly, the formulation of the optimisation problem for the track inspection plan does not always consider uncertainties. These events are assumed to have a low probability of occurrence, but their presence can negatively affect on-going track plan operations. Also categorised as high-impact, low-probability events, they are sources of disruptive risk for which an additional cost is required for risk mitigation. As track inspection activities are performed to avoid infrastructure failures in the railway business, it is vital to make sure the plan survives disruptive events, or that resilience actions are taken to mitigate consequences/losses that may arise.

In this research, the aim is to develop response and recovery models showcasing resilience functions in a disrupted periodic plan for track geometry inspection. All research questions are answered under a specific theme; disruptions in track inspection plans and development and valuation of resilience models for disrupted inspections. Based on the results obtained in Chapters 5–7, a discussion of the findings is conducted to determine whether the proposed hypotheses are supported. It is then determined whether the research objectives have been met. Finally, the contributions of the study are discussed based on theoretical, methodological, and practical approaches, and suggestions for future research are offered.

8.1 Research questions regarding the issue of disruption in track inspection plan

Three research questions are raised in this study to provide a solid understanding of the property of track inspection plans and sources of disruptions.

The first research question is to what extent the scheduling approach is applied to construct a periodic schedule for track inspection. Prior research suggests that a periodic track inspection plan benefits rail infrastructure companies in terms of track availability and safety and inspection costs while satisfying operational, technical, and business constraints. Additionally, the allocation of company resources such as equipment, inspection vehicles, and manpower is well organised, and track possessions are arranged effectively. Section 2.2.1 showed that track inspection plans are conveniently formulated and solved under an optimisation model. The reviewed literature also revealed that the existence of complex environments in association with train and track operations has posited some model inputs with a high degree of uncertainties being neglected when formulating a track inspection problem. Trading the model complexity of an inspection plan with uncertainties exposes the inspection plan to disruption risks. When analysing the formulation of a track inspection plan, sources of disruption links to uncertainties, possibly due to workers’ strikes, climate-related extreme weather events such as flash floods and landslides, terrorist attacks, train recording car breakdowns, and special inspection requests. Because of the nature of the occurrence of disruptions, previous studies have emphasized on focusing on reducing consequences of disruption risks rather than lowering the probability of occurrence.

The second research question is what characteristics determine the uniqueness and/or similarities of the track inspection schedule when compared to other periodic schedules. The reviewed literature revealed that the condition of the rail track is inspected through various inspection tasks, which can range from human visual inspection to machine-aided measurement. Unlike road inspection plans, an inspection plan for rail tracks can only be endorsed after the disclosure of freight planning, passenger timetable, and major track possessions. This policy states that track possession for inspections is determined last when tracks are unattended by both passengers and freight trains. A common strategy used to prepare a track inspection plan under constraints is to formulate it as a constrained optimisation problem. While analysing previous studies (in Section 2.2.1), it was fond that various objective functions have been proposed, such as track possession costs, total inspection times, inspection and failure costs, delays in train service, train derailment cost, and total deadhead distance.

The third research question is how uncertainties in an operation of TIP can be determined by formulating the inspection problem as a mathematical optimisation problem. The influence diagram developed for the formulation of optimisation problems for track inspection (Fig. 2.1) captures the links among (potential) input variables, structure, and decision values. Any input variables that have a low probability of occurrence and a high degree of uncertainty are highly likely to be excluded from the model. When analysing the characteristics of each input variable, it was found that workers' strikes, terrorist attacks, climate-related events such as landslides, machine breakdowns, track re-classification, and special inspection requests are uncertainties in the optimisation model for track inspection. It is worth mentioning that excluding uncertainties from the model formulation corresponds to reduction of model complexity, defined in Eqn. 2.1. Disruption risks linked to these uncertainties may trigger not only economic losses but also reputation damage.

8.2 Research questions regarding the development of resilience models

The fourth research question is to what extent the components of the resilience concept are meaningful to an existing risk management framework, particularly when dealing with low-probability events. Disruptions in a track inspection plan are beyond the control of track engineers. This unique characteristic is the reason that most of them prefer to deal with the consequences associated with disruptive events after their occurrence. The reviewed literature on disruption management provided evidence that it is necessary to understand decision makers' attitudes towards risk to persuade them to make additional investments in track inspection. In practice, it can be convenient to offer flexibility in decision-making regarding how and when to respond to disruptive events subject to real-time evaluations that suit the decision makers' degree of risk aversion. This was found to be fulfilled under the resilience concept, in which the performance of an inspection plan, such as availability of resources, safety, or train delay minutes, were defined as the resilience metric. Depending on the decision maker's preferences, deterioration in the resilience metric can be managed using different resilience strategies at different system phases: steady phase, disruptive phase, recovery phase, and post-recovery phase (an extensive discussion was conducted in Section 2.4.1). Interestingly, there is the possibility of having a newly attained steady state for a recovered inspection plan that is lower or higher than the initial steady state.

The fifth research question is which platform is reliable to develop a prediction model for a temporal data series. A large body of evidence suggests that a series of track measurements is significant to trace the evolution of track geometry condition over tonnage (or time). In this study, historic track measurements were manipulated to predict 'missing' track measurements in the event of the disruption of a track inspection. This manipulation can be performed because track engineers have the privilege to access the RIM's data repository for past track measurements. Taking advantage of the fact that there is no requirement to model the evolution of track conditions explicitly, a non-linear autoregressive model and a neural network were jointly employed to develop a prediction model called iNARXNN. To obtain a prediction model with good prediction performance while generalize well, a regularization method was adopted in the network's learning process. The prediction model, which is fully described in Chapter 6, was used to predict longitudinal levels for both left and right rails.

The sixth research question is what the requirements are for incorporating a rescheduling framework into the recovery phase of a resilient system. Knowledge of resilience phases is essential for not only monitoring performance loss but also estimating individual resilience costs. Because recovery capability, knowledge of which is required for cost calculation, is assessed during the recovery phase, the rescheduling must explicitly incorporate costs incurred for rescheduling when providing the configuration for a disrupted inspection schedule. The configuration was determined through an analysis of all dimensions of rescheduling: environment, strategy, policy, and method with respect to the nature of the track inspection schedule. The findings showed that for a periodic schedule in track inspection, rescheduling was conveniently performed upon the occurrence of disruptive events. This was mainly caused by uncertainties that the schedulers did not consider when formulating an initial schedule. Therefore, a preferable method of rescheduling a disrupted track inspection is the recovery method known as partial rescheduling. This method aims to keep the degree of deviation from the existing schedule as low as possible when the part of the schedule that is directly or indirectly affected by disruption is rescheduled. Compared to total rescheduling, this recovery method takes less computational effort to assess recovery capability in terms of costs of rescheduling. The cost increment should be kept as low as possible when performing partial rescheduling because it contributes to overall resilience costs.

The seventh research question is what time-varying value function should be used to compare the costs involved and the expected benefits of re-planning an affected inspection plan. The proposed value function was formulated on the basis of the resilience metric to quantitatively assess resilience restoration to a disruption-affected inspection plan by comparing costs involved and the expected benefits of rescheduling. The total cost incurred for rescheduling was formulated as the sum of differences of the track inspection cost and penalty cost. The reviewed literature revealed that the time-varying effect of inspection in a track inspection cost model is represented by the track possession cost. Further, the cost drivers attached to the direct cost of vehicle-based inspection, such as the cost of operating the track recording car, are influenced by the time band during which an inspection is performed. Consequently, this effect is transferable to a penalty cost because it is the proportion of the direct inspection cost. Through an analysis of decision makers' attitude towards safety-related investment (in Section 2.3), it was concluded that a track engineer expects a reduction in the risk of the missing opportunity to repair a defective track in the track inspection reschedule. The potential loss of profit due to a missed opportunity in relation to undetected track defects was defined by the difference in the total tamping cost between planned and unplanned tamping operations. In addition to the total direct cost of tamping, a time-dependent delay cost model was used to explicitly differentiate the effects of planned and unplanned repairs. The results of the tamping data analysis presented in Section 5.3.1 permitted the value function to use the property of the NHPP model to estimate the number of failures occurring in a track segment corresponding to longitudinal-level defects occurring during an inspection interval.

The eighth research question asks what the trade-offs are between the level of managerial decisions and the model values. Restoring resilience to a disrupted plan for track inspection requires cost-restoration-level trade-off. As safety regulators in track operations, track managers are risk averse, which indicates that the affected plan is most likely restored beyond its performance level before disruptions can occur. This in turn implies that the denominator of the value model in Section 7.4 must be greater than zero. Otherwise, the recovery action is irrelevant for the disrupted track inspection plan even if the action contributes to a large reduction in inspection costs.

8.3 Findings

This section describes the major findings stemming from the development and evaluation of resilience models, including the responses to the hypotheses studied in this research.

Rail infrastructure managers are often risk averse rather than risk neutral when confronted with the consequences of disruptions. However, the additional investments required to restore resilience to a disrupted track inspection make them search for a solution they can afford. To discuss affordability with respect to the nature of unexpected events, this study examined decision makers’ utility towards disruption risks by determining the tipping point of risk aversion parameters, as evidenced by the conclusion to Hypothesis (1):

The tipping point of affordability capability in the development of the response model can be determined from an expected utility theory.

This hypothesis was confirmed. From the analysis of the risk aversion model, it was concluded that the affordability of restoring resilience to a disrupted plan is determined by the lowest value of risk aversion that a track manager assigns in compliance with the track’s safety and operability. This indicated that the expected benefits from any proposals for a resilient inspection plan should outweigh the probability of unplanned maintenance. The results of the statistical analysis of tamping maintenance records suggested that the delay time in track maintenance follows Weibull distribution. These results demonstrated the adequacy of deterioration analysis for examining the interaction between a deterioration process and failure time prediction.

A track inspection plan faces performance deterioration upon the occurrence of disruption. Framing state transitions in the context of system resilience allows for the resilience metric to be characterised by the evolution of the state of a disrupted plan over time. This allows track managers flexibility in deciding which tracks should be restored and when, as evidenced by the conclusion to Hypothesis (2):

The use of time-varying resilience metrics as risk-acceptance criteria provides flexibility to decision makers about whether to immediately implement or to delay the formulation of a response to a disrupted inspection plan, depending on the real-time evaluation of a company's strengths and limitations.

This hypothesis was supported. The results from the scenario analysis (in Section 7.4.2) revealed that a steady increment in rescheduling cost is observed when the delay minute per passenger journey is increased. The effect is more pronounced when a rescheduling decision is taken during the morning off-peak time band. This represents purely the effect of track possessions.

Dealing with low-probability events demands specific innovations in affordable risk-resilience response strategies. Fortunately, an affordable, but reliable, strategy was developed through a smart partnership between artificial intelligence and a spatiotemporal prediction model, as evidenced by the conclusion to Hypothesis (3):

An application of artificial intelligence for spatiotemporal data prediction is reliably proven to respond to the absence of track measurement data.

This hypothesis was confirmed. According to the comparison results of predicted and actual longitudinal levels obtained from the correlation plot and PSD analysis, the iNARXNN provided excellent predictions for a disrupted inspection plan. Additionally, post-processing evaluation powered by the Bayes linear method verified its readiness for track quality assessment. No significant variation was found in the value of information. The predicted track measurements substantiated the effectiveness of the proposed prediction model. Statistical tests were conducted, with the results indicating that the deterioration rate of track quality followed an inverse Weibull distribution.

The flexibility of restoring resilience to a disrupted track inspection plan enables the track manager to reconfigure the original inspection interval. This was confirmed by the conclusion to Hypothesis (4):

Extending an inspection interval with or without a reduction in inspection frequency creates positive value in an inspection scheduling system.

A sensitivity analysis was used to determine how the value of rescheduling reacts to a change in inspection interval. The optimal value of rescheduling can be achieved by extending an initial inspection interval by 50 to 60 minutes after disruption. However, these findings cannot be universally applied to all disrupted track inspections.

Restoring resilience will result in the allocation of a certain amount of time and resources to ensuring that the predetermined schedule can adapt to the impact of disruption with minimal actions. The situation becomes more challenging for decision makers when several potential rescheduling strategies are presented for comparison. Thus, resilience trade-off is vital in ensuring appropriate participation of track managers in managing disrupted track inspection, as argued by Hypothesis (5):

The degree of changes in the established TIP depends upon the managerial decision preferences in a cost-benefit trade-off analysis.

This hypothesis was supported. According to the comparison results for the different rescheduling strategies introduced in Section 7.4.7, the value of rescheduling is positively related to the number of inspection reductions in the original schedule and to the selection of a new (temporary) inspection interval. The reduction in the number of inspections generates a refund, which reduces the penalty cost. This interesting pattern also appeared after an adjustment of the inspection interval.

8.4 Contributions

Recent developments in respect to railway operations have shown that disruptions follow an increasing trend in the number of occurrences, and these occurrences manifest in new and various ways. This is a strong incentive for infrastructure managers to rethink the importance of the resilience model to complement conventional methodologies for planning track inspection. The risk-return relationship is a core aspect of business decision-making. Therefore, the benefits of the resilience model are formulated in terms of the risk aversion of low-probability events and of monetary value. In a specific situation, the resilience model can recommend a reduction of the number of inspections without compromising track safety requirements that should be presented over a period of time. In parallel, infrastructure managers can enjoy substantial benefits from the reduction, such as increased track access for freight train services and the elimination of some contributors to train delays.

The proposed resilience model potentially adds more value to the money invested in track geometry inspection. This argument is, however, not enough to motivate infrastructure managers to become more risk averse about disruption management. It is still a low-probability risk; thus, more motivation is needed to maximise the desire to invest. In this research, an index of which is derived from the evolution of uncertainty in model parameters, which reflect the potential of detecting sudden shift event is also proposed. The higher the index, the greater the motivation for a decision maker to reach the maximum limit for risk aversion. Interestingly, this situation is allowed to occur in a natural way and is treated as a disruption to the on-going track inspection schedule.

Recent statistics show that tracks associated with long inspection intervals (and low track capacity utilisation) has the highest rate of train derailment caused by track geometry defects (Liu, Saat and Barkan, 2012). These tracks are typically constrained by physical investments owing to relatively small revenue (Andersson, 2011). In the event of track re-classification, incorporating iNARXNN in place of an on-site track inspection could be a low-cost strategy to reduce this particular type of risk. The reader should bear in mind that this proposal is not a replacement for the practice of track inspection but rather an innovation to combat low-probability risk in managing railway track, such as train derailment and unplanned maintenance, particularly on low priority tracks.

Updating model parameters as often as required can be tedious. Therefore, the model should be presented as beneficial in overcoming exponential computational costs. This element is among the reasons why executives/organisations hesitate to invest in soft-computing innovations, leading us to propose the model in practical terms (Panda and Gupta, 2014). The Bayes linear method used in the resilience model takes into account known information with the confidence in the parameters of the track degradation model that characterise track geometry condition over time (or tonnage). This effort can be seen as a short-term initiative, i.e. taking place within an inspection interval, to averse from confronting with a sudden shift in track deterioration.

The resilience model proposed in this study has two components: iNARXNN and a risk-aversion-based value measure, respectively, for predicting "missing" track measurements and recovering from disruptions in real time. With these capabilities, a track engineer who is responsible for making a decision at an operational level is provided with a logical explanation of the valuation of model outputs based on information model carries. Interestingly, no capital investment for installing, replacing, or purchasing any new equipment is required to embed the proposed model into the planned track inspection plan.

Finally, the development of the resilience model, iNARXNN, and the response model promotes data innovation in railway management maintenance. The use of freely available track measurement data reduces the time and cost of resilience.

8.5 Limitations and further research

The post-processing phase supports outputs of iNARXNN with a smart verification system powered by the Bayes linear method. This Bayesian-type method, which relaxes the need for the heavy use of probabilistic computations, can be used perform a series of checks in terms of uncertainty reduction of outputs of iNARXNN. This feature determines how much the predicted track measurements would deviate from the performance of a track deterioration model. For further exploration using the same data, I extend variance learning from a static linear combination of observations to multiple linear combinations. This might create a more prolonged process due to evaluations of variance and covariance between those linear combinations. Additionally, the measures used in this study should be weighted based on class type and location of rail tracks. By having weighting functions, the relative importance of each quantity can be represented more adequately while taking the complexity of decision-making into practical consideration. Further, a performance comparison between the two types of Bayesian approaches to assessing uncertainty propagation should be conducted to demonstrate practicality when dealing with large sizes of components.

In the future, I aim to demonstrate the proposed methods with a larger railway network size. This might help explain how a cascade effect can potentially originate from localised rescheduling decisions. As the size of the problem increases, the time taken to execute the proposed methodology is also expected to increase. To accelerate the process of selecting the rescheduling strategy, the original model is necessary to formulate track inspection rescheduling as an optimisation problem. As different track sections are attributed to specific operational and safety constraints, the related multi-criteria optimisation model can be expanded to a constrained optimisation model. Additionally, the process of determining an optimal interval ratio in a rescheduling strategy involving iNARXNN's outputs can be systematically carried out by applying the utility function attached to Fig. 7.8. Finally, future works will address the interconnected risks of other rail defects.

LIST OF REFERENCES

Abadie, L. M., Galarraga, I. and de Murieta, E. S. (2017) 'Understanding risks in the light of uncertainty: low-probability, high-impact coastal events in cities', *Environmental Research Letters*, 12(1), p. 14017. doi: 10.1088/1748-9326/aa5254.

Acuña-Agost, R. (2009) *Mathematical modeling and methods for rescheduling trains under disrupted operations*. Université d'Avignon. Available at: https://tel.archives-ouvertes.fr/tel-00453640.

Adjetey-Bahun, K., Birregah, B., Châtelet, E. and Planchet, J. L. (2016a) 'A model to quantify the resilience of mass railway transportation systems', *Reliability Engineering and System Safety*, 153, pp. 1–14. doi: 10.1016/j.ress.2016.03.015.

Adjetey-Bahun, K., Birregah, B., Châtelet, E. and Planchet, J. L. (2016b) 'Railway transportation system's resilience: integration of operating conditions into topological indicators', in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*. Istanbul, 25-29 April: IEEE, pp. 1163–1168. doi: 10.1109/NOMS.2016.7502981.

Al-Nazer, L., Raslear, T., Patrick, C. and Gertler, J. (2011) *Track inspection time study*. Washington, D.C: FRA, U.S. Department of Transportation. Available at: www.fra.dot.gov/us/content/185.

Alwadood, Z., Shuib, A. and Hamid, N. A. (2012) 'A review on quantitative models in railway rescheduling', *International Journal of Scientific & Engineering Research*, 3(6), pp.

1–7.

Andersson, M. (2011) 'Marginal cost of railway infrastructure wear and tear for freight and passenger trains in Sweden', *European Transport - Trasporti Europei*, (48), pp. 3–23.

Andrade, A. R. and Teixeira, P. F. (2011) 'Uncertainty in rail-track geometry degradation: Lisbon-Oporto line case study', *Journal of Transportation Engineering*, 137(3), pp. 193–200. doi: 10.1061/(ASCE)TE.1943-5436.0000206.

Andrade, A. R. and Teixeira, P. F. (2013) 'Optimal maintenance and renewal strategy due to rail track geometry', in Macario, R. (ed.) *13th World Conference on Transport Research (WCTR)*. Rio de Janeiro, July 15-18: Institute for Transport Studies.

Andrade, A. R. and Teixeira, P. F. (2015) 'Statistical modelling of railway track geometry degradation using Hierarchical Bayesian models', *Reliability Engineering and System Safety*, 142, pp. 169–183. doi: 10.1016/j.ress.2015.05.009.

Andrade, A. R. and Teixeira, P. F. (2016) 'Exploring different alert limit strategies in the maintenance of railway track geometry', *Journal of Transportation Engineering*, 142(9), pp. 1–9. doi: 10.1061/(ASCE)TE.1943-5436.0000867.

Ayyub, B. M. (2015) 'Practical resilience metrics for planning, design and decision making', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(3), p. 4015008. doi: 10.1061/AJRUA6.0000826.

Bae, S. J., Kuo, W. and Kvam, P. H. (2007) 'Degradation models and implied lifetime distributions', *Reliability Engineering and System Safety*, 92(5), pp. 601–608. doi: 10.1016/j.ress.2006.02.002.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M. and White, P. (2004) 'The demand for public transport: a practical guide', *Transport Policy*, 13, pp. 295–306. doi: 10.1016/j.tranpol.2005.12.004.

Barkan, C. P. L., Dick, C. T. and Anderson, R. (2003) 'Railroad derailment factors affecting hazardous materials transportation risk', *Transportation Research Board of the National Academies*, 1825, pp. 64–74.

BBC UK (2015) *King's Cross and Paddington christmas rail chaos affected 115,000 passengers*. Available at: http://www.bbc.co.uk/news/uk-england-london-31435821 (Accessed: 18 February 2006).

Beddington, J. (2012) 'Blackett review of high impact low probability risks'. Edited by Government Office of Science. United Kingdom: Government, The UK, p. 46.

Bedford, T., Denning, R., Revie, M. and Walls, L. (2008) 'Applying bayes linear methods to support reliability procurement decisions', in *2008 Annual Reliability and Maintainability Symposium*. Las Vegas, 28-31 January: IEEE, pp. 341–346.

Bell, S. A. and Percy, D. F. (2012) 'Modelling uncertainty in preventive maintenance scheduling', *Quality and Reliability Engineering International*, 28(6), pp. 604–615. doi: 10.1002/qre.1409.

Berawi, A. R. B. (2013) *Improving railway track maintenance using power spectral density (PSD)*. University of Porto. Available at: https://sigarra.up.pt/feup/en/pub_geral.pub_view?pi_pub_base_id=27042.

Bergquist, B. and Söderholm, P. (2015) 'Data analysis for condition-based railway infrastructure maintenance', *Quality and Reliability Engineering International*, 31(5), pp.

773–781. doi: 10.1002/qre.1634.

Birkie, S. E. (2016) ‘Operational resilience and lean: in search of synergies and trade-offs’, *Journal of Manufacturing Technology Management*, 27(2), pp. 185–207. doi: 10.1108/JMTM-07-2015-0054.

Blann, D. R. (2013) *Maximizing the P-F interval through condition-based maintenance*. Available at: http://www.maintworld.com/Applications/Maximizing-the-P-F-Interval-Through-Condition-Based-Maintenance (Accessed: 28 August 2019).

Blos, M. F. and Miyagi, P. E. (2015) ‘Modeling the supply chain disruptions focusing on the supply chain interdependencies’, in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*. Cambridge, 24-25 July: IEEE, pp. 604–609. doi: 10.1109/INDIN.2015.7281803.

Brugnach, M., Pahl-Wostl, C., Lindenschmidt, K. E., Janssen, J. A. E. B. and Filatova, T. (2008) ‘Complexity and uncertainty: rethinking the modelling activity’, in Jakeman, A. J., Voinov, A. A., Rizzoli, A. E., and Chen, S. H. (eds) *Environmental modelling, software and decision support: state of the art and new perspectives*. Amsterdam: Elsevier, pp. 49–68.

Bruneau, M. *et al.* (2003) ‘A framework to quantitatively assess and enhance the seismic resilience of communities’, *Earthquake Spectra*, 19(4), pp. 733–752. doi: 10.1193/1.1623497.

Budai, G., Huisman, D. and Dekker, R. (2006) ‘Scheduling preventive railway maintenance activities’, *The Journal of the Operational Research Society*, 57(9), pp. 1035–1044. Available at: http://www.jstor.org/stable/4102318.

Burden, F. and Winkler, D. (2008) ‘Bayesian regularization of neural networks’, in

Livingstone, D. J. (ed.) *Artificial Neural Networks: Methods and Protocols*. 1st edn. Sandown, UK, pp. 25–44.

Cacchiani, V., Hemmelmayr, V. C. and Tricoire, F. (2014) ‘A set-covering based heuristic algorithm for the periodic vehicle routing problem’, *Discrete Applied Mathematics*, 163, pp. 53–64. doi: 10.1016/j.dam.2012.08.032.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L. and Wagenaar, J. (2014) ‘An overview of recovery models and algorithms for real-time railway rescheduling’, *Transportation Research Part B: Methodological*, 63, pp. 15–37. doi: http://dx.doi.org/10.1016/j.trb.2014.01.009.

Caetano, L. F. and Teixeira, P. F. (2013) ‘Availability approach to optimizing railway track renewal operations’, *Journal of Transportation Engineering*, 139(9), pp. 941–948. doi: 10.1061/(ASCE)TE.1943-5436.0000575.

Camerer, C. F. and Kunreuther, H. (1989) ‘Decision processes for low probability events: policy implications’, *Journal of Policy Analysis and Management*, 8(4), pp. 565–592. doi: 10.2307/3325045.

Carretero, J. *et al.* (2003) ‘Applying RCM in large scale systems: a case study with railway networks’, *Reliability Engineering and System Safety*, 82(3), pp. 257–273. doi: 10.1016/S0951-8320(03)00167-4.

Cha, E. J. and Ellingwood, B. R. (2012) ‘Risk-averse decision-making for civil infrastructure exposed to low-probability, high-consequence events’, *Reliability Engineering and System Safety*, 104, pp. 27–35. doi: 10.1016/j.ress.2012.04.002.

Cha, E. J. and Ellingwood, B. R. (2013a) ‘Acceptance of risk due to competing wind and earthquake hazards', in Deodatis, G., Ellingwood, Bruce R., and Frangopol, D. M. (eds) *Safety, reliability, risk and life-cycle performance of structures and infrastructures*. London: CRC Press, Taylor & Francis Group, pp. 1221–1226.

Cha, E. J. and Ellingwood, B. R. (2013b) 'Seismic risk mitigation of building structures: the role of risk aversion', *Structural Safety*, 40, pp. 11–19. doi: 10.1016/j.strusafe.2012.06.004.

Cha, E. J. and Ellingwood, B. R. (2014) 'Attitudes towards acceptance of risk to buildings from extreme winds', *Structure and Infrastructure Engineering*. Taylor & Francis, 10(6), pp. 697–707. doi: 10.1080/15732479.2012.758642.

Chang, H., Liu, R. and Li, Q. (2010) 'A multi-stage linear prediction model for the irregularity of the longitudinal level over unit railway sections', *WIT Transactions on the Built Environment*, 114, pp. 641–650. doi: 10.2495/CR100591.

Chattopadhyay, G. and Kumar, S. (2009) 'Parameter estimation for rail degradation model', *International Journal of Performability Engineering*, 5(2), pp. 119–130.

Chen, Y., Lawrence, F. V, Barkan, C. P. L. and Dantzig, J. A. (2006) 'Weld defect formation in rail thermite welds', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. IMECHE, 220(4), pp. 373–384. doi: 10.1243/0954409JRRT44.

Cheng, Y. S. and Peng, C. Y. (2015) 'Integrated degradation models in R using iDEMO', *Journal of Statistical Software*, 49(2). doi: 10.18637/jss.v049.i02.

Choi, I. Y., Um, J. H., Lee, J. S. and Choi, H. H. (2013) 'The influence of track irregularities on the running behavior of high-speed trains', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(1), pp. 94–102. doi:

10.1177/0954409712455146.

Chopra, S. and Sodhi, M. S. (2004) ‘Managing risk to avoid supply-chain breakdown’, *MIT Sloan management review*. Massachusetts: Massachusetts Institute of Technology, p. 53.

Chopra, S. and Sodhi, M. S. (2014) ‘Reducing the risk of supply chain disruptions’, *MITSloan Management Review*. Massachusetts: Massachusetts Institute of Technology, pp. 73–80.

Clausen, J., Larsen, A., Larsen, J. and Rezanova, N. J. (2010) ‘Disruption management in the airline industry—concepts, models and methods’, *Computers & Operations Research*, 37(5), pp. 809–821. doi: http://dx.doi.org/10.1016/j.cor.2009.03.027.

CNN (2010) *Nine dead in Italy train derailment*. Available at: http://edition.cnn.com/2010/WORLD/europe/04/12/italy.train.derailment/ (Accessed: 25 August 2019).

Coetzee, J. L. (1997) ‘The role of NHPP models in the practical analysis of maintenance failure data’, *Reliability Engineering & System Safety*, 56(2), pp. 161–168. doi: 10.1016/S0951-8320(97)00010-0.

D’Lima, M. and Medda, F. (2015) ‘A new measure of resilience: an application to the London underground’, *Transportation Research Part A: Policy and Practice*. Elsevier Ltd, 81, pp. 35–46. doi: 10.1016/j.tra.2015.05.017.

Daniels, L. E. (2008) *Track Maintenance Costs on Rail Transit Properties*. Washington, DC: The National Academies Press. doi: 10.17226/23033.

Davies, R. (2015) *Malaysia floods-Kelantan flooding worst recorded as costs rise to RM1* *billion*. Available at: www.floodlist.com/asia/malaysia-floods-kelantan-worst-recorded-costs (Accessed: 7 June 2019).

Dindar, S., Kaewunruen, S., An, M. and Osman, M. H. (2016) 'Natural hazard risks on railway turnout systems', *Procedia Engineering*, 161, pp. 1254–1259. doi: 10.1016/j.proeng.2016.08.561.

Doherty, A., Dora, J. and Newsome, C. (2012) 'Enhancing resilience in Britain's railway infrastructure', *Proceedings of the Institution of Civil Engineers - Civil Engineering*, 165(6), pp. 20–26. doi: 10.1680/cien.11.00064.

Ebersöhn, W. and Selig, E. T. (1994) 'Use of track geometry measurements for maintenance planning', *Transportation Research Record: Journal of the Transportation Research Board*, (1470), pp. 84–92.

Eksioglu, B., Vural, A. V. and Reisman, A. (2009) 'The vehicle routing problem: a taxonomic review', *Computers & Industrial Engineering*, 57(4), pp. 1472–1483. doi: 10.1016/j.cie.2009.05.009.

Esveld, C. (2001) 'Track maintenance and renewal', in *Modern railway track*. 2nd edn. The Netherlands: MRT-Production, pp. 349–406.

European Standards (2008) *EN 13848-5:2008 Railway applications – track – track geometry quality – Part 5: geometric quality levels*.

European Standards (2009) *EN 13848-1:2009 Railways applications-track-track geometry quality-Part 1: characterization of track geometry*.

Fang, W., Yang, S. and Yao, X. (2015) 'A survey on problem models and solution approaches to rescheduling in railway networks', *IEEE Transactions on Intelligent Transportation Systems*, 16(6), pp. 2997–3016. doi: 10.1109/TITS.2015.2446985.

Farhangi, H., Konur, D., Long, S., Qin, R. and Harper, J. (2015) 'Bi-objective track inspection scheduling: formulation and solution analysis', in *Transportation Research Board 94th Annual Meeting*. Washington, DC, 11-15 January: The National Academies of Sciences, Engineering, and Medicine, pp. 1–18.

Farrow, S. (2017) 'Risk-informed benefit-cost analysis', in Abbas, A. E., von Winterfeldt, D., and Tambe, M. (eds) *Improving homeland security decisions*. Cambridge: Cambridge University Press, pp. 238–258. doi: DOI: 10.1017/9781316676714.011.

Fekete, S. P., Kroeller, A., Lorek, M. and Pfetsch, M. (2011) 'Disruption management with rescheduling of trips and vehicle circulations', in *5th ASME/ASCE/IEEE Joint Rail Conference*. Virginia, 16-18 March: American Society of Civil Engineers, pp. 1–14.

Filipe, C. L. and Teixeira, P. F. (2016) 'Predictive maintenance model for ballast tamping', *Journal of Transportation Engineering*. American Society of Civil Engineers, 142(4), p. 4016006. doi: 10.1061/(ASCE)TE.1943-5436.0000825.

Ford, D. (2018) *Schedule 8 - National Recalibration Methodology (Control Period 6)*. London.

Fox, M., Gerevini, A., Long, D. and Serina, I. (2006) 'Plan stability: replanning versus plan repair', in Derek Long, Smith, S. F., Borrajo, D., and McCluskey, L. (eds) *Proceedings of International Conference on AI Planning and Scheduling (ICAPS)*. Menlo Park, California, USA.: The AAAI Press, pp. 212–221.

FRA (2005) *Development of objective track quality indices*. Washington, DC.

Francis, R. and Bekera, B. (2014) ‘A metric and frameworks for resilience analysis of engineered and infrastructure systems’, *Reliability Engineering and System Safety*. Elsevier, 121, pp. 90–103. doi: 10.1016/j.ress.2013.07.004.

Freitas, M. A., Colosimo, E. A., Santos, T. R. Dos and Pires, M. C. (2010) ‘Reliability assessment using degradation models: Bayesian and classical approaches’, *Pesquisa Operacional*, 30(1), pp. 195–219. doi: 10.1590/S0101-74382010000100010.

Gilmore, L. and Valaika, J. (1992) *Operating and support cost-estimating guide*. Washington, D.C: U.S. Department of Defence.

Gligorijevic, D., Stojanovic, J. and Obradovic, Z. (2016) ‘Uncertainty Propagation in Long-Term Structured Regression on Evolving Networks’, in *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)*. CA, United States: AAAI Press, pp. 1603–1609.

Goldstein, M. and Wooff, D. (2007) *Bayes linear statistics: Theory and methods*, *Bayes Linear Statistics: Theory and Methods*. doi: 10.1002/9780470065662.

Guler, H. (2012) ‘A Decision Support System for Railway Track Maintenance and Renewal Management’, *Journal of Computing in Civil Engineering*, 27(June), pp. 292–306. doi: 10.1061/(ASCE)CP.1943-5487.0000221.

Haigermoser, A., Luber, B., Rauh, J. and Gräfe, G. (2015) ‘Road and track irregularities: Measurement, assessment and simulation’, *Vehicle System Dynamics*, 53(7), pp. 878–957. doi: 10.1080/00423114.2015.1037312.

Halcrow (2013) *ORR schedule 8 payment rates recalibration-Phase A.*

He, Q., Li, H., Bhattacharjya, D., Parikh, D. P. and Hampapur, A. (2013) ‘Railway Track

Geometry Defect Modeling: Deterioration, Derailment Risk, and Optimal Repair', *Transportation Research Board 92st Annual Meeting*, pp. 1–20. Available at: http://trid.trb.org/view.aspx?id=1242877.

Henry, D. and Emmanuel Ramirez-Marquez, J. (2012) 'Generic metrics and quantitative approaches for system resilience as a function of time', *Reliability Engineering and System Safety*. Elsevier, 99, pp. 114–122. doi: 10.1016/j.ress.2011.09.002.

Higgins, A., Ferreira, L. and Lake, M. (1999) 'Scheduling Rail Track Maintenance to Minimise Overall Delays', in Ceder, A. (ed.) *Transportation and Traffic Theory*. Amsterdam: Elsevier, pp. 779–796.

Holling, C. S. (1973) 'Resilience and Stability of Ecological Systems', *Annual Review of Ecology and Systematics*. Annual Reviews, 4(1), pp. 1–23. doi: 10.1146/annurev.es.04.110173.000245.

Hollnagel, E., Woods, D. and Leveson, N. (2006) *Resilience Engineering : Concepts and Precepts*, *Resilience Engineering: Concepts and Precepts*. Edited by E. Hollnagel, D. Woods, and N. Leveson. Ashgate Publishing Limited.

Hornik, K., Stinchcombe, M. and White, H. (1989) 'Multilayer feedforward networks are universal approximators', *Neural Networks*. doi: 10.1016/0893-6080(89)90020-8.

Hosseini, S. (2016) *Modeling and measuring resilience: Applications in supplier selection and critical infrastructure*. University of Oklahoma.

Hox, J. J. and Boeijie, H. R. (2005) 'Data collection, primary vs. secondary', *Encyclopedia of Social Measurement*. Elsevier Inc.

International Union of Railways (2008) *Best practice guide for optimum track geometry durability*. Paris, France: ETF-Railway Technical Publications.

Ishak, M. F., Dindar, S. and Kaewunruen, S. (2016) 'Safety-based maintenance for geometry restoration of railway turnout systems in various operational environments', in *Proceedings of The 21st National Convention on Civil Engineering*. Songkla, 28-30 June.

Iyer, M. and Rhinehart, R. R. (1999) 'A method to determine the required number of neural-network rraining repetitions', *Neural Networks, IEEE Transactions on*, 10, pp. 427–432. doi: 10.1109/72.750573.

Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G. and Nielsen, M. N. (2009) 'Disruption management in passenger railway transportation', in Ahuja, R. K., Möhring, R. H., and Zaroliagis, C. D. (eds) *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 399–421. doi: 10.1007/978-3-642-05465-5_18.

Johansson, P. and Nilsson, J. E. (2004) 'An economic analysis of track maintenance costs', *Transport Policy*, 11(3), pp. 277–286.

Kaewunruen, S. and Remennikov, A. M. (2006) 'Sensitivity analysis of free vibration characteristics of an in situ railway concrete sleeper to variations of rail pad parameters', *Journal of Sound and Vibration*, 298(1–2), pp. 453–461. doi: 10.1016/j.jsv.2006.05.034.

Kaewunruen, S. and Remennikov, A. M. (2009) 'Impact fatigue responses of pre-stressed concrete sleepers in railway track systems', *The IES Journal Part A: Civil & Structural Engineering*, 2(1), pp. 47–58. doi: 10.1080/19373260802479377.

Kaewunruen, S., Sussman, J. M. and Matsumoto, A. (2016) 'Grand challenges in transportation and transit systems', *Frontiers in Built Environment*, 2(4), pp. 1–5. doi: 10.3389/fbuil.2016.00004.

Kahneman, D. and Tversky, A. (1979) 'Prospect theory: an analysis of decision under risk', *Econometrica*. Econometric Society, 47(2), pp. 263–291. doi: 10.2307/1914185.

Karis, T. (2009) *Track irregularities for high-speed trains*. KTH Engineering Sciences.

Kashima, T. (2004) *Reliability-based optimization of rail inspection*. Massachusetts Institute of Technology.

Kendall, A., Keoleian, G. A. and Helfand, G. E. (2008) 'Integrated Life-Cycle Assessment and Life-Cycle Cost Analysis Model for Concrete Bridge Deck Applications', *Journal of Infrastructure Systems*, 14(3), pp. 214–222. doi: 10.1061/(ASCE)1076-0342(2008)14:3(214).

Khadem Sameni, M. (2012) *Railway track capacity: measuring and managing*. University of Southampton. Available at: https://orbit.dtu.dk/en/publications/railway-track-capacity-measuring-and-managing.

Khan, F. I. and Haddara, M. M. (2003) 'Risk-based maintenance (RBM): a quantitative approach for maintenance/inspection scheduling and planning', *Journal of Loss Prevention in the Process Industries*, 16(6), pp. 561–573. doi: 10.1016/j.jlp.2003.08.011.

Khan, M. S. and Pasha, G. (2009) 'The plotting of observations for the inverse Weibull distribution on probability paper', *Journal of Advanced Research in Statistics and Probability*, 1, pp. 11–22.

Khouy, I. A. (2013) *Cost-effective maintenance of railway track geometry: a shift from* *safety limits to maintenance limits*. Luleå University of Technology.

Khouy, I. A., Larsson-Kråik, P.-O., Nissen, A. and Kumar, U. (2016) 'Cost-effective track geometry maintenance limits', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 230(2), pp. 611–622. doi: 10.1177/0954409714542859.

Khouy, I. A., Larsson-Kråik, P. O., Nissen, A., Juntti, U. and Schunnesson, H. (2014) 'Optimisation of track geometry inspection interval', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 228(5), pp. 546–556. doi: 10.1177/0954409713484711.

Khouy, I. A., Schunnesson, H., Juntti, U., Nissen, A. and Larsson-Kråik, P.-O. (2013) 'Evaluation of track geometry maintenance for a heavy haul railroad in Sweden: a case study', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. IMECHE, 228(5), pp. 496–503. doi: 10.1177/0954409713482239.

Kim, S. and Frangopol, D. M. (2011) 'Cost-based optimum scheduling of inspection and monitoring for fatigue-sensitive structures under uncertainty', *Journal of Structural Engineering*, 137(11), pp. 1319–1331. doi: doi:10.1061/(ASCE)ST.1943-541X.0000365.

Klein, R. J. T., Nicholls, R. J. and Thomalla, F. (2003) 'The resilience of coastal megacities to weather related hazards: a review', in Kreimer, A., Arnold, M., and Carlin, A. (eds) *Building Safer Cities: The Future of Disaster Risk*. 3rd edn. Washington D.C, USA: The World Bank, pp. 111–137. Available at: http://eprints.soton.ac.uk/53205.

Kohl, N., Larsen, A., Larsen, J., Ross, A. and Tiourine, S. (2007) 'Airline disruption management—perspectives, experiences and outlook', *Journal of Air Transport*

Management, 13(3), pp. 149–162. doi: http://dx.doi.org/10.1016/j.jairtraman.2007.01.001.

Konur, D., Farhangi, H., Long, S., Qin, R., Elmore, C. and National Technical Information Center (2014) *Track inspection planning and risk measurement analysis*. Missouri: Missouri University of Science and Technology.

Kroon, L., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., Schrijver, A., Steenbeek, A. and Ybema, R. (2009) 'The new dutch timetable: the OR revolution', *Interfaces*, 39(1), pp. 6–17. doi: doi:10.1287/inte.1080.0409.

Kumar, R. and Gardoni, P. (2013) '16 - Stochastic modeling of deterioration in buildings and civil infrastructure', in Tesfamariam, S. and Goda, K. (eds) *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*. Woodhead Publishing (Woodhead Publishing Series in Civil and Structural Engineering), pp. 410–434. doi: https://doi.org/10.1533/9780857098986.3.410.

Kundu, D. and Howlader, H. (2010) 'Bayesian inference and prediction of the inverse Weibull distribution for Type-II censored data', *Computational Statistics and Data Analysis*, 54(6), pp. 1547–1558. doi: 10.1016/j.csda.2010.01.003.

Lake, M., Ferreira, L. and Murray, M. (2000) 'Minimising costs in scheduling railway track maintenance', in Brebbia, C. A., Allan, J., Hill, R. J., Sciutto, G., and Sone, S. (eds) *Seventh International Conference on Computers in Railways*. Bologna, 11-13 September: WIT Press, pp. 895–902.

Lam, J. Y. J. and Banjevic, D. (2015) 'A myopic policy for optimal inspection scheduling for condition based maintenance', *Reliability Engineering and System Safety*, 144, pp. 1–11. doi: 10.1016/j.ress.2015.06.009.

Lander, S. and Petersson, J. (2012) *Tamping planning in railway maintenance: improvement potential for Optram as decision support*. Chalmers University of Technology.

Lannez, S., Artigues, C., Damay, J. and Gendreau, M. (2015) 'A railroad maintenance problem solved with a cut and column generation matheuristic', *Journal Nerwork*, 66(1), pp. 40–56. doi: 10.1002/net.21605.

Lasisi, A. and Attoh-Okine, N. (2018) 'Principal components analysis and track quality index: a machine learning approach', *Transportation Research Part C: Emerging Technologies*, 91, pp. 230–248. doi: 10.1016/j.trc.2018.04.001.

Lawrence, S., Giles, C. L. and Tsoi, A. C. (1997) 'Lessons in neural network training: overfitting may be harder than expected', in *Proceedings of the Fourteenth National Conference on Artificial Intelligence*. Menlo Park, California, 27-31 July: AAAI Press, pp. 540–545.

Lee, B., Preston, F. and Green, G. (2012) *Preparing for high-impact, low-probability events: lessons from Eyjafjallajökull*. London. doi: ISBN 978 1 86203 253 8.

Lesley, L. (2009) 'Generalised costs and the value of time as a method of patronage forecasting', *Acta Technica Jaurinensis*, 2(1), pp. 57–68.

Leturiondo, U., Mishra, M., Galar, D. and Salgado, O. (2015) 'Synthetic data generation in hybrid modelling of rolling element bearings', *Insight: Non-Destructive Testing and Condition Monitoring*, 57(7), pp. 395–400. doi: 10.1784/insi.2015.57.7.395.

Levin, J. (1989) 'Choice under uncertainty', *European Journal of Operational Research*, 41(3), p. 369. doi: 10.1016/0377-2217(89)90259-2.

Li, J., Pollard, S., Kendall, G., Soane, E. and Davies, G. (2009) 'Optimising risk reduction: an expected utility approach for marginal risk reduction during regulatory decision making', *Reliability Engineering and System Safety*, 94(11), pp. 1729–1734. doi: 10.1016/j.ress.2009.05.005.

Lidén, T. (2014) *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*. Communications and Transport Systems, Department of Science and Technology, Linköping University. Available at: http://liu.diva-portal.org/smash/get/diva2:754942/FULLTEXT01.pdf.

Lidén, T. (2015) 'Railway infrastructure maintenance - a survey of planning problems and conducted research', *Transportation Research Procedia*, 10, pp. 574–583. doi: http://dx.doi.org/10.1016/j.trpro.2015.09.011.

Lin, Z. L., Huang, Y. S. and Fang, C. C. (2015) 'Non-periodic preventive maintenance with reliability thresholds for complex repairable systems', *Reliability Engineering and System Safety*, 136, pp. 145–156. doi: 10.1016/j.ress.2014.12.010.

Liu, R., Xu, P. and Wang, F. (2010) 'Research on a short-range prediction model for track irregularity over small track lengths', *Journal of Transportation Engineering*, 136(12), pp. 1085–1092. doi: 10.1061/(ASCE)TE.1943-5436.0000192.

Liu, X., Barkan, C. P. L. and Saat, M. R. (2011) 'Analysis of derailments by accident cause evaluating railroad track upgrades to reduce transportation risk', *Transportation Research Record Journal of the Transportation Research Board Transportation Research Board of the National Academies*, (2261), pp. 178–185. doi: 10.3141/2261-21.

Liu, X., Lovett, A. H., Dick, C. T., Saat, M. R. and Barkan, C. P. L. (2014) 'Optimization of ultrasonic rail-defect inspection for improving railway transportation safety and efficiency', *Journal of Transportation Engineering*, 61801(217), pp. 1–10. doi: 10.1061/(ASCE)TE.1943-5436.0000697.

Liu, X., Saat, M. and Barkan, C. (2012) 'Analysis of causes of major train derailment and their effect on accident rates', *Transportation Research Record: Journal of the Transportation Research Board*, 2289(December 2012), pp. 154–163. doi: 10.3141/2289-20.

Liu, X., Saat, M. R. and Barkan, C. P. L. (2011) 'Benefit-cost analysis of heavy haul railway track upgrade for safety and efficiency', in *Proceedings of the International Heavy Haul Association Conference*. Calgary, Canada, 19-22 June: IHHA, pp. 1–8. Available at: http://ict.uiuc.edu/railroad/oldsite_2011/CEE/pdf/Liu et al IHHA 2011 Track Class Upgrade Final.pdf.

Louwerse, I. and Huisman, D. (2014) 'Adjusting a railway timetable in case of partial or complete blockades', *European Journal of Operational Research*, 235(3), pp. 583–593. doi: http://dx.doi.org/10.1016/j.ejor.2013.12.020.

Lu, C. J. and Meeker, W. Q. (1993) 'Using degradation measures to estimate a time-to-failure distribution', *Technometrics*, 35(2), pp. 161–174. doi: 10.1080/00401706.1993.10485038.

Lu, Y. and Madanat, S. (1994) 'Bayesian updating of infrastructure deterioration models', *Transportation Research Record*, (1442), pp. 110–114. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0028530771&partnerID=40.

Mahmoudi, M., Elwany, A., Shahanaghi, K. and Gholamian, M. R. (2017) 'A delay time model with multiple defect types and multiple inspection methods’, *IEEE Transactions on Reliability*, 66(4), pp. 1073–1084. doi: 10.1109/TR.2017.2754519.

Meeker, W. Q., Escobar, L. A. and Lu, C. J. (1998) ‘Accelerated degradation tests: modeling and analysis’, *Technometrics*, 40(2), pp. 89–99. doi: 10.1080/00401706.1998.10485191.

Menezes, J. M. P. and Barreto, G. A. (2008) ‘Long-term time series prediction with the NARX network: an empirical evaluation’, *Neurocomputing*, 71(16–18), pp. 3335–3343. doi: 10.1016/j.neucom.2008.01.030.

Nan, C. and Sansavini, G. (2017) ‘A quantitative method for assessing resilience of interdependent infrastructures’, *Reliability Engineering and System Safety*, 157, pp. 35–53. doi: 10.1016/j.ress.2016.08.013.

Narayanaswami, S. and Rangaraj, N. (2011) ‘Scheduling and rescheduling of railway operations: a review and expository analysis’, *Technology Operation Management*, 2(2), pp. 102–122. doi: 10.1007/s13727-012-0006-x.

Network Rail (2008) *Rail infrastructure cost benchmarking*.

Network Rail (2014) *How we will get there: our operating environment, Annual report and accounts*. Network Rail.

Network Rail Limited (2016) *Network Rail Limited’s annual report and accounts 2016*. Available at: www.networkrail.co.uk/annualreport.

Von Neumann, J. and Morgenstern, O. (1944) *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Nielsen, L. K. (2011) *Rolling stock rescheduling in passenger railways: applications in* *short-term planning and in disruption management*. Erasmus University Rotterdam.

Nielsen, L. K., Kroon, L. and Maróti, G. (2012) ‘A rolling horizon approach for disruption management of railway rolling stock’, *European Journal of Operational Research*, 220(2), pp. 496–509. doi: http://dx.doi.org/10.1016/j.ejor.2012.01.037.

Nissen, A. (2009a) *Development of life cycle cost model and analyses for railway switches and crossings*. Luleå University of Technology.

Nissen, A. (2009b) ‘LCC-analysis for switches and crossings: a case study from the Swedish railway network’, *International Journal of COMADEM*, 12(2), pp. 10–19.

O’Connor, A. N., Modarres, M. and Ali Mosleh (2016) *Probability distributions used in reliability engineering*. Maryland, USA: Center for Risk and Reliability, University of Maryland.

Obrenovic, M., Jaeger, B. and Lemmer, K. (2006) ‘Methodology for the LCC-analysis and the optimal migration of the railway operations control on the example of ETCS’, *WIT Transactions on the Built Environment*, 88, pp. 255–264. doi: 10.2495/CR060261.

Office for Research and Experiments (1981) *Quantitative evaluation of geometry track parameters determining vehicle behavior: introductory study to the problem of assessing track geometry on the basis of vehicle response*. Utrecht, Netherlands: ORE Publisher.

Office of Rail Regulation (2012) *Review of European renewal and maintenance methodologies technical appendix number 1 -asset inspection, condition assessment and decision making*. Surrey. Available at: http://orr.gov.uk/__data/assets/pdf_file/0015/4812/pr08-konsins-290508.pdf.

Okoh, P. and Haugen, S. (2015) ‘Improving the robustness and resilience properties of maintenance’, *Process Safety and Environmental Protection*, 94, pp. 212–226. doi: 10.1016/j.psep.2014.06.014.

Osman, M. H. and Kaewunruen, S. (2018) ‘Execution time estimation of recovery actions for a disrupted railway track inspection schedule’, in Caspeele, R., Taerwe, L., and M. Frangopol, D. (eds) *Life-cycle analysis and assessment in civil engineering: Towards an integrated vision*. Leiden: CRC Press, Taylor & Francis Group, pp. 445–450.

Osman, M. H., Kaewunruen, S., Ann, M. and Dindar, S. (2016) ‘Disruption management of resource schedule in transportation sector: understanding the concept and strategy’, *Procedia Engineering*, 161, pp. 1295–1299. doi: 10.1016/j.proeng.2016.08.592.

Osman, M. H., Kaewunruen, S. and Jack, A. (2017) ‘Optimisation of schedules for the inspection of railway tracks’, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. IMECHE, p. 095440971772163. doi: 10.1177/0954409717721634.

Ottomanelli, M., Pace, P. and Pascoschi, G. (2002) *Intelligent decision support tools for optimal planning of rail track maintenance*. Italy: Polytechnic University of Bari.

Panda, A. and Gupta, R. K. (2014) ‘Making academic research more relevant: a few suggestions’, *IIMB Management Review*, 26(3), pp. 156–169. doi: 10.1016/j.iimb.2014.07.008.

Pant, R., Barker, K., Ramirez-Marquez, J. E. and Rocco, C. M. (2014) ‘Stochastic measures of resilience and their application to container terminals’, *Computers and Industrial Engineering*, 70(1), pp. 183–194. doi: 10.1016/j.cie.2014.01.017.

Pant, R., Barker, K. and Zobel, C. W. (2014) ‘Static and dynamic metrics of economic resilience for interdependent infrastructure and industry sectors’, *Reliability Engineering and System Safety*, 125, pp. 92–102. doi: 10.1016/j.ress.2013.09.007.

Peng, F. (2011) *Scheduling of track inspection and maintenance activities in railroad networks*. University of Illinois.

Peng, F., Ouyang, Y. and Somani, K. (2013) ‘Optimal routing and scheduling of periodic inspections in large-scale railroad networks’, *Journal of Rail Transport Planning & Management*, 3(4), pp. 163–171. doi: 10.1016/j.jrtpm.2014.02.003.

Piotrowski, A. P. and Napiorkowski, J. J. (2013) ‘A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling’, *Journal of Hydrology*, 476, pp. 97–111. doi: 10.1016/j.jhydrol.2012.10.019.

Pisoni, E., Farina, M., Carnevale, C. and Piroddi, L. (2009) ‘Forecasting peak air pollution levels using NARX models’, *Engineering Applications of Artificial Intelligence*, 22(4–5), pp. 593–602. doi: 10.1016/j.engappai.2009.04.002.

Podofillini, L., Zio, E. and Vatn, J. (2006) ‘Risk-informed optimisation of railway tracks inspection and maintenance procedures’, *Reliability Engineering & System Safety*, 91(1), pp. 20–35. doi: http://dx.doi.org/10.1016/j.ress.2004.11.009.

Putallaz, Y. and Rivier, R. (2003) ‘Strategic maintenance and renewal policy of a railway corridor, taking into account the value of capacity’, in *World Conference on Rail Research, WCRR 03*. Edinburgh, Scotland, 28 September - 1 October, pp. 1–18.

Rail Accident Investigation Branch (2014) *Derailment at Castle Donington, Leicestershire 21 January 2013, RAIB report*. Available at: https://aibcms.co.uk/sites/raib/cms_resources/110228_R042011v2_Moreton_on_Lugg.pdf.

Rail Safety and Standards Board (2014) ‘Taking safe decisions-how Britain’s railways take decisions that affect safety’. London: Rail Safety and Standards Board, p. 32. Available at: http://www.rssb.co.uk/Library/risk-analysis-and-safety-reporting/2014-guidance-taking-safe-decisions.pdf.

Randell, D., Goldstein, M., Hardman, G., Jonathan, P. and Troffaes, M. (2010) ‘Bayesian linear inspection planning for large-scale physical systems’, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 224(4), pp. 333–345.

Remennikov, A. M. and Kaewunruen, S. (2008) ‘A review of loading conditions for railway track structures due to train and track vertical interaction’, *Structural Control and Health Monitoring*, 15(2), pp. 207–234. doi: 10.1002/stc.227.

Renooij, S. and Van der Gaag, L. C. (1998) ‘Decision making in qualitative influence diagrams’, in Diane, J. C. (ed.) *Eleventh International Florida Artificial Intelligence Research Society Conference*. Snibel Island, Florida, 18-20 May: AAAI Press, pp. 410–414.

Revie, M., Bedford, T. and Walls, L. (2010) ‘Evaluation of elicitation methods to quantify Bayes linear models’, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 224(4), pp. 322–332. doi: 10.1243/1748006XJRR304.

Ribeiro, J. (2014) *Impact of vertical track stiffness variations in railway infrastructure maintenance*. Technico Lisbao.

Richmond, B. (2006) *Introduction to data analytics handbook*. Washington, D.C: Academy for Educational Development.

Roberts, C. *et al.* (2014) *Modular, self-inspecting infrastructure*. Paris: International Union of Railways.

Royal, A. (2017) ‘Dynamics in risk taking with a low-probability hazard’, *Journal of Risk and Uncertainty*, 55(1), pp. 41–69. doi: 10.1007/s11166-017-9263-1.

RSSB (2014) ‘Taking Safe Decisions-How Britain’s Railways Take Decisions That Affect Safety’. London: Rail Safety and Standards Board, pp. 1–32. Available at: http://www.rssb.co.uk/Library/risk-analysis-and-safety-reporting/2014-guidance-taking-safe-decisions.pdf.

Santos, R., Fonseca Teixeira, P. and Pais Antunes, A. (2015) ‘Planning and scheduling efficient heavy rail track maintenance through a decision rules model’, *Research in Transportation Economics*, 54, pp. 20–32. doi: 10.1016/j.retrec.2015.10.022.

Sato, K. and Fukumura, N. (2012) ‘Real-time freight locomotive rescheduling and uncovered train detection during disruption’, *European Journal of Operational Research*, 221(3), pp. 636–648. doi: http://dx.doi.org/10.1016/j.ejor.2012.04.025.

Setsobhonkul, S., Kaewunruen, S. and Sussman, J. M. (2017) ‘Lifecycle assessments of railway bridge transitions exposed to extreme climate events’, *Frontiers in Built Environment*, 3(35), pp. 1–20. doi: 10.3389/fbuil.2017.00035.

Sheffi, Y. (2005) *The resilient enterprise: overcoming vulnerability for competitive advantage*. Cambridge, Massachusetts: MIT Press.

Snowling, S. (2000) *Evaluation of modelling uncertainty and the role of model complexity in risk assessment, PhD Thesis*. McMaster University.

Soleimanmeigouni, I., Ahmadi, A. and Kumar, U. (2016) 'Track geometry degradation and maintenance modelling: a review', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, pp. 1–30. doi: 10.1177/0954409716657849.

Soleimanmeigouni, I., Ahmadi, A., Letot, C., Nilsson, A. and Kumar, U. (2016) 'Cost-based optimization of track geometry inspection', in *11th World Conference on Transport Research*. Milan, Italy, 29 May-2 June: UIC, pp. 1–7.

Soleimanmeigouni, I., Xiao, X., Ahmadi, A., Xie, M., Nissen, A. and Kumar, U. (2018) 'Modelling the evolution of ballasted railway track geometry by a two-level piecewise model', *Structure and Infrastructure Engineering*, 14(1), pp. 33–45. doi: 10.1080/15732479.2017.1326946.

Spliet, R., Gabor, A. F. and Dekker, R. (2014) 'The vehicle rescheduling problem', *Computers & Operations Research*, 43, pp. 129–136. doi: http://dx.doi.org/10.1016/j.cor.2013.09.009.

Steen, R. and Aven, T. (2011) 'A risk perspective suitable for resilience engineering', *Safety Science*, 49(2), pp. 292–297. doi: 10.1016/j.ssci.2010.09.003.

Stenström, C., Norrbin, P., Parida, A. and Kumar, U. (2015) 'Preventive and corrective maintenance – cost comparison and cost–benefit analysis', *Structure and Infrastructure Engineering*. Taylor & Francis, 2479(April), pp. 1–15. doi: 10.1080/15732479.2015.1032983.

Stenström, C., Norrbin, P., Parida, A. and Kumar, U. (2016) 'Preventive and corrective maintenance – cost comparison and cost–benefit analysis', *Structure and Infrastructure Engineering*, 12(5), pp. 603–617. doi: 10.1080/15732479.2015.1032983.

Stolker, R., Karydas, D. and Rouvroye, J. (2008) 'A comprehensive approach to assess operational resilience', in *Proceedings of the Third Resilience Engineering Symposium*. Antibes-Juan-les-Pins, France, 28-30 October: Presses Des Mines, pp. 247–253.

Sunstein, C. R. and Zeckhauser, R. (2011) 'Overreaction to fearsome risks', *Environmental and Resource Economics*, 48(3), pp. 435–449. doi: 10.1007/s10640-010-9449-3.

Tanaev, V., Gordon, W. and Shafransky, Y. M. (2012) *Scheduling theory single-stage systems*. Netherlands: Springer. Available at: https://books.google.co.uk/books?id=ITDyCAAAQBAJ.

Tanaev, V. S., Gordon, V. S. and Shafransky, Y. M. (1994) 'Introduction', in *Scheduling theory single-stage systems*. Dordrecht: Springer Netherlands, pp. 1–39. doi: 10.1007/978-94-011-1190-4_1.

The Guardian (2016) *Train crash in Belgium leaves three dead and nine injured*. Available at: https://www.theguardian.com/world/2016/jun/06/belgium-train-crash-dead-injured-passenger-freight (Accessed: 22 October 2019).

Thoft-Christensen, P. (2012) 'Infrastructures and life-cycle cost-benefit analysis', *Structure and Infrastructure Engineering*, 8(5), pp. 507–516. doi: 10.1080/15732479.2010.539070.

Thomas, P. J. (2016) 'Measuring risk-aversion: the challenge', *Measurement: Journal of the International Measurement Confederation*, 79, pp. 285–301. doi: 10.1016/j.measurement.2015.07.056.

Turner, C., Tiwari, A., Starr, A. and Blacktop, K. (2015) 'A review of key planning and scheduling in the rail industry in Europe and UK', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 230(3), pp. 984–998. doi:

10.1177/0954409714565654.

Tversky, A. and Kahneman, D. (1992) ‘Advances in prospect theory: cumulative representation of uncertainty’, *Journal of Risk and Uncertainty*, 5, pp. 297–323. doi: 10.1007/BF00122574.

UK Office of Rail and Road (2014) *Final determination of Network Rail’s outputs and funding for 2014-19*. Available at: http://orr.gov.uk/what-and-how-we-regulate/regulation-of-network-rail/how-we-regulate-network-rail/periodic-review-2013/pr13-publications/final-determination (Accessed: 12 June 2019).

UK Office of Rail and Road (2017) *Revenue per passenger kilometres and per passenger journey -Table 12.10*. Available at: https://dataportal.orr.gov.uk/statistics/usage/passenger-rail-usage/revenue-per-passenger-kilometres-and-per-passenger-journey-table-1210/ (Accessed: 13 July 2018).

Veelenturf, L. P., Kroon, L. G. and Mar´oti, G. (2014) ‘Passenger oriented railway disruption management by adapting timetables and rolling stock schedules’, in *10th International Conference of the Practice and Theory of Automated Timetabling*. York, United Kingdom, 26-29 August: Springer, pp. 11–34.

Vieira, G. E., Herrmann, J. W. and Lin, E. (2003) ‘Rescheduling manufacturing systems: a framework of strategies, policies, and methods’, *J. of Scheduling*, 6(1), pp. 39–62. doi: 10.1023/a:1022235519958.

Vugrin, E. D., Warren, D. E. and Ehlen, M. A. (2011) ‘A resilience assessment framework for infrastructure and economic systems: quantitative and qualitative resilience analysis of petrochemical supply chains to a hurricane’, *Process Safety Progress*, 30(3), pp. 280–290.

doi: 10.1002/prs.10437.

Walker, C. G., Snowdon, J. N. and Ryan, D. M. (2005) ‘Simultaneous disruption recovery of a train timetable and crew roster in real time’, *Computers & Operations Research*, 32(8), pp. 2077–2094. doi: http://dx.doi.org/10.1016/j.cor.2004.02.001.

Wang, X., Wu, X. and Hu, X. (2010) ‘A study of urgency vehicle routing disruption management problem’, in *Proceedings - 2010 WASE International Conference on Information Engineering, ICIE 2010*. Hebei, China, 14-15 August: Elsevier, pp. 452–455.

Wardman, M. and Batley, R. (2014) ‘Travel time reliability: a review of late time valuations, elasticities and demand impacts in the passenger rail market in Great Britain’, *Transportation*, 41(5), pp. 1041–1069. doi: 10.1007/s11116-014-9526-5.

Wardman, M. and Whelan, G. (2001) ‘Valuation of improved railway rolling stock: a review of the literature and new evidence’, *Transport Reviews*, 21(4), pp. 415–447. doi: 10.1080/01441640010020115.

Weber, E. U., Blais, A.-R. and Betz, N. E. (2002) ‘A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors’, *Journal of Behavioral Decision Making*, 15(4), pp. 263–290. doi: 10.1002/bdm.414.

Xu, P., Jia, C., Li, Y., Sun, Q. and Liu, R. (2015) ‘Developing an enhanced short-range railroad track condition prediction model for optimal maintenance scheduling’, *Mathematical Problems in Engineering*, 2015, p. 12. doi: 10.1155/2015/796171.

Xu, P., Sun, Q., Liu, R. and Wang, F. (2011) ‘A short-range prediction model for track quality index’, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 225(3), pp. 277–285. doi: 10.1177/2041301710392477.

Yan, H., Zhou, J. and Pang, C. K. (2015) 'Cost optimization on warning threshold and non-fixed periodic inspection intervals for machine degradation monitoring', in *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*. Yokohama, Japan, 9-12 November: IEEE, pp. 1079–1084. doi: 10.1109/IECON.2015.7392243.

Yang, L., Li, K. and Gao, Z. (2009) 'Train timetable problem on a single-line railway with fuzzy passenger demand', *IEEE Transactions on Fuzzy Systems*, 17(3), pp. 617–629. doi: 10.1109/TFUZZ.2008.924198.

Yang, L., Zhou, X. and Gao, Z. (2014) 'Credibility-based rescheduling model in a double-track railway network: a fuzzy reliable optimization approach', *Omega*, 48, pp. 75–93. doi: http://dx.doi.org/10.1016/j.omega.2013.11.004.

Yue, Z., Songzheng, Z. and Tianshi, L. (2011) 'Bayesian regularization BP neural network model for predicting oil-gas drilling cost', *BMEI 2011 - Proceedings 2011 International Conference on Business Management and Electronic Information*, 2, pp. 483–487. doi: 10.1109/ICBMEI.2011.5917952.

Zhang, R. and Mahadevan, S. (2000) 'Model uncertainty and Bayesian updating in reliability-based inspection', *Structural Safety*, 22(2), pp. 145–160. doi: 10.1016/S0167-4730(00)00005-9.

Zhang, R. and Mahadevan, S. (2003) 'Bayesian methodology for reliability model acceptance', *Reliability Engineering and System Safety*, 80(1), pp. 95–103. doi: 10.1016/S0951-8320(02)00269-7.

Zhang, T., Andrews, J. and Wang, R. (2013) 'Optimal scheduling of track maintenance on a railway network', *Quality and Reliability Engineering International*, 29(2), pp. 285–297.

doi: 10.1002/qre.1381.

Zhao, J., Chan, A. H. C. and Burrow, M. P. N. (2007) ‘Reliability analysis and maintenance decision for railway sleepers using track condition information’, *Journal of the Operational Research Society*, 58(8), pp. 1047–1055. doi: 10.1057/palgrave.jors.2602251.

Zhen, X., Li, Y., Cai, G. and Shi, D. (2016) ‘Transportation disruption risk management: business interruption insurance and backup transportation’, *Transportation Research Part E: Logistics and Transportation Review*, 90, pp. 51–68. doi: http://dx.doi.org/10.1016/j.tre.2016.01.005.

Zoeteman, A. (2001) ‘Life cycle cost analysis for managing rail infrastructure’, *European Journal of Transport and Infrastructure Research*, 1(4), pp. 391–413. Available at: http://www.ejtir.tbm.tudelft.nl/issues/2001_04/index.asp.